Department of Economics Working Paper

Number 21-07 | July 2021

Hazardous Waste and Home Values: An Analysis of Treatment and Disposal Sites in the U.S.

Dennis Guignet
Appalachian State University

Christoph Nolte
Boston University

Department of Economics
Appalachian State University
Boone, NC 28608
Phone: (828) 262-2148
Fax: (828) 262-6105
www.business.appstate.edu/economics

Hazardous Waste and Home Values: An Analysis of Treatment and Disposal Sites in the U.S.

Dennis Guignet*

Department of Economics

Appalachian State University

and

Christoph Nolte

Department of Earth and Environment

Boston University

Last Revised: July 6, 2021

*Corresponding Author: Department of Economics, Appalachian State University, 416 Howard Street, ASU Box 32051, Boone, NC, 28608-2051. Ph: 828-363-2117. guignetdb@appstate.edu.

Hazardous Waste and Home Values:
An Analysis of Treatment and Disposal Sites in the U.S.

By:

Dennis Guignet and Christoph Nolte

Abstract:

The Resource Conservation and Recovery Act (RCRA) is a cornerstone of environmental policy in the United States (US). It regulates the generation, use, transportation, and eventual disposal of hazardous chemicals. Focusing on the 2,389 treatment, storage, and disposal facilities (TSDFs) in the contiguous US, we frame a difference-in-differences and triple differences quasi-experiment that exploits the temporal and spatial variation in contamination and cleanup events. Hedonic property value regressions are estimated using a sample of over 9.6 million single-family home transactions from 2000-2018. The discovery of contamination and subsequent investigation is associated with up to a 5% depreciation in the value of homes within 750 meters of a TSDF, but the evidence is mixed. In contrast, we find robust, causal evidence that the completion of cleanup leads to an average 6-7% increase in the value of homes within 750 meters. This implies that a total increase in housing stock value of $323 million can be attributed to the 195 TSDFs that have been remediated since the inception of the RCRA cleanup program. The completion of cleanup at a TSDF is estimated to yield an average lower bound, ex post benefit of about $8,400 per household.

This research is supported in part by the first author's appointment to the U.S. Environmental Protection Agency (EPA) Research Participation Program administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy (DOE) and the EPA. ORISE is managed by ORAU under DOE contract number DE-SC0014664. Property transaction data were provided by Zillow through the Zillow Transaction and Assessment Dataset (ZTRAX). More information on accessing the data can be found at http://www.zillow.com/ztrax. Additional data support was provided by the Private-Land Conservation Evidence System (PLACES) at Boston University. We thank Shelby Sundquist for excellent research assistance. We are grateful to the EPA's Office of Resource Conservation and Recovery, and especially to Norman Birchfield, David Hockey, Joseph Krahe, Jennifer McLeod, and Scott Palmer for sharing their insights of the RCRA program and data. Finally, we thank seminar participants at San Diego State University and the University of Calgary, as well as participants of the 2020 and 2021 PLACES webinars and NAREA 2021 virtual conference, for helpful comments. The results and opinions are those of the authors and do not reflect the position or policies of the US EPA, DOE, ORAU/ORISE, or the Zillow Group.

I. INTRODUCTION

The Resource Conservation and Recovery Act (RCRA) is a cornerstone of environmental policy in the United States (US). It regulates the generation, use, transportation, and eventual disposal of hazardous chemicals. The 1984 amendments to RCRA established requirements that facilities investigate, and if necessary, cleanup, any hazardous releases into the environment. Such investigations and cleanup activities are collectively referred to as a Corrective Action. Releases of RCRA regulated hazardous chemicals are truly ubiquitous, with over 35 million people (or 12% of the US population) living within one mile of a RCRA Corrective Action site (US EPA, 2013). Cleaning up and reducing exposure to these hazardous chemicals (including arsenic, benzene, lead, mercury, solvents, and many others) reduces risks to human health and the environment.

Despite the fact that RCRA and the Corrective Action program have been in place for over three and a half decades, there is little known about the monetized benefits of the program. Many benefit endpoints of US Environmental Protection Agency (EPA) regulations under the authority of RCRA go unmonetized in benefit-cost analyses informing these interventions. As summarized in Table 1, over the last ten years, with the exception of the Regulatory Impact Analysis (RIA) of the 2015 Coal Combustion Residuals Rule, the only monetized benefits entail avoided regulatory or cleanup costs to regulators and industry. Welfare impacts to surrounding residents go unmonetized, and are represented solely in a qualitative fashion.

The objective of this nationwide study is to provide insight towards the monetized benefits of RCRA programs. Focusing on all hazardous waste treatment, storage, and disposal facilities (TSDFs) across the contiguous US, we estimate a series of hedonic property value models. We set out to answer three primary questions. First, is proximity to a TSDF associated with lower home values? Second, does the discovery of hazardous contamination and opening of a Corrective Action (CA) investigation further decrease surrounding home values? And if so, does the subsequent cleanup and completion of a CA lead to a rebound in home values? Third, what are the welfare implications to households? We adapt an approach recently proposed by Banzhaf (Forthcoming) to obtain lower bound estimates of the ex post benefits to impacted households.

To facilitate a causal interpretation of the estimated price and welfare effects, we exploit temporal and spatial variation in contamination and cleanup events, framing the empirical setting as two alternative difference-in-differences comparisons – utilizing both intra- and inter-neighborhood counterfactuals – as well as a triple differences quasi-experiment. In subsequent models we use exact covariate and coarsened exact matching (CEM) techniques to make the treatment and control groups more comparable, and ultimately to provide more defensible causal estimates of how the opening of a CA, and subsequent cleanup and completion of the CA, impact home values and households.

A national dataset of 2,389 TSDFs, and details about the CAs and cleanup activities occurring at 689 of these facilities, is compiled from RCRAInfo (an extensive and novel dataset maintained by the EPA). These data are spatially and temporally linked to individual residential property transactions data from Zillow (2020), and supplemented with high resolution spatial data of residential parcels and nearby amenities compiled by the Private-Land Conservation Evidence

System (PLACES) at Boston University. Our dataset covers the contiguous US, and entails 9.6 million transactions from 2000 to 2018.

We find clear evidence that proximity to a TSDF is associated with lower home values. This negative association is strongest nearest a TSDF (i.e., within 250 meters) – suggesting an 8% lower value, all else constant. Although this negative association diminishes with distance, it remains significant out to at least 4.5 kilometers. Furthermore, the quasi-experimental results suggest that the value of residential properties within 750 meters of a TSDF are impacted by CA activities. Our base models suggest an additional 5% decline in value when a CA is opened, but the evidence is mixed and interpreting this as a causal relationship is questionable. In contrast, we find robust evidence that once cleanup is complete and the CA is resolved, home values rebound, with an average appreciation of 5-7%. Data diagnostics support a causal interpretation of this post-CA increase in home values. This translates to an average increase in value of $12,690 (2019$) per home. Extrapolating this result to all single-family homes within 750 meters of one of the 195 TSDFs that has been cleaned up since the inception of the RCRA CA program suggests a total increase in housing stock value of $323 million. Finally, we estimate that the completion of cleanup leads to an average lower bound, ex post benefit of $8,395 per household.

Next, we briefly review the relevant hedonic property value literature on proximity to contaminated sites and cleanup, and then in section III provide details on the RCRA CA program and data. Section IV describes the empirical methods, and the results are presented in section V. Section VI outlines the approach for estimating a lower bound of the ex post welfare effects, and presents the results. Concluding remarks, caveats, and directions for future research are discussed in section VII.

II. LITERATURE

Kohlhase (1991) and Michaels and Smith (1990) were among the first to study how proximity to waste sites impacts home values. The ensuing literature on the topic is large and well-established, with several literature reviews conducted (Farber, 1998; Jackson, 2001; Boyle and Kiel, 2001). A recent meta-analysis by Schütt (2021) revisited the literature and synthesized the results across 83 different hedonic studies. Schütt concludes that the literature generally suggests that proximity to waste sites does lower home values, all else constant, and that this adverse effect is strongest among the most severe and hazardous sites. Schütt also finds that cleanup of the contamination reduces these adverse price effects.

The rise of quasi-experimental methods in environmental economics (Greenstone and Gayer, 2009), and in particular to hedonic property value applications (Parmeter and Pope, 2013), has led several researchers to re-examine the causality of the empirical relationship between home values and proximity to contaminated sites. Such efforts include a series of national-scale studies on various cleanup programs headed by the EPA's Office of Land and Emergency Management (OLEM). Gamper-Rabindran and Timmins (2013) use decadal census tract data to examine the impact of Superfund site cleanups on home values. Superfund sites are those that fall under the jurisdiction of, and are subsequently cleaned up under, the Comprehensive Environmental Response, Compensation and Liability Act (CERCLA), and are among the most notorious and severe hazardous waste sites in the US. Gamper-Rabindran and Timmins find that homes within three miles (about 4.8 km) appreciate by roughly 20% when a superfund site is cleaned up, and that this effect is strongest among the lowest quantiles of the tract-specific price distributions.

A difference-in-differences (DID) hedonic property value study by Haninger et al. (2017) utilized individual home transaction data to examine how remediation through EPA's brownfield cleanup grants program impacts nearby home values. They find that cleanup leads to a 5-15% appreciation among homes within two kilometers. The DID and triple differences estimation approaches we later describe in this study are similar to those implemented by Haninger et al. (2017), as well as to a study by Muehlenbachs et al. (2015) that examined the residential property value impacts from unconventional natural gas wells (i.e., fracking sites).

Focusing on the most severe and well-publicized releases of petroleum contamination from underground storage tanks at gas stations, Guignet et al. (2018) estimate a series of site-specific DID hedonic price equations, and then synthesize the results using an internal meta-analysis. They found an average depreciation of 6% in value among homes within three kilometers of a release, and that home values fully rebound, on average, after the contamination is cleaned up.

Together, these studies suggest that EPA's various programs to clean up hazardous waste and contamination do benefit surrounding communities. Nonetheless, a comprehensive nationwide study of the property value impacts of the broadest law regulating the use, transportation, storage, and disposal of hazardous material remains absent from the literature. Focusing on the highest risk and most hazardous category of RCRA sites – Treatment, Storage, and Disposal Facilities (TSDFs) – we set out to fill this gap.

A working paper by Cassidy et al. (2020) also examines how cleanups under RCRA's CA program impacts home values. Using decadal census tract-level data and a methodology similar to Gamper-Rabindran and Timmins (2013), Cassidy et al. find that after cleanup home values in the same census tract as the RCRA site increase in value by up to 11%. Similar to Gamper-Rabindran and Timmins, they also find that these price effects are strongest among homes falling in the lowest portions of the tract-specific price distributions.

Our study and Cassidy et al.'s (2020) working paper are useful complements. Although Cassidy et al. provide a comprehensive analysis, their use of temporally and geographically coarse decadal Census tract-level data naturally inhibits identification of very local and possibly dynamic impacts on home values. Another potential drawback is that the Census data is based on self-reported home values.[1] In contrast, our analysis uses actual prices from individual residential property transactions. We know the location of each individual home and the date of the transaction. This

[1] Kiel and Zabel (1999) found that self-reported home values tend to be 5.1% higher than actual values, on average. Although to be fair, Kiel and Zabel conclude that such differences are not related to house or neighborhood characteristics, and so regression models based on self-reported values can yield reliable implicit price estimates of house and location attributes.

finer temporal and spatial resolution allow us to identify very local effects, and to examine how both the opening of a CA, and subsequent cleanup and completion of the CA, impact home values. A higher spatial resolution also lets us to better isolate the price effects of interest from other spatially correlated factors affecting home values.

We utilize data on all 2,389 TSDFs in the contiguous US, and not just those where a CA occurs. This allows us to construct and compare various counterfactuals. More specifically, we illustrate an inter-area DID estimate that compares home values around a "control" set of TSDFs where a CA does not occur. Those DID estimates are compared to more conventional intra-area DID estimates that use homes at farther distances around the same TSDF as the counterfactual (e.g., Linden and Rockoff, 2008; Haninger et al., 2017; Guignet et al., 2018; Muehlenbachs et al., 2015). Comparison of these intra- versus inter-area DID estimates can inform future applications where data or context may be more conducive to one framework over the other. Finally, we exploit both layers of controls in a series of triple differences estimates.

III. BACKGROUND AND DATA

To answer our research questions and provide insights toward the monetized benefits of RCRA programs, we compile and combine two detailed and comprehensive nationwide datasets – one of all RCRA sites and Corrective Actions (CAs), and the other of the majority of residential properties and transactions in the US from 2000 through 2018.

III.A. TSDFs and Corrective Actions

A hazardous waste treatment, storage, and disposal facility (TSDF) is a permanent establishment in operation for the purpose of waste management (US EPA, 2005). Examples of TSDFs include active landfills, lead battery recycling plants, storage facilities for cleaning bi-products, and former steel mills, brass foundries, and plumbing manufacturers. We focus on TSDFs because they are the riskiest and most hazardous type of RCRA sites. About 43% of all CAs in our data involve a facility that is registered as a TSDF. Almost 30% of all TSDFs are involved in a CA. TSDFs face increased requirements to prevent releases and exposures (US EPA, 2005), and are among the highest priority for EPA's Office of Resource Conservation and Recovery (ORCR).[2]

A dataset of all RCRA regulated facilities was first compiled from RCRAInfo – an extensive multi-relational database compiled and maintained by the EPA. The database includes eight distinct "modules", each containing numerous data tables.[3] Our dataset pulls facility-level information from the "Handler" module, which includes information on whether the facility generates,

---

[2] Although we initially set out to examine all 129,670 RCRA sites, computational constraints when merging these data to all residential transactions throughout the US made data processing and analysis impractical.
[3] EPA, "RCRAInfo Public Extract", https://rcrapublic.epa.gov/rcra-public-export/, data downloaded in June, 2019.

transports, stores, or disposes of hazardous materials, what hazardous materials the facility is registered to handle, and the type of business or industry the facility is in, among other things. Information on the timing and severity of CAs that occurred at these RCRA sites was obtained from the "Corrective Action" module. This included the opening and completion dates of a CA, and the type of remediation that took place. The initial investigations under a CA can potentially reveal that no contamination issue needs to be addressed. We draw focus to CAs that involve actual pollution concerns that needed to be addressed in some fashion. Such CAs are identified as those that later required physical remediation, groundwater or non-groundwater controls (e.g., caps, physical barriers), and/or institutional controls. Geographic coordinates for the RCRA sites were obtained from EPA's Federal Registry Service (FRS) database.[4] In the end a dataset of 129,670 RCRA sites across the entire US was compiled. This includes all sites that were confirmed small or large quantity generators, recyclers, and/or TSDFs for at least one federally regulated hazardous material.[5] CAs involving contamination releases were identified at 1,610 of these sites.

In this analysis we focus on the 2,389 TSDFs in the contiguous US, and the CAs that occurred at 689 of these sites (see Figure 1). The 1,700 TSDFs where a CA involving actual contamination did *not* occur, serve as a control group. At 195 of the 689 CA sites, the CA was completed by the end of 2018. This means that the site-specific remediation and/or exposure mitigation goals were achieved, or were sufficiently underway so that the contamination was no longer deemed a threat.

Table 2 shows that many of the 2,389 TSDFs (43%) are also registered as large and/or small quantity generators of hazardous materials.[6] About 14% of the TSDFs accept hazardous materials from other (offsite) facilities for treatment, storage, or disposal. Roughly 10% of the TSDFs also recycle one or more hazardous materials for subsequent re-use. For example, spent solvents (e.g., acetone) and hazardous metals (e.g., lead) are sometimes processed in order to reclaim usable product.[7] On average, a TSDF is registered as dealing with 57 different hazardous materials, but this varies greatly across sites and is heavily skewed. The median facility is registered as using, storing, disposing of, or transporting eight different hazardous chemicals.

As shown in Table 3, based on the available two-digit North American Industry Classification System (NAICS) codes, the majority of TSDFs are involved in the manufacturing industry (60%), followed by services (31%), waste management (25%), and trade, transportation and warehousing (24%). These categories are not mutually exclusive; some facilities report multiple NAICS codes. There are some, but far fewer facilities associated with utilities, construction, mining, and agriculture/forestry.

---

[4] EPA, "FRS Data Resources", https://rcrapublic.epa.gov/rcra-public-export/, data downloaded in August, 2019.

[5] The dataset excludes 657 facilities falling under the conditional subpart K exemptions, which apply to academic institutions, such as teaching hospitals and universities (US EPA, 2008). We also exclude 48 facilities dealing with international imports and exports of hazardous waste.

[6] A large quantity generator is defined as a facility that generates 1,000 kg or more of RCRA-regulated hazardous material in a single month, accumulates 1 kg or more of an acute hazardous material in a single month, or accumulates more than 100kg of spill material that contains an acute hazardous material (US EPA, 2021b).

[7] In 2019, over one million tons of hazardous metals and 234,000 tons of solvents were recovered and reused for productive purposes (US EPA, 2021b).

Under the 1984 congressional amendments to RCRA, a CA is required when there is an identified release of hazardous material, or when EPA is considering an application for a TSDF RCRA permit (US EPA, 2020). The RCRAInfo data include several different types of events and milestones associated with the CA process. Although the database is standardized, event dates are sometimes missing, or certain events are not applicable or always inputted in a consistent fashion. After numerous discussions with EPA's ORCR, we define the opening of a Corrective Action as the earliest of the following three events: when regulators were first notified of contamination, the date an investigation was imposed, or when the initial assessment was complete.[8] We focus only on CAs where the investigation revealed actual contamination that required clean up or active efforts to minimize exposure, as described below.

Descriptive statistics for the CAs that occurred at 689 of the TSDFs are shown in Table 4. Almost all CA sites (98%) are also on the EPA and state regulators' list of high priority RCRA cleanups (US EPA, 2019). The average CA has a risk prioritization between "medium" and "high". By definition, all CAs of interest involved one or more of the following three approaches to address the hazardous waste contamination. In 76% of the cases, active remediation technologies are constructed and implemented (e.g., excavation of contaminated soil, pump-and-treat groundwater). Physical and engineering controls to minimize exposure and migration of contaminants are employed at 41% of the CA sites. Such controls include routine monitoring and testing of groundwater, concrete caps, and physical barriers. Institutional controls (e.g., zoning restrictions on future land uses) are put in place at 49% of the CA sites.[9] It is not uncommon to have two (172 CAs) or even all three (141 CAs) of the approaches in place to remediate contamination and minimize exposure. For 195 of the 689 CAs (28%), the investigation and cleanup was complete by 2018 (the end of our study period). We define the completion of a CA as the earliest recorded date for one of the following events, and only when applicable to the entire site – when the CA was officially terminated, or when the required performance standards were achieved.[10] Although the cleanup goals have been achieved and active cleanup efforts have ceased by this point, some institutional or physical controls, as well as routine monitoring, may still remain in place. The average (and median) duration of a CA was 15 years, but there is noticeable variation in this timing, ranging from 0 to almost 31 years.

The opening and completion of the CAs over time are shown in Figure 2. There is noticeable variation in when these investigations and cleanup activities are first initiated and subsequently resolved across the sites. This temporal variation, coupled with the spatial variation in the location of these CA sites (see Figure 1), bolsters the quasi-experimental design discussed in section IV. Given the large number of sites spread across such a vast study area, with CAs being initiated and completed in different years, it is increasingly less plausible that potential confounders could be

---

[8] This corresponds to the RCRAInfo event codes CA060, CA100, and CA200, respectively.

[9] The implementation of active remediation technologies, physical controls, and institutional controls are identified by the following event codes: CA550RC (remedy construction); CA770GW and CA770NG (groundwater and non-groundwater controls); and CA772EP, CA772GC, CA772ID, and CA772PR (institutional controls).

[10] This corresponds to the RCRAInfo event codes CA999, CA999NF, CA999RM; and CA900CR and CA900NC, respectively. We thank Jennifer McLeod at EPA for providing data of when these events applied to the entire facility (Personal Communication, 19 Nov 2019).

systematically correlated with both the location *and* timing of all contamination and cleanup events.

Finally, although we do not know what hazardous materials were released at each of the 689 CA sites, we do know what chemicals were registered for use. Figure 3 shows the 15 most common chemicals registered at the 689 CA sites. The general categories of ignitable and corrosive waste are the most common, followed by spent nonhalogenated solvents (e.g., xylenes, acetone, ethyl benzene), which are associated with increased risks of neurological damage, nausea and headaches, kidney damage, and cancer (US EPA, 2021a; US HHS, 1994, 2010). Lead is the most common individual chemical, and is associated with numerous adverse health outcomes, including cancer, kidney and cardiovascular problems, and even mortality in adults (US HHS, 2020; US EPA, 2013). Lead is also well-known to cause cognitive deficits in children (Aizer et al., 2018; US HHS, 2020; Lanphear et al., 2005; Miranda et al., 2007), and recent evidence suggests such deficits can persist for years (Shadbegian et al., 2019).[11] Other heavy metals like cadmium and mercury are associated with an increased risk of kidney damage; and contaminants like benzene, tetrachloroethylene, and arsenic are carcinogenic and associated with increased risks of cardiovascular issues (US EPA, 2013, 2021a).

In order to examine whether perceived health risks, and other factors like displeasing aesthetics and degraded environmental quality, are capitalized into local home prices, we spatially and temporally link the TSDF and CA data to residential property transactions.

III.B. Residential Properties and Transactions

Information on all residential parcels and transactions across most of the contiguous US come from Zillow's ZTRAX program (ZTRAX, 2020). The ZTRAX data include information on transaction prices and dates, types of transactions, addresses, types of homes, and housing characteristics (e.g., number of bathrooms, interior square footage, year built). The geographic coordinates provided by ZTRAX are sometimes incomplete, and depending on the spatial scale of the environmental commodity of interest, these coordinates may not be sufficiently precise. Given the localized nature of disamenities like RCRA sites, the ZTRAX data are supplemented with higher-precision digital parcel maps obtained from twelve open-source state-level datasets and by data from two commercial providers (Loveland and Boundary Solutions, Inc.). Details on these data and data sources are provided by Nolte (2020).

The geographic data was processed using the Private-Land Conservation Evidence System (PLACES). PLACES uses assessor parcel numbers to link ZTRAX data to parcel boundaries based on county and town-specific string pattern matching and geographic quality controls (for details see Nolte, 2020). Within a Geographic Information System (GIS), residential parcels are spatially linked to the corresponding 2010 census tract. Distance to the nearest highway, and the proportion

[11] Focusing on a hazardous waste site in Chile that contained several of the aforementioned contaminants, Rau et al. (2015) found that children attending school within one-kilometer experience a significant decline in academic performance, and that this translates to a $60,000 loss in lifetime earnings for each child.

of land within a 200m and 500m circular buffer around each individual home, are also calculated.[12] Most importantly, the Euclidean distance from each residential parcel to each TSDF within 10km was calculated.

Our analysis focuses on all arms-length, single-family home transactions in the contiguous US from 2000-2018; where the home was within five kilometers of a TSDF. After eliminating outliers, the final sample size is n=9,655,158 transactions.[13] The study area covers 836 counties across 47 States and Washington, D.C. (see Figure 4).

Descriptive statistics of the residential transactions are presented in Table 5. The mean sales price is about $273,000 (2019$). The average home has a lot size of about 0.25 acres, 1.4 stories, 1.86 bathrooms, an interior square footage of 2,953 sq ft., and is about 46 years in age. There are missing values for all of these house and parcel characteristics. As can be seen by the companion missing value dummies, data on acreage and interior square footage is only missing for 2.6% and 5.1% of the sample, respectively. House age is missing for about 6.8% of the sample. A larger proportion of the sample is missing data on the number of stories (19.0%) and bathrooms (23.1%). When applicable, observations with missing values are excluded from the corresponding statistics in Table 5. In the later analysis, missing values are coded as zero and companion missing dummy variables are included as right-hand side variables in the regression models.

Turning to the location variables in Table 5, the proportion of land area that is developed within 200m and 500m of a parcel is 55% and 50%, on average; and about 39% of transactions are of homes within 500m of a highway. On average, there are 0.03 TSDFs within 750m of a home, and 0.14 and 1.73 within 1,500m and 5,000m, respectively. Figure 5 displays the number of sales in each 250m incremental bin from 0-250m to 4,750-5,000m. There are clearly a large number of transactions in each distance bin. Even among the smallest bins nearest to a TSDF, there are several thousand sales. For example, there are 18,988 sales within 0-250m of a TSDF, and 84,103 within 250-500m. Beyond that, there are well over 100,000 transactions in each distance bin.

Figure 6 shows the number of transactions in each 250-meter incremental distance bin during each of the three CA stages – pre-Corrective Action, mid-Corrective Action, and post-Corrective Action. There is a large number of observations in many of the bins, especially during an open CA. This is not surprising since the average CA is opened for 15 years. Even in the nearest distance bins, there are often at least a few hundred transactions observed. The relatively small sample size in some of the nearest distance-by-stage bins illustrates the benefit of conducting such a large nationwide study. Using a large-scale dataset allows us to obtain a sufficient number of identifying observations for statistical analysis, while still maintaining a flexible specification of the distance

---

[12] Census tract data comes from the National Historical Geographic Information System (Manson et al., 2018). Land cover data comes from the 2011 National Land Cover Database (Dewitz, 2019). The highways data is from the US Census Bureau's TIGER/Line shapefiles (2019).

[13] Outlying observations were dropped as follows: homes with a real price (2019$) less than $20k or more than $1M (corresponds roughly to the 5th and 95th percentiles), less than 0.05 acres or greater than 2 acres (corresponds roughly to the 5th and 95th percentiles), less than one or greater than three stories, an interior size of less than 750 sqft. or greater than 12,000 sqft. (corresponds roughly to the lowest and highest percentiles), less than one or greater than six bathrooms, and if the home was greater than 150 years in age.

gradient. In the empirical analysis we use a series of 250-meter incremental bin indicator variables to allow the distance gradient to freely vary across bins without imposing any functional form assumptions. Such a flexible and detailed price gradient specification would not be possible in a more local analysis where less transactions are observed.

IV. METHODS

IV.A. Difference-in-differences and Triple Differences Framework

The usefulness of the DID strategy for causal inference in the hedonic pricing framework was made clear through early applications by Davis (2004) and Linden and Rockoff (2008), among others. There have since been numerous hedonic applications of the DID approach to estimate the impact of local environmental commodities on surrounding home values, including studies of contaminated site cleanups (Haninger et al., 2017; Guignet et al., 2018), aquatic vegetation and invasive species (Guignet et al., 2017; Horsch and Lewis, 2009), flood and hurricane risks (Atreya et al., 2013; Bin and Landry, 2013), and shale gas extraction (Muehlenbachs et al., 2015). The approach remains at the forefront of hedonic pricing methods (Guignet and Lee, 2021).

The intuitive nature of the DID and triple differences strategies in mimicking a classical experimental design make the approaches particularly attractive. Consider the simplified illustration in Figure 7. The top row represents homes in the neighborhood around a TSDF where contamination and a CA will eventually occur but has not yet (top-left panel), or has occurred (top-right panel). The bottom row represents homes near a counterfactual TSDF where contamination and a CA does not occur. The price effect of interest is experienced by homes in group *B*. These are homes in close proximity to a TSDF after a CA is initiated. In quasi-experimental terms, group *B* denotes the treated group, post-treatment. The objective is to identify how contamination and the opening of a CA impacts home values. A simple cross-sectional first difference estimate would be just *B-B'*. However, the TSDFs and surrounding neighborhoods where contamination and CAs do and do not occur could be systematically different.

Our *inter-area* DID estimate addresses this by exploiting temporal variation in the initiation of CAs. This second differencing comes from comparing the change in home values before and after the CA is opened *(B-A)*, and the change in home values during this same time period around the counterfactual TSDFs where a CA does not occur *(B'-A')*. The inter-area DID estimates are: *(B-A)-(B'-A')*.

One could speculate the possibility that *trends* in the neighborhoods where contamination does and does not occur also differ over time. Our triple differences estimate addresses that possibility by further differencing out broader neighborhood trends based on homes around the same TSDFs, but that are presumably far enough away from the TSDF to not be impacted by the contamination and CA activities. Our triple differences estimate in this simple illustration is represented as: *(B-A)-(B'-A')-{(D-C)-(D'-C')}*.

In a more focused examination, we later drop transactions that are only near TSDFs where a CA never occurs (i.e., we disregard the bottom row in Figure 7). This provides an alternative and arguably cleaner quasi-experiment by focusing on a more homogenous set of homes and neighborhoods. In doing so, we estimate what we label an *intra-area* DID estimate, which is represented as: *(B-A)-(D-C)*. This intra-area DID estimate is the same as the conventional DID strategy implemented in preceding hedonic studies (e.g., Davis, 2004; Linden and Rockoff, 2008; Haninger et al., 2017; Muehlenbachs et al., 2015).

IV.B. Hedonic Property Value Model

Our DID and triple differences identification strategies are implemented within a hedonic regression framework, where the dependent variable is the natural log of the price of home i, in neighborhood *j*, in housing market *m*, at time *t* (p_{ijmt}). More formally:

$$\text{(1)} \qquad ln(p_{ijmt}) = \boldsymbol{x}_{ijmt}\boldsymbol{\beta}_{mt} + \mathbb{1}(\boldsymbol{TSD}_{ijm} > 0)\boldsymbol{\theta} + \boldsymbol{TSD}_{ijm}\boldsymbol{\varphi}$$
$$+\boldsymbol{pre}_{ijmt}\boldsymbol{\gamma}^{pre} + \boldsymbol{mid}_{ijmt}\boldsymbol{\gamma}^{mid} + \boldsymbol{post}_{ijmt}\boldsymbol{\gamma}^{post} + \boldsymbol{\tau}_{mt} + \upsilon_{jm} + \varepsilon_{ijmt}$$

The right-hand side control variables include a vector of house and neighborhood characteristics $(\boldsymbol{x}_{ijmt})$, which includes the number of stories and bathrooms, the natural log of the lot size (in acres) and interior square footage, measures of the percent of the immediate area that is developed, and proximity to highways. As reflected by the subscript on the $\boldsymbol{\beta}_{mt}$ coefficient to be estimated, the hedonic price surface with respect to these attributes is allowed to vary by housing market (i.e., county) and year. In other words, we include *county* × *year* interaction terms with all elements of $\boldsymbol{x}_{ijmt}$. Although we are assuming a single nationwide hedonic price surface when estimating equation (1), these interaction terms allow the hedonic equilibrium surface to vary across markets and over time. Additionally, county-specific year and quarter fixed effects $(\boldsymbol{\tau}_{mt})$ are included to capture broader trends and seasonal effects in a particular market. Finally, υ_{jm} denotes spatial fixed effects at the census tract level to account for all time-invariant price factors associated with a specific location. ε_{ijmt} is a normally distributed disturbance term, which we allow to be correlated for all transactions within the same housing market.

The parameter estimates that are of direct interest are $\boldsymbol{\theta}$, $\boldsymbol{\varphi}$, $\boldsymbol{\gamma}^{pre}$, $\boldsymbol{\gamma}^{mid}$, and $\boldsymbol{\gamma}^{post}$. The vector $\boldsymbol{TSD}_{ijm}$ denotes the number of TSDFs within each 250-meter incremental bin, and $\mathbb{1}(\boldsymbol{TSD}_{ijm} > 0)$ is a corresponding indicator denoting the presence of at least one TSDF in each bin. Estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$ are used to estimate the price-distance gradient associated with proximity to a TSDF, irrespective of a CA. The variable $\boldsymbol{pre}_{ijmt}$ is a vector of indicator variables denoting the presence of a TSDF in each of the incremental distance bins around a home where a CA will occur, but has not yet occurred as of the time of the sale. Similarly, $\boldsymbol{mid}_{ijmt}$ denotes the presence of an open CA within each bin around a home, and $\boldsymbol{post}_{ijmt}$ denotes the presence of a completed CA in each bin. Therefore, $\boldsymbol{\gamma}^{pre}$, $\boldsymbol{\gamma}^{mid}$, and $\boldsymbol{\gamma}^{post}$ reflect the incremental difference in home values for a CA at each distance bin and in each respective CA stage, relative to a TSDF where a CA never occurs. These parameters by themselves reflect the first difference estimates described in section IV.A. In the most flexible model, 250-meter incremental bins are assumed, i.e., 0-250m, 250-500m, …, 4,750-5,000m. For the main regression models, as informed by initial data diagnostics (see section IV.D), broader bins are assumed for the control group (i.e., 750-1,500m).

Let *d* denote each respective distance bin in the treated group (e.g., 0-250m). The inter-area DID estimates of the percent change in price due to the opening of a CA is:

(2) $$DDTE_{[d]}^{1} = \exp\left(\gamma_{[d]}^{mid} - \gamma_{[d]}^{pre}\right) - 1$$

This inter-area DID estimate can be interpreted as the Average Treatment Effect on the Treated (ATT). It reflects the average percent change in price among homes located in distance bin *d* from a CA, relative to before that CA was initiated. The "1" superscript denotes that this is the first treatment event (i.e., the opening of a CA).

The subsequent effect of the second treatment event – completion of cleanup and resolution of the CA – is calculated as:

(3) $$DDTE_{[d]}^{2} = exp\left(\gamma_{[d]}^{post} - \gamma_{[d]}^{mid}\right) - 1$$

The vectors $\boldsymbol{pre}_{ijmt}$, $\boldsymbol{mid}_{ijmt}$, and $\boldsymbol{post}_{ijmt}$ also include indicator variables denoting the presence of a TSDF in each of the respective CA stages at farther control group distance bins (i.e., 750-1,500m). This is needed for our triple differences and intra-area DID estimates of the ATT. The triple differences treatment effect estimates of the opening and completion of a CA within distance bin *d* are:

(4) $$D3TE_{[d]}^{1} = \exp\left(\left[\gamma_{[d]}^{mid} - \gamma_{[d]}^{pre}\right] - \left[\gamma_{[ctrl]}^{mid} - \gamma_{[ctrl]}^{pre}\right]\right) - 1$$

(5) $$D3TE_{[d]}^{2} = \exp\left(\left[\gamma_{[d]}^{post} - \gamma_{[d]}^{mid}\right] - \left[\gamma_{[ctrl]}^{post} - \gamma_{[ctrl]}^{mid}\right]\right) - 1$$

where $\gamma_{[ctrl]}^{pre}$, $\gamma_{[ctrl]}^{mid}$, and $\gamma_{[ctrl]}^{post}$ are the coefficients corresponding to the presence of a CA in each respective stage in the farther control group distance bin.

In our more focused analysis of just homes within five kilometers of a CA, we drop all home transactions that are only near TSDFs where no CA occurs. However, due to the sometimes-clustered nature of TSDFs, we must still control for the presence and number of nearby TSDFs. Therefore, the intra-area DID estimates are also calculated following equations (4) and (5). Although the calculations are the same, dropping that initial control group of homes near a TSDF where no CA occurs diminishes a formal triple differences interpretation.

IV.C. Covariate and Coarsened Exact Matched Sample

To examine the robustness of our results, in later models we perform a pre-regression matching of the data that prunes and re-weights the sample so that the defined treatment and control groups have more balanced distributions in terms of the observed attributes. We use a combination of exact covariate matching and coarsened exact matching (CEM) techniques (Blackwell et al., 2009; Iacus et al., 2012). The CEM approach first divides the continuous attribute space for the relevant variables into discrete bins. Then home transactions that fall within the same set of discrete attribute bins (i.e., have the same value for *all* "coarsened" attributes) are matched. Homes are matched based on coarsened variables of the number of bathrooms, age, interior square footage, lot acreage, and the percent of the surrounding area within 500m that is developed.[14] In addition, and perhaps more importantly, we match homes in the treatment and control groups only if they are in the same county and sold in the same year. As a result, transactions in the control group are dropped from the matched sample if there are no comparable homes in the treated group in that county and sold in that same year, and vice versa.

In addition to the pruning of the estimating sample, CEM entails the assignment of weights to the treatment and control groups. A weight of one is given to all treated homes that could be matched to at least one control home. A treated home can potentially be matched to more than one control home, and one control home can potentially be matched to more than one treated home, and so all maintained control homes are given a positive weight that may be less than or greater than one. Despite its advantages, the CEM procedure has been used in only a few other hedonic property value studies (Groves and Rogers, 2011; Guignet et al., 2018; Qiu et al., 2017).

IV.D. Determining the Treatment Group

To determine the spatial extent of the treatment (i.e., the property value effects from the opening and completion of a CA), we adopt a strategy similar to the less parametric approach of using local polynomial regressions to graph the pre- and post-treatment distance gradients (Haninger et al., 2017; Muehlenbachs et al., 2015). We could not directly undertake the approach used by Haninger et al. (2017), Muehlenbachs et al. (2015), and others, due to the clustered nature of TSDFs and CAs. A home transaction could be near more than one CA, and it is possible that nearby CA sites could be in different stages of the investigation and cleanup process.

Using the sample of 2,512,354 homes within 5,000 meters of a CA, we estimate a hedonic price regression following equation (1) where $\boldsymbol{pre}_{ijmt}$, $\boldsymbol{mid}_{ijmt}$, and $\boldsymbol{post}_{ijmt}$ are defined using 250-meter incremental distance bins from 0-250 meters out to 4,500-4,750 meters (the farthest distance bin of 4,750-5,000 meters is the omitted category). This diagnostic regression includes all other variables in equation (1).

The estimates of $\boldsymbol{\gamma}^{pre}$, $\boldsymbol{\gamma}^{mid}$, and $\boldsymbol{\gamma}^{post}$ are plotted in Figure 8, thus displaying the price gradient with respect to distance to a CA in each respective stage. Comparing the pre- versus mid-CA

[14] The coarsened attribute values for bathrooms are one bathroom, 1-2 bathrooms, 2-3, bathrooms, and more than 3 bathrooms. Three values are allowed for the coarsened variables for age, interior square footage, acres, and percent of land developed within 500m. The cutoffs for these coarsened values correspond roughly to the 25th and 75th percentiles of the sample distributions. A fourth "missing" category is included for each of these coarsened variables, when applicable.

gradients shows that the homes nearest to the CA are lesser in value after the CA is opened. Comparison of the post-CA gradient suggests that homes nearest the CA experience a rebound in value back to the pre-CA levels once the CA is complete. Otherwise, the value of homes farther from the site do not vary significantly across the CA stages. The treated group is thus assumed to be all homes within 0-750 meters of a TSDF. The corresponding control group is assumed to be homes within 750-1,500 meters of a TSDF.

V. HEDONIC REGRESSION RESULTS

We estimate numerous variants of the hedonic price regression in equation (1). All models include census tract fixed effects, county-by-year and county-by-quarter fixed effects, and county-by-year interactions with all house and location attributes unrelated to TSDFs and CAs. Initially, to allow for spatial heterogeneity in the treatment effects at various distances, the CA stage vectors ($\boldsymbol{pre}_{ijt}$, $\boldsymbol{mid}_{ijt}$, and $\boldsymbol{post}_{ijt}$) include indicator variables denoting the presence of a CA in each respective stage within 0-250m, 250-500m, and 500-750m. Two aggregated bins are also included, denoting the presence of a pre-, mid-, or post-CA site within 750-1,500m and 1,500-5,000m. The former is used as the control group when calculating the intra-area DID and triple differences estimates.

The discussion here focuses on the estimated treatment effects expressed in equations (2) through (5), but the full hedonic regression results are presented in the Appendix (Tables A1 and A2). The control variables that are not of primary interest generally have the expected sign and magnitude, and are statistically significant across all models. For example, in early models that constrained $\boldsymbol{\beta}$ to be the same across markets and years, we found that house prices were increasing with parcel acreage, number of stories, interior square footage, the number of bathrooms, and the percent of the surrounding land area that is developed. Home values decrease with age and proximity to a major highway.

V.A. Price Gradient Associated with Proximity to a TSDF.

Following equation (1), the base hedonic model (model 1) is estimated with the over 9.6 million home transactions within five kilometers of at least one TSDF.

Before discussing how CAs and the related contamination and cleanup activities impact home values, we first examine the general association between house prices and proximity to a TSDF. Using the estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$, we can illustrate the price-distance gradient associated with proximity to a TSDF. More formally, this is calculated as:

$$\%\Delta p_{[d]}^{s} = \left\{\exp\left(\theta_{[d]} + \varphi_{[d]} s\right) - 1\right\} \times 100 \quad (6)$$

where $\theta_{[d]}$ and $\varphi_{[d]}$ are elements of the vectors $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$ that correspond to distance bin *d* (e.g., 0-250 meters), and *s* denotes the number of TSDF sites in distance bin *d*, which we vary in this illustrative exercise. $\%\Delta p^{s}_{[d]}$ is the percent difference in price associated with a home at a distance d from s TSDFs (e.g., s=*1, 2, 3*), holding all else constant. These estimates are relative to the omitted farthest distance bin (4,750-5,000 meters).

The price gradient results in Figure 9 clearly demonstrate a strong negative association between proximity to a TSDF and house prices. The top black line suggests that a single TSDF located within 0-250 meters of a home corresponds to an 8% decrease in value. This negative association diminishes with distance, but remains statistically significant. Although the use of indicator variables for each incremental 250-meter bin allows for a flexible distance gradient, the results actually align with expectations, suggesting a smooth price gradient that diminishes with distance.

Figure 9 also shows evidence of rather large cumulative effects, particularly in the nearest distance bins. The dark grey middle gradient in Figure 9 suggests that having two TSDFs within 0-250 meters is associated with a 19% depreciation in home values, and the presence of three is associated with a 29% decline.[15] These negative associations also diminish with distance, but the price decrements associated with proximity to multiple TSDFs are still greater than 5% until a distance of at least 2,000 meters, and remain statistically significant out to the full spatial extent examined (4,750 meters). Although we cannot argue these estimates as causal, it is clear homes near a TSDF already experience a significant discount in value, irrespective of any contamination issues.

V.B. Price Effects of Corrective Actions.

The remainder of the analysis aims to estimate the causal effects of a CA on these already value-depressed homes. The ATT estimates from model 1 are shown in Table 6. Both inter-area DID and triple differences estimates can be derived from the estimated coefficients from model 1. Following equations (2) through (5), we estimate the treatment effects for each 250-meter bin in the treated group separately, i.e., *d=0-250m*, *250-500m*, and *500-750m*. The control group bin for the triple differences estimates in equations (4) and (5) is 750-1,500 meters (see section IV.D).

Although we have less confidence in the inter-area DID estimates compared to the triple differences and later intra-area DID estimates, which utilize nearby homes as part of the control group, we still include them for comparison. For model 1 (Table 6), the DID and triple differences estimates are strikingly similar. Both suggest a 6% decline among homes in the nearest 0-250m distance bin. This effect is statistically insignificant, however, which could be partly attributed to the relatively small sample size in this nearest bin (see Figure 6). Although not statistically different, the point estimates decrease in magnitude at farther distances, suggesting a marginally

[15] We are extrapolating out of sample when inferring the price effect associated with three TSDFs in the 0-250m bin. Although we observe 18,762 and 226 sales within 0-250m of one and two TSDFs, respectively, no transactions are observed in such close proximity to three TSDFs. Such cases are observed, however, in the farther distance bins, with 14, 102, and 268 sales within 250-500m, 500-750m, and 750-1000m of three TSDFs, for example.

significant 5.6% decline among homes in the 250-500m bin, and then a significant 4.6% decline in the 500-750m bin.

According to the inter-area DID estimates from model 1 (Table 6), cleanup and the completion of the CA yield a marginally significant 5.4% appreciation, but only in 500-750m. These estimated positive effects become stronger in magnitude and significance with the triple differences estimates, which better account for unobserved local trends that could otherwise confound the results. Interestingly, both sets of estimates suggest small and insignificant effects of CA completion on homes in the nearest 0-250m bin. These homes are extremely close to these hazardous waste sites, and so it is not necessarily surprising that we see little change in home values after cleanup is complete. Perceived risks may not change much because residents are extremely close to these sites. Buyers may still perceive risks, and these nearest homes may be permanently stigmatized despite any cleanup activities.[16]

Model 2 in Table 6 is similar to the previous model, but is estimated using only transactions of homes that are within five kilometers of a TSDF where a CA has occurred as of the time of the sale, or will occur later in the study period. Focusing on this subset of homes allows us to derive intra-area DID estimates that are similar to the DID approaches in other hedonic studies (e.g., Linden and Rockoff, 2008; Haninger et al., 2017; Muehlenbachs et al., 2015). The TSDF sites where CAs occur, and the homes and neighborhoods around these sites, may be systematically different from those around TSDFs where a CA does not occur. Focusing on this more homogenous sample in model 2 facilitates a cleaner quasi-experimental comparison. The results are quite similar to the previous model, suggesting a 4.5% decline in value when a CA is opened, and a subsequent 6.1-7.9% rebound in price once the CA is completed. The results for the nearest bins, however, are again insignificant or only marginally significant at best.

A series of Wald tests suggest that the estimates across the 250-meter treatment bins are not statistically different.[17] Therefore, in Table 7 we re-estimate the previous two models but pool the treatment bins into a single 0-750m bin, thus constraining the estimated treatment effects to be the same for all homes within 750m. The DID and triple differences estimates in both models 3 and 4 suggest a statistically significant 4.6-5.0% decrease in value, on average, when a CA is initiated at a TSDF within 750m. When the contamination issue is resolved and the CA completed, there is an average appreciation of about 5.3-7.2% among homes within 750m.

To provide a cleaner quasi-experiment, we next estimate the models using only homes in the treatment (0-750m) or control (750-1500m) distance bins. In doing so, model 5 in Table 7 suggests a smaller 2.5% depreciation from the opening of a CA, and the effect is now insignificant. The estimated 6.5% increase in value after completion of a CA remains robust.

---

[16] See Messer et al. (2006) for a discussion of stigma and property value impacts.

[17] For the opening of a CA, we fail to reject the null hypothesis that $D3TE^1_{[0-250]} = D3TE^1_{[250-500]} = D3TE^1_{[500-750]}$ for model 1 ($\chi^2(2) = 0.26, p = 0.8793$) and model 2 ($\chi^2(2) = 0.05, p = 0.9737$). Similarly, for the closing of a CA we fail to reject the null hypothesis that $D3TE^2_{[0-250]} = D3TE^2_{[250-500]} = D3TE^2_{[500-750]}$ for model 1 ($\chi^2(2) = 0.64, p = 0.7274$) and model 2 ($\chi^2(2) = 0.31, p = 0.8547$).

Finally, to further examine the robustness of our results, we parse and re-weight this 0-1500m sample using exact covariate matching and CEM techniques (Blackwell et al., 2009; Iacus et al., 2012). The matching procedure substantially reduces the sample size, but also results in more comparable treatment and control groups.[18] Model 6 in Table 7 shows that the opening of a CA is again statistically insignificant. However, the average price impact from the cleanup and resolution of the contamination issue remains robust, suggesting a slightly larger 7.5% increase in the value of homes within 750m when the CA is completed.

V.C. Assessing a Causal Interpretation.

A causal interpretation of the estimated treatment effects from the initiation and completion of a CA depends on the validity of the assumed counterfactual. To assess the appropriateness of a causal interpretation we first compare the mean values of observed covariates across the treated and control groups (i.e., homes within 0-750m and 750-1,500m of a CA, respectively). We do this for the full sample of homes in these groups, and then for the CEM-weighted sample. Second, we compare the univariate distributions of observed characteristics using the *L1* statistics (Blackwell et al., 2009; Iacus et al., 2012). Third, we evaluate the validity of the common trends assumption (Angrist and Pischke, 2009) by first conducting parametric regression-based tests, and then by examining the pre-treatment trends using non-parametric and semi-parametric local polynomial regression techniques.

A series of two-sample t-tests comparing the full, unweighted samples of transactions within 0-750m and 750-1,500m of a CA reveals statistically significant differences in the mean values for all covariates (see Table A3 in the Appendix). However, these sample sizes are fairly large, resulting in extremely small standard errors. The differences in means are not practically or economically significant in most cases. For example, the average acres are 0.245 versus 0.236 acres, the average number of bathrooms are 1.74 versus 1.72, and the difference in interior square footage is only 51 square feet. The differences in means of the location variables are slightly starker; for example, 50% of the land area within a 500m buffer of a home in the control group is developed, on average, compared to just 43% for the average treated home. Similarly, 47% of the control group is within 500 meters of a major highway, compared to 58% of the treated group. The average number of TSDFs in close proximity are quite similar across the two groups (1.21 versus 1.23 TSDFs within 1,500m), but again, this difference is statistically significant.

As expected, the mean values among the CEM weighted sample are more comparable across the treated and control groups, with statistically significant differences ($p \leq 0.10$) now only among six of the ten covariates. And even when there are statistically significant differences, there is again little practical difference in these mean values (see Table A3 in the Appendix). The percent difference across all variable means is 3% or less, with the exception of proximity to a highway

[18] See section V.C. and Tables A3-A5 in the appendix for details.

(51% of the control group is within 500m of a highway, compared to 58% of the treated group – a 14% difference).

Comparison of the multivariate and univariate *L1* statistics (Blackwell et al., 2009; Iacus et al., 2012) shows that the univariate and multivariate distributions across the treated and control groups are more similar after matching (see Tables A4 and A5 in the Appendix). This is true even among most of the covariates that were not included in the CEM algorithm.[19] It is worth emphasizing that we restrict matches to only homes in the same county *and* that sold in the same year (in addition to matching on the coarsened variables discussed in section IV.C). Doing so draws focus to CA sites where both treated and control transactions are observed, and in general helps mitigate biases associated with a particular county and year by ensuring an equal balance of treated and control observations in each county-year combination.

Finally, in a well-defined quasi-experiment the trajectory of the outcome variable experienced by the treated group in the absence of treatment (a state-of-the-world that is unobserved) must be the same as the outcome trajectory experienced among the assumed control group in the post-treatment period. This is commonly referred to as the common or parallel trends assumption (Angrist and Pischke, 2009). Given that we do not observe the true counterfactual (i.e., the treated group absent the treatment), researchers often turn to the pretreatment trends. If the outcome of interest (i.e., house prices) for the treated and control groups follow similar trajectories in the pretreatment period (i.e., before the opening or completion of a CA), then it seems more reasonable to assume those trajectories would have remained similar in the absence of treatment.

We estimate a series of regressions using only the subsamples of transactions that are in the treated or control group, and that occurred prior to the respective treatment events. A trend variable *years before* is included to capture the linear trend in home prices leading up to the treatment event, holding all else constant. Time is normalized across sites so that $t=0$ denotes the day of the treatment event (i.e., the first CA being opened or the first CA being completed). Column (1) in Table 8 shows a positive but insignificant trend in house prices leading up to the opening of a CA. Most importantly, these models include an interaction term between the trend variable and an indicator denoting the treatment group (homes within 0-750m). The corresponding coefficient captures the incremental difference in the linear trend leading up to the opening of a CA. The small and insignificant 0.0024 coefficient corresponding to the interaction term *years before × 0-750m* suggests that the trends among the treated and control groups are parallel. A similar result is found in column (2), which is based on the CEM sample of pre-CA homes. Columns (3) and (4) focus on the pre-treatment trends relative to the completion of a CA, and are estimated using the full and CEM samples of transactions observed within 1,500m and prior to completion. The small and insignificant -0.0022 and 0.0018 coefficients on *years before × 0-750m* suggest the price trends leading up to completion of a CA are parallel.

A critical assumption with these tests is that the conditional pre-treatment trends are linear, which may not necessarily be the case. As has been done in several other hedonic studies (Haninger et

[19] The two exceptions are that there was a slight increase in the univariate *L1* statistics for housing age and the number of TSDFs within 0-5,000m.

al., 2017; Linden and Rockoff, 2008; Muehlenbachs et al., 2015), we next examine the pre-treatment trends using less parametric local polynomial regression techniques. The house price trends for the treated (0-750m of a CA) and control (750-1,500m of a CA) groups are presented in Figures 10 and 11 for the opening and completion of a CA, respectively. We first do this non-parametrically by simply graphing the real transaction prices (2019$) over time. Of course, many factors can affect house prices over time, and so a semi-parametric variant is also estimated where we first estimate a regression similar to equation (1), but that excludes: $\boldsymbol{pre}_{ijt}$, $\boldsymbol{mid}_{ijt}$, and $\boldsymbol{post}_{ijt}$. This regression is estimated using the full sample of transactions within 5 km of a TSDF. The predicted residuals $\left(\hat{\varepsilon}_{ijmt}\right)$ for home sales in the treated and control groups are then graphed. $\hat{\varepsilon}_{ijmt}$ reflects a cleaner price signal where otherwise confounding factors have been conditioned out. The local polynomial regressions of the prices and residuals are graphed separately for the opening and completion of a CA.

In Figures 10 and 11, the top panel shows the price trends and the bottom panel shows the residual trends. The left column is for the unweighted sample of all home sales within 1,500 meters of a CA, and the right column corresponds to the CEM-weighted sample. The focus for assessing the appropriateness of the common trends assumption is on the pretreatment trends (i.e., the portion of the graphs to the left of zero on the x-axis), but the post-treatment trends in event-study graphs like these can also be insightful.

Figure 10 shows the trend graphs relative to the opening of a CA. The top-left panel shows the raw prices with no assumed regression or matching taking place. Visual inspection suggests that the pre-treatment trajectories follow similar broader paths, but may not necessarily be parallel, especially when considering 5,000 or more days (almost 14 years) prior to the CA opening. The CEM sample trends in the top-right of Figure 10 also show similar broader trajectories, but the curves are not parallel, and even cross prior to the opening of the CA. The price graphs are useful for examining broader trends, but given the relatively small magnitude of the estimated price effects from CAs, the residual graphs are needed for a more focused examination. Although the bottom graphs suggest a clear price decline a few years after the opening of a CA, the pre-treatment trends are not parallel. There is a clear divergence in trends when looking at more than 4,000 days (about 11 years) before the CA opens. Limiting the temporal window around the treatment events to a period where the pre-Corrective Action trends are more parallel results in a statistically insignificant effect.[20] Overall, visual inspection of the pre-treatment trends brings into question the validity of a causal interpretation of our estimated price effects from the opening of a CA. And in cases where causal inference may be more appropriate (e.g., when using the CEM matched sample, and/or focusing on a tighter temporal window) we find statistically insignificant effects.

The price trends relative to the completion of a CA are displayed in Figure 11. The top row shows that prices in the treatment and control groups followed similar trends prior to the completion of a

[20] Models 5 and 6 were re-estimated with the subset of home transactions that took place within 10 years before or after the opening or closing of a CA. The results suggest that the opening of a CA corresponded to an insignificant 0.9% and 2.2% decrease in home values, respectively. The post-closure appreciation among homes within 0-750m remained robust, suggesting a significant 5.7% and 8.6% increase.

CA, but that the trends are not perfectly parallel. Particularly striking is the large divergence in the price trajectories after the completion of the CA (top-left panel). The graphs of the predicted residuals in the bottom row of Figure 11 provide a cleaner look at the price trends because otherwise confounding factors are conditioned out. The pre-treatment trends here do appear fairly parallel. Based on these diagnostics, a causal interpretation of the estimated 5-7% average appreciation after the completion of a CA seems valid.

VI. WELFARE IMPLICATIONS

Considering the entire contiguous US, our data suggests there are 25,415 single-family homes within 750 meters of one of the 195 TSDFs where the CA was fully complete by 2018. The mean transaction price for the corresponding 620 sales of these homes in 2018 is $207,307. We calculate the mean capitalization effects of cleanup as: $207{,}307\left(\frac{D3TE^2_{[0-750]}}{1+D3TE^2_{[0-750]}}\right)$, where $D3TE^2_{[0-750]}$ is the percent change in price expressed by equation (5). Based on the 6.52% appreciation estimated from model 4, this translates to a mean capitalization effect of $12,691 per home (with a 95% confidence interval of $3,843 to $21,540). We select model 4 for this illustration because it is the most conservative estimate among the more credible triple differences and intra-area DID estimates. Multiplying the mean capitalization effect by the 25,415 single-family homes suggests that the completion of cleanup at these 195 TSDFs increased the total housing stock value by $323 million (with a 95% confidence interval of $97 million to $547 million).

Estimating the capitalization effects of hazardous waste cleanup, as done above, is empirically appealing, but these estimates generally lack a formal welfare interpretation if the hedonic price gradients with respect to the variables of interest are changing over time (Klaiber and Smith, 2013; Kuminoff and Pope, 2014). A Wald test based on the regression model described in equation (7) below confirms that is the case with our application and data. We reject the null hypothesis that the coefficients $\gamma^{mid}_{[0-750]}$ and $\gamma^{post}_{[0-750]}$ are constant from 2000-2018 ($F(36, 376)=2.04$, $p=0.0006$).

Following Banzhaf et al. (Forthcoming), we estimate a variant of equation (1) that allows us to derive a formally valid, lower-bound ex post estimate of the benefits of cleanup completion. Doing so requires that the entire hedonic surface, including the slopes with respect to the TSDF and CA variables of interest, be allowed to vary over time. Banzhaf demonstrates that a change in price along the same ex post price gradient is a lower bound of the Hicksian equivalent surplus for an improvement in quality.

We estimate the following variant of equation (1), where interaction terms with year indicators are included to allow the CA coefficient vectors $\boldsymbol{\gamma}^{pre}_t$, $\boldsymbol{\gamma}^{mid}_t$, and $\boldsymbol{\gamma}^{post}_t$ to vary freely from year to year:

(7) $$ln(p_{ijmt}) = \boldsymbol{x}_{ijmt}\boldsymbol{\beta}_{mt} + \mathbb{1}(\boldsymbol{TSD}_{ijm} > 0)\boldsymbol{\theta}_m + \mathbb{1}(\boldsymbol{TSD}_{ijm} > 0)\boldsymbol{\delta}_t + \boldsymbol{TSD}_{ijm}\boldsymbol{\varphi}_m$$
$$+\boldsymbol{TSD}_{ijm}\boldsymbol{\alpha}_t + \boldsymbol{pre}_{ijmt}\boldsymbol{\gamma}_t^{pre} + \boldsymbol{mid}_{ijmt}\boldsymbol{\gamma}_t^{mid} + \boldsymbol{post}_{ijmt}\boldsymbol{\gamma}_t^{post}$$
$$+\boldsymbol{\tau}_{mt} + \upsilon_{jm} + \varepsilon_{ijmt}$$

An additional complication in a nationwide context such as ours is that the price surface must not only be allowed to vary temporally, but ideally it would also be allowed to vary spatially across housing markets. From a theoretical standpoint, the ideal model for our nationwide analysis would, for example, include county-by-year interactions with the TSD and CA variables. However, given the highly localized nature of this disamenity, the resulting distance-by-county-by-year bins have very few and often zero identifying observations. We attempted to estimate models that allow for *separate* interactions by state and by year with $\boldsymbol{pre}_{ijmt}$, $\boldsymbol{mid}_{ijmt}$, and $\boldsymbol{post}_{ijmt}$, but the models could not be estimated due to multicollinearity. The hedonic model proposed in equation (7) at least allows for heterogeneity in the baseline effects of TSDFs across spatially-defined markets. The coefficients corresponding to the presence and number of TSDFs in each 250-meter bin are allowed to vary separately by county ($\boldsymbol{\theta}_m, \boldsymbol{\varphi}_m$) and over time ($\boldsymbol{\delta}_t, \boldsymbol{\alpha}_t$). Given the empirical constraints in this particular context, equation (7) is the best we could do in applying Banzhaf's (Forthcoming) framework to a national setting. The most necessary feature for welfare calculations is maintained – the price gradients with respect to CAs are allowed to vary over time.

The coefficients from equation (7) are then used to estimate treatment effects specific to the ex post hedonic price surface in year $\tilde{t}$. These estimates can be interpreted as a direct unmediated effect (DUE), expressed as the percentage change in price.

(8) $$DUE_{[d]\tilde{t}} = \exp\left\{\left[\left(\gamma_{[d]\tilde{t}}^{post}\right) - \left(\gamma_{[d]\tilde{t}}^{mid}\right)\right] - \left[\left(\gamma_{[ctrl]\tilde{t}}^{post}\right) - \left(\gamma_{[ctrl]\tilde{t}}^{mid}\right)\right]\right\} - 1$$

$DUE_{[d]\tilde{t}}$ is an unmediated effect because all other attributes are held constant, and it is a direct effect because it only considers movement on the *same* price surface (Banzhaf, Forthcoming).

We set out to estimate a formal lower bound of the benefits of cleanup and completion of a CA. For a chosen ex post year $\tilde{t}$, we calculate the direct unmediated effect on the treated (DUET) as:

(9) $$\overline{DUET}_{[d]\tilde{t}} = \bar{p}_{[d]\tilde{t}}^{post}\left(\frac{DUE_{[d]\tilde{t}}}{1+DUE_{[d]\tilde{t}}}\right)$$

where $\bar{p}_{[d]\tilde{t}}^{post}$ denotes the ex post mean price among home sales observed in the treated group in year $\tilde{t}$. The total DUET (TDUET) is then calculated by multiplying $\overline{DUET}_{[d]\tilde{t}}$ by the total number of impacted single-family homes $(N_{[d]\tilde{t}})$:

(10) $$TDUET_{[d]\tilde{t}} = \overline{DUET}_{[d]\tilde{t}} \times N_{[d]\tilde{t}}$$

A key question in some policy settings, including that here where CAs and cleanups are completed at different times at different sites, is what is the appropriate ex post hedonic price gradient to use for calculating changes in welfare. We find the results are quite sensitive to what year is assumed for $\tilde{t}$.

To illustrate this, consider the 123 TSDFS (across 30 different states) where the CA and cleanup were completed between 2000 and 2015. We chose this time period because we wanted to ensure that we had a few years of ex post data to examine the sensitivity of the results, and because we wanted to account for the noticeable spike in the number of CA completions in 2015 (see Figure 2). There are $N_{[0-750]\tilde{t}}$ =18,179 single-family residences within 750 meters of one of the 123 remediated TSDFs. Since this policy illustration only considers cleanup completions through 2015, and we use the 2015 ex post price surface (at least initially) in our welfare calculations, we base our calculations on the mean price of $\bar{p}^{post}_{[0-750]\tilde{t}}$=\$165,780 among the post-CA homes that are within 0-750m and sold in 2015. We also judge this to be a conservative assumption because it is the lowest mean price observed across all ex post years in our data.[21]

First for comparison, we calculate the corresponding capitalization effects based on model 4, and apply the estimated 6.52% appreciation to the mean price and total number of impacted single-family residences. As shown in Table 9, this yields a mean price increase of \$10,149 and a total capitalization effect of \$184.5 million.

Following equation (7), a variant of model 4 is estimated so that we can estimate the direct effects of the policy on a fixed hedonic price surface. The next four columns under model 4′ in Table 9 show the results depending on which ex post price gradient is used, $\tilde{t}$ = 2015, 2016, 2017, or 2018. For this purposefully chosen policy illustration of CA completions through 2015, the 2015 price gradient seems like the most appropriate to use for an ex post welfare calculation. In this case we see the *DUE* of CA completion is an 8.58% increase in price along the fixed 2015 price gradient. Applying this to the mean post-CA price in 2015 suggests an average DUET of \$13,105. This is a lower bound estimate of how much the average household benefited from cleanup and completion of a CA. Multiplying this by the total number of single-family homes within 750m of a completed CA as per equation (10), suggests a lower bound, ex post total benefit of \$238.2 million to all impacted households.

Now suppose we assumed a different year for the ex post price gradient. As can be seen in Table 9, the DUE, $\overline{DUET}$, and *TDUET* estimates vary substantially depending on the assumed ex post gradient. Holding the number of impacted parcels and mean price constant, we see statistically insignificant estimates based on the 2016 and 2018 gradients, but actually see significant and larger effects based on the 2017 price gradient, suggesting a $\overline{DUET}$ of \$16,902 and a *TDUET* of \$307.3 million. The estimated total capitalization effects from the original model 4, where the price effects of a CA are assumed to be time invariant, falls squarely in between the TDUET estimates across the different ex post price gradients that are assumed.

Two key points come out of this welfare exercise. First, at least in this context, the lower bound benefit estimates are very sensitive to the assumed ex post price gradient. Second, in contrast to the empirical applications of Banzhaf (Forthcoming) and Kuminoff and Pope (2014), we find that

[21] The mean price (and number) of post-CA transactions within 0-750m in each ex post year are \$165,750 in 2015 (n=282), \$179,320 in 2016 (n=453), \$219,526 (n=512) in 2017, and \$207,307 in 2018 (n=620).

the estimated capitalization effects can be lower or higher than the estimated welfare effects, depending on which ex post price gradient is assumed for welfare estimation.

We believe the instability in the price gradients with respect to CA completion is likely more an artifact of the empirical analysis and data limitations, rather than an actual shift in the true price gradient from year to year. Many parameters are being estimated in these models, and there are relatively few identifying observations in close proximity to TSDFs in each CA stage, especially when considering these observations separately by year. Figure 12 shows the price effect estimates from year to year, and although they jump up and down, the longer-term trend is relatively flat. A trend line is fitted by re-estimating the model and constraining $\boldsymbol{\gamma}_t^{pre}$, $\boldsymbol{\gamma}_t^{mid}$, and $\boldsymbol{\gamma}_t^{post}$ to vary linearly over time. The linear trend coefficients are jointly insignificant ($F(6, 376) = 1.36$, $p = 0.2303$), suggesting that although the estimates jump up and down from year to year, in the long-run the price effects of CA completion have not changed.

In order to still apply Banzhaf's (Forthcoming) framework and focus on the ex post price surface, but at the same time account for data constraints and the instability of the *estimated* ex post price gradients from year to year, we assume $\boldsymbol{\gamma}_t^{mid}$, and $\boldsymbol{\gamma}_t^{post}$ are constant from 2015-2018. Doing so essentially yields an average over the available ex post gradients to smooth out any instabilities. A Wald test based on the estimates from model 4′ supports this assumption – we fail to reject the null hypothesis that $\boldsymbol{\gamma}_t^{mid}$ and $\boldsymbol{\gamma}_t^{post}$ are constant from 2015-2018 ($F(6, 376) = 1.10$, $p = 0.3616$).

Model 4″ in Table 9 is similar to model 4′, but groups the $\boldsymbol{\gamma}_t^{mid}$ and $\boldsymbol{\gamma}_t^{post}$ from 2015-2018 into a single temporal bin. Doing so yields a marginally significant *DUE* of 5.33%, and the average DUET is $8,395 per household. Our preferred lower bound estimate of the ex post total benefits to nearby residents from the completion of cleanup at the 123 TSDFs from 2000-2015 is $152.6 million.

VII. CONCLUSION

There are 2,389 treatment, storage, and disposal facilities (TSDFs) of hazardous waste in the contiguous US, and over 487 open Corrective Actions with ongoing investigation and incomplete cleanup activities at these sites. Considering RCRA sites more broadly, the US EPA (2013) estimates that over 35 million people (or 12% of the US population) live within one mile of a Corrective Action (CA) site. Releases and improper disposal and treatment of hazardous waste clearly remain a threat to local residents. At the same time, monetized benefits to nearby residents have generally not been accounted for in benefit-cost analyses informing federal regulations under the RCRA. To fill this gap, we framed a series of quasi-experimental comparisons using a novel dataset that combines nationwide data on housing transactions and TSDFs. We find that CAs do affect home values up to 750-meters away. This raises potential equity concerns because home values in communities around TSDFs are already significantly depressed, irrespective of any contamination or cleanup events.

We find that the discovery of contamination and opening of a CA investigation may lead to a price decline of up to 5% among homes within 0-750 meters. This result is not robust across all models, however. In addition, some differences in the pre-treatment trends suggest that causal inference of the average price effects from the opening of a CA may not be appropriate. These findings are not necessarily surprising. The opening of a CA investigation may not signal new information that changes local residents' and potential homebuyers' risk perceptions. Any perceived risks may already be capitalized in surrounding home values due to visual cues associated with proximity to a TSDF in general. In fact, often a new CA involves an investigation of historical contamination issues (US EPA, 2013), and not necessarily a new release.

We do find robust, causal evidence that the completion of a CA leads to a 5% to 7.5% increase in value among homes within 0-750m. The completion of a CA signals resolution of the contamination issue, either through the completion of active cleanup efforts, placement of physical barriers to prevent exposure and pollutant migration, and/or future land use restrictions to minimize exposure. These results are robust to a variety of approaches to control for spatially and temporally correlated factors that could otherwise confound the analysis, including: implementation of difference-in-differences (DID) and triple differences methods, the use of census tract-fixed effects, allowing for county-specific year and quarter fixed effects, and through the use of exact covariate and coarsened exact matching (CEM) techniques. Examination of the pre-treatment trends further supports causal inference.

The estimated 5% to 7.5% post-cleanup appreciation among homes within 750m is in line with previous studies of similar disamenities. Hedonic property value studies of other cleanup programs under EPA's Office of Land and Emergency Management generally suggest that cleanup leads to a 5-20% increase in surrounding home values (Gamper-Rabindran and Timmins, 2013; Guignet et al., 2018; Haninger et al., 2014); although those studies find farther extending effects, out to 2-4 kilometers. Our results are also in agreement with Cassidy et al.'s (2020) working paper that uses Census tract-level data to examine the impact of RCRA site cleanups through the CA program. They find that cleanup leads to a roughly 7-11% appreciation among homes in the census tract where the RCRA site is located. They find these price effects to be the strongest among homes in the lowest percentiles of the tract-specific price distributions, which as suggested by our study, likely corresponds to the homes nearest the TSDFs.

Our treatment effect estimates can be used to better inform policy. Our preferred model suggests a 6.5% appreciation upon completion of a CA, which translates to a mean capitalization effect of \$12,691 per home. Considering the 25,415 single-family homes with 750m of one of 195 TSDFs where a CA was complete by 2018 (the end of our study period), this suggests that cleanup of these sites led to a total increase in housing stock value of \$323 million.

Following a similar exercise, we apply an approach proposed by Banzhaf (Forthcoming) to estimate a lower bound of the ex post benefits of cleanup completion, which we find to be about \$8,400 per household. The estimated benefit of cleanup at a single site will vary greatly depending on the number of nearby residents, among other things, but as a back-of-the-envelope comparison, this suggests an average lower bound benefit of \$1.24 million for each TSDF that was cleaned up. The costs of cleanup can also vary substantially from site to site, but in general would tend to outweigh our estimated average lower bound of the benefits. Case studies described by EPA (2013) suggest costs ranging from \$2.7 million to \$1.9 billion (2019\$). Tonn et al. (1991) simulated remediation costs across 2,000 different scenarios and found a base case average cost of \$12.0 million per solid waste management unit.[22] Of course, our estimated benefits to nearby residents are just one part of the benefit calculus, and there are other endpoints and impacted populations to consider.[23]

Our estimated price and welfare impacts reflect the average effects of the opening and completion of a CA. However, TSDFs are truly local disamenities and there is likely significant variation in the impacts around any one site. TSDFs can differ in baseline characteristics, community awareness, and perceived risks. The CA events can also vary in terms of contamination severity, presence of exposure pathways, and the extent of subsequent cleanup activities. Examination of such heterogeneity is a fruitful direction for future research, and could yield important implications for benefit-transfer and policy analysis.

Hedonic property value methods compose an increasingly large portion of the nonmarket valuation literature, but still are seldom used in policy analysis (Petrolia et al., Forthcoming). One deterrent is the lack of a formal welfare interpretation in many contexts. Further research and guidance are needed to solidify a welfare interpretation of estimates produced from hedonic pricing methods. Theoretical work building on that of Banzhaf (Forthcoming, 2020), Kuminoff and Pope (2014), and others to provide both an upper and lower bound of the welfare impacts, and ideally for both ex post and ex ante settings, will help further the application of hedonic pricing methods to inform policy decisions. Nonetheless, our average capitalization effects and lower bound benefit estimates serve as a useful intermediate step to inform policy on the cleanup of hazardous waste sites.

[22] Cleanup cost estimates adjusted to 2019\$ using the Bureau of Labor Statistics's annual average consumer price index for all urban consumers (https://www.bls.gov/cpi/tables/supplemental-files/historical-cpi-u-202105.pdf, accessed 14 June 2021).
[23] A brief, high-level review of benefit categories is provided by US EPA (2013). For more detailed discussions see the Regulatory Impact Analyses cited in Table 1.

WORKS CITED

Aizer, Anna, Janet Currie, Peter Simon, and Patrick Vivier. 2018. Do low levels of blood lead reduce children's future test scores? *American Economic Journal: Applied Economics* 10 (1): 307–341.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

Atreya, Ajita, Susana Ferreira, and Warren Kriesel. 2013. Forgetting the Flood? An Analysis of the Flood Risk Discount over Time. *Land Economics* 89(4): 577-596.

Banzhaf, H. Spencer. 2020. Panel data hedonics: Rosen's first stage as a "sufficient statistic." *International Economic Review* 61(2): 973-1000.

Banzhaf, H. Spencer. Forthcoming. Difference-in-Differences Hedonics. *Journal of Political Economy*: https://doi.org/10.1086/714442.

Bin, Okmyung, and Craig Landry. 2013. Changes in Implicit Flood Risk Premiums: Empirical Evidence from the Housing Market. *Journal of Environmental Economics and Management* 65: 361-376.

Blackwell, Matthew, Stefano Iacus, Gary King, and Guiseppe Porro. 2009. Cem: Coarsened exact matching in Stata. *The Stata Journal 9*(4): 524-546.

Boyle, Melissa A., and Katherine A. Kiel. 2001. A Survey of House Price Hedonic Studies of the Impact of Environmental Externalitie. *Journal of Real Estate Literature* 9(2): 117-144.

Cassidy, Alecia, Elaine L. Hill, and Lala Ma. 2020. Housing Impacts of Hazardous Waste Cleanup. SSRN Working Paper #3766930. Available at: https://privpapers.ssrn.com/sol3/papers.cfm?abstract_id=3766930.

Davis, Lucas W. 2004. The Effect of Health Risk on Housing Values: Evidence from a Cancer Cluster. *American Economic Review*: 94(5): 1693-1704.

Dewitz, J., 2019. National Land Cover Database (NLCD) 2016 Products. U.S. Geological Survey data release,. https://doi.org/10.5066/P96HHBIE.

Farber, Stephen. 1998. Undesirable facilities and property values: a summary of empirical studies. *Ecological Economics* 24: 1-14.

Gamper-Rabindran, Shanti, and Christopher 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.

Greenstone, Michael. and Ted Gayer. 2009. Quasi-experimental and experimental approaches to environmental economics. *Journal of Environmental Economics and Management.* 57: 21-44.

Groves, Jeremy R., and William H. Rogers. 2011. Effectiveness of RCA Institutions to Limit Local Externalities: Using Foreclosure Data to Test Covenant Effectiveness. *Land Economics 87*(4): 559-581.

Guignet, Dennis, Charles Griffiths, Heather Klemick, and Patrick J. Walsh. 2017. The Implicit Price of Aquatic Grasses. *Marine Resource Economics* 32(1): 21-41.

Guignet, Dennis, Robin Jenkins, Matthew Ranson, and Patrick J. Walsh. 2018. Contamination and Incomplete Information: Bounding Implicit Prices using High-Profile Leaks. *Journal of Environmental Economics and Management 88*: 259-282.

Guignet, Dennis, and Jonathan Lee. 2021. State of the Art of Hedonic Pricing. In *Oxford Research Encyclopedia of Environmental Science*, ed. James R. Kahn. Oxford University Press. doi: https://doi.org/10.1093/acrefore/9780199389414.013.555.

Haninger, Kevin, Lala Ma, and Chrisopher Timmins. 2017. The Value of Brownfield Remediation. *Journal of the Association of Environmental and Resource Economists 4*(1): 197-241.

Horsch, Eric J., and David J. Lewis. 2009. The Effects of Aquatic Invasive Species on Property Values: Evidence from a Quasi-Experiment. *Land Economics* 85(3): 391-409.

Iacus, Stefano, Gary King, and Giuseppe Porro. 2012. Causal Inference without Balance Checking: Coarsened Exact Matching. *Political Analysis 20*: 1-24.

Jackson, Thomas. 2001. The Effects of Environmental Contamination on Real Estate: A Literature Review. *Journal of Real Estate Literature* 9(2): 93-116.

Kiel, Katherine A., and Jeffrey E. Zabel. 1999. The Accuracy of Owner-Provided House Values: The 1978-1991 American Housing Survey. *Real Estate Economics* 27(2): 263-298.

Klaiber, H. Allen, and V. Kerry Smith. 2013. Quasi-Experiments, Hedonic Models, and Estimating Trade-offs for Local Amenities. *Land Economics* 89(3): 413-431.

Kohlhase, Janet E. 1991. The Impact of Toxic Waste Sites on Housing Values. *Journal of Urban Economics* 30: 1-26.

Kuminoff, Nicolai V., and Jaren C. Pope. 2014. Do ‘Capitalization Effects’ for Public Goods Reveal the Public’s Willingness to Pay. *International Economic Review* 55(4): 1227-1250.

Lanphear, Bruce P., Richard Hornung, Jane Khoury, Kimberly Yolton, Peter Baghurst, David C. Bellinger, et al., 2005. Low-Level environmental lead exposure and children's intellectual function: an international pooled analysis. *Environmental Health Perspectives* 113 (7): 894–899.

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

Manson, Steven, Jonathan Schroeder, David Van Riper, and Steven Ruggles. 2018. IPUMS National Historical Geographical Information System: Version 13.0 [Database]. Minneapolis. https://doi.org/10.18128/D050.V13.0

Messer, Kent D., William D. Schulze, Katherine F. Hackett, Trudy A. Cameron, and Gary H. McClelland. 2006. Can Stigma Explain Large Property Value Losses? The Psychology and Economics of Superfund. *Environmental and Resource Economics* 33: 299-324.

Michaels, R. Gregory and V. Kerry Smith. 1990. Market Segmentation and Valuing Amenities with Hedonic Models: The Case of Hazardous Waste Sites. *Journal of Urban Economics* 28: 223-242.

Miranda, Marie Lynn, Dohyeong Kim, M. Alicia Overstreet Galeano, Christopher J. Paul, Andrew P. Hull, and S. Philip Morgan. 2007. The relationship between early childhood blood lead levels and performance on end-of-grade tests. *Environmental Health Perspectives* 115(8): 1242–1247.

Muehlenbachs, Lucija, Elisheba Spiller, and Christopher Timmins. (2015). The Housing Market Impacts of Shale Gas Development. *American Economic Review 105*(12): 3633-3659.

Nolte, Christoph. 2020. High-resolution land value maps reveal underestination of conservation costs in the United States. *Proceedings of the National Academny of Sciences* 117: 29577-29583.

Parmeter, Christopher F. and Jaren C. Pope. 2013. Quasi-experiments and Hedonic Property Value Methods. In *Handbook on Experimental Economics and the Environment*, eds. John A. List and Michael K. Price. Cheltenham, United Kingdom: Edward Elgar Publishing.

Petrolia, Daniel R., Dennis Guignet, John Whitehead, Cannon Kent, Clay Caulder, and Kelvin Amon. Forthcoming. Nonmarket Valuation in the Environmental Protection Agency's Regulatory Process. *Applied Economic Perspectives and Policy*. doi: 10.1002/aepp.13106.

Qiu, Y., Yi David Wang, and Jianfent Wang. 2017. Soak up the sun: Impact of solar energy systems on residential home values in Arizona. *Energy Economics 66*: 328-336.

Rau, Tomás, Urzua Sergio, and Loreto Reyes. 2015. Early Exposure to Hazardous Waste and Academic Achievement: Evidence from a Case of Environmental Negligence. *Journal of the Association of Environmental and Resource Economists* 2(4): 527-563.

Schütt, Marvin. 2021. Systematic Variation in Waste Site Effects on Residential Property Values: A Meta-Regression Analysis and Benefit Transfer. *Environmental and Resource Economics* 78: 381-416.

Shadbegian, Ron, Dennis Guignet, Heather Klemick, and Linda Bui. 2019. Early Childhood Lead Exposure and the Persistence of Educational Consequences into Adolescence. *Environmental Research* 178: 108643.

Tonn, Bruce, Milton Russell, Ho-Ling Hwang, Richard Goeltz, and John Warren. 1991. Costs of RCRA Corrective Action: Interim Report. Oak Ridge National Laboratory, ORNL/TM-11864.

US Census Bureau, 2019. 2019 TIGER/Line Shapefiles (machine-readable data files) [WWW Document]. URL https://www.census.gov/cgi-bin/geo/shapefiles/index.php (accessed 5.19.20).

US EPA. 2005. U.S. Environmental Protection Agency. Introduction to Treatment, Storage and Disposal Facilities. EPA530-K-05-017, September. Accessed at: https://www.epa.gov/sites/production/files/2015-07/documents/tsdf05.pdf, 6 Apr 2021.

US EPA. 2008. Standards Applicable to Generators of Hazardous Water; Alternative Requirements for Hazardous Waste Determination and Accumulation of Unwanted Material at Laboratories Owned by Colleges and Universities and Other Eligible Academic Entities Formally Affiliated with Colleges and Universities; Final Rule, 40 Fed. Reg. 72912.

US EPA. 2013. RCRA Corrective Action: Case Studies Report. EPA 530-R-13-002, April. Accessed at: https://www.epa.gov/hw/case-studies-report-about-resource-conservation-and-recovery-act-rcra-corrective-action, 11 Feb 2021.

US EPA. 2019. Baselines for Resource Conservation and Recovery Act (RCRA) Corrective Action Sites. Accessed at: https://www.epa.gov/hw/baselines-resource-conservation-and-recovery-act-rcra-corrective-action-sites,7 Apr 2021.

US EPA. 2020. Learn about Corrective Action. Accessed at: https://www.epa.gov/hw/learn-about-corrective-action#whatis,7 Apr 2021.

US EPA. 2021a. National Primary Drinking Water Regulations. Accessed at: https://www.epa.gov/ground-water-and-drinking-water/national-primary-drinking-water-regulations, 3 Mar 2021.

US EPA. 2021b. RCRAInfo Web: Biennial Report Summary. Accessed at: https://rcrapublic.epa.gov/rcrainfoweb/action/modules/br/summary/view, 6 Apr 2021.

US HSS (Department of Health and Human Services). 1994. Toxicological Profile for Acetone. Accessed at: https://www.atsdr.cdc.gov/toxprofiles/tp21.pdf, 7 Apr 2021.

US HSS. 2010. Toxicological Profile for Ethylbenzene. Accessed at: https://www.atsdr.cdc.gov/toxprofiles/tp110.pdf, 7 Apr 2021.

US HSS. 2020. Toxicological Profile for Lead. Accessed at: https://www.atsdr.cdc.gov/toxprofiles/tp13.pdf, 7 Apr 2021.

Zillow. 2020. ZTRAX: Zillow Transaction and Assessor Dataset, 2019-Q4. Accessed at: https://www.zillow.com/research/ztrax/, 26 Feb 2020.

FIGURES AND TABLES

Figure 1. Treatment, Storage, and Disposal (TSD) Facilities and Corrective Actions.

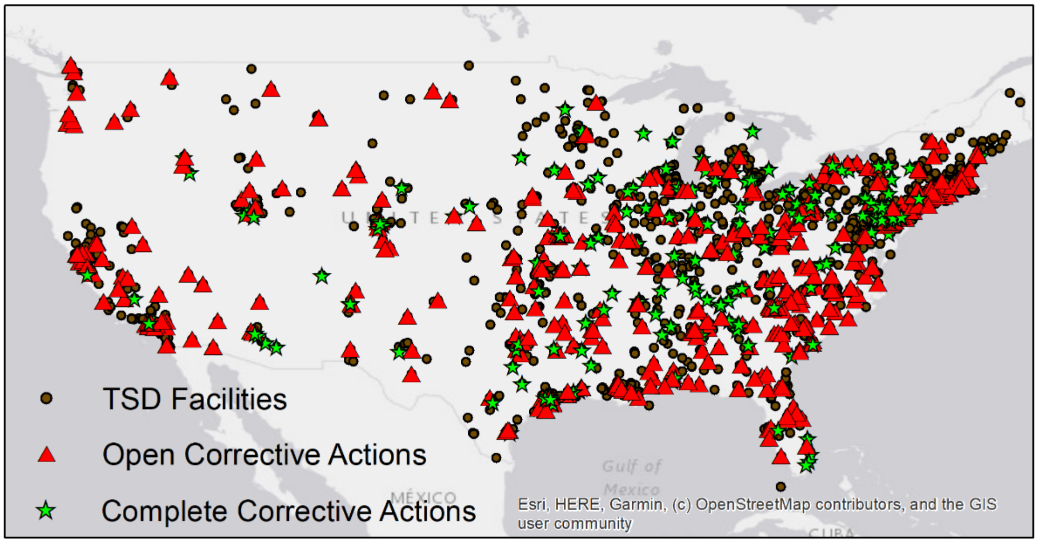


Figure 2. Number of Corrective Actions (CAs) Opened and Completed over Time.

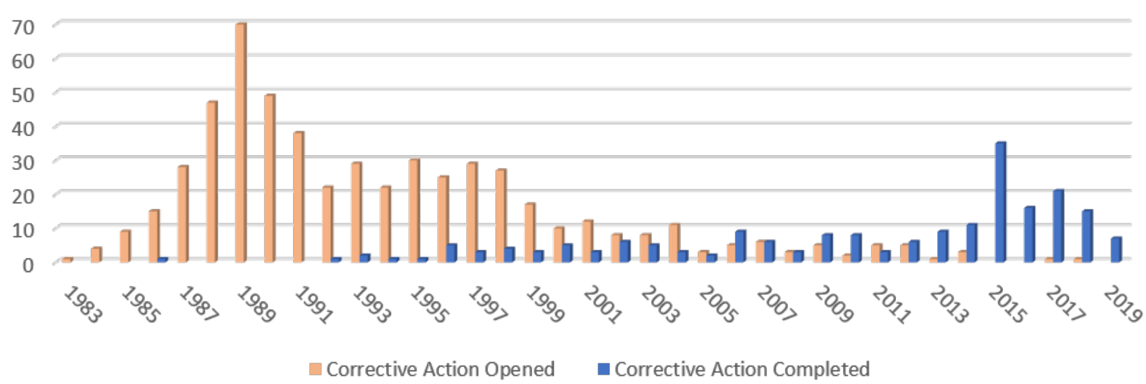


Figure 3. Most Common Hazardous Materials among the 689 TSDFs with a Corrective Action.

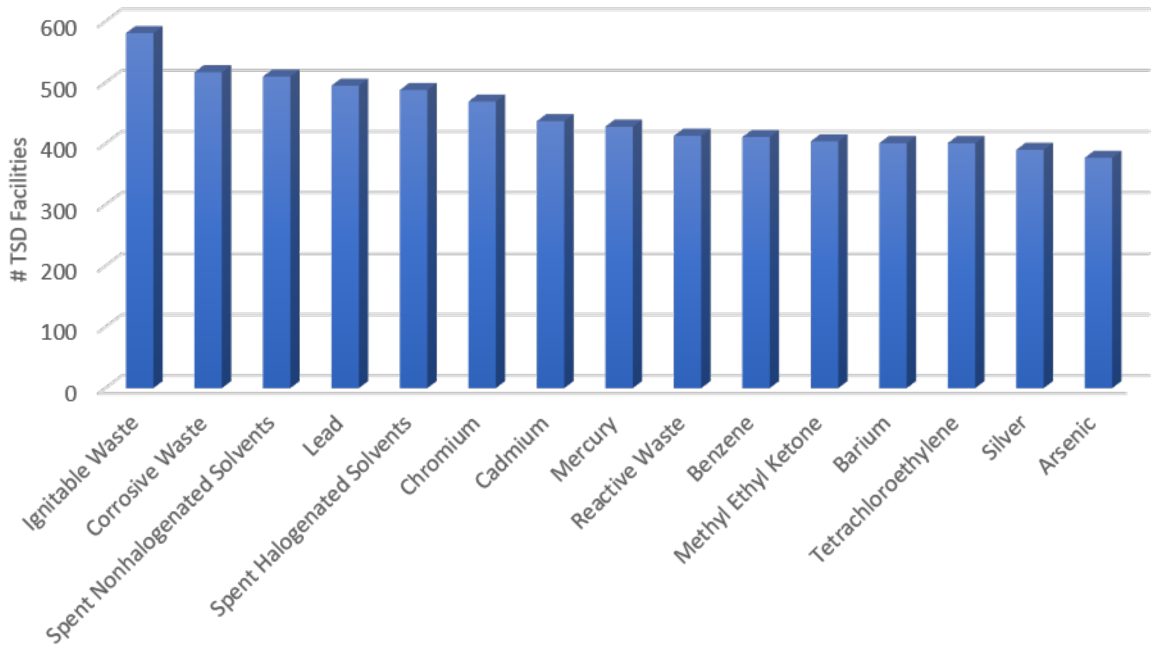


Figure 4. Counties included in Study Area.

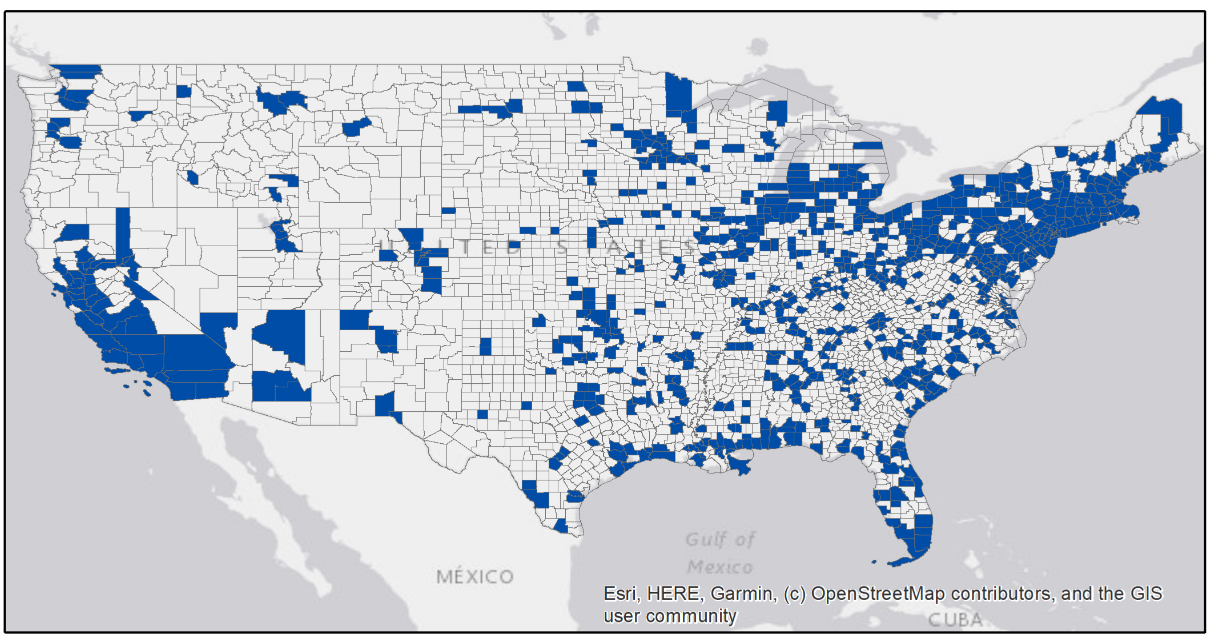


Figure 5. Number of Transactions in each 250m Distance Bin from a TSDF.

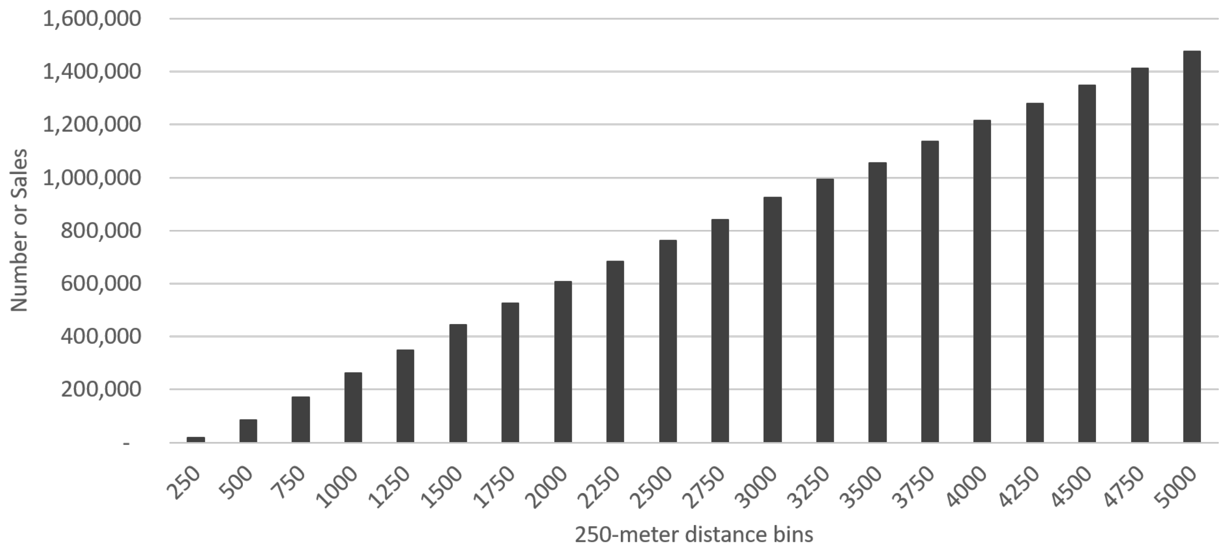


Figure 6. Number of Transactions in each 250-meter Distance Bin from a Corrective Action Site, by Stage.

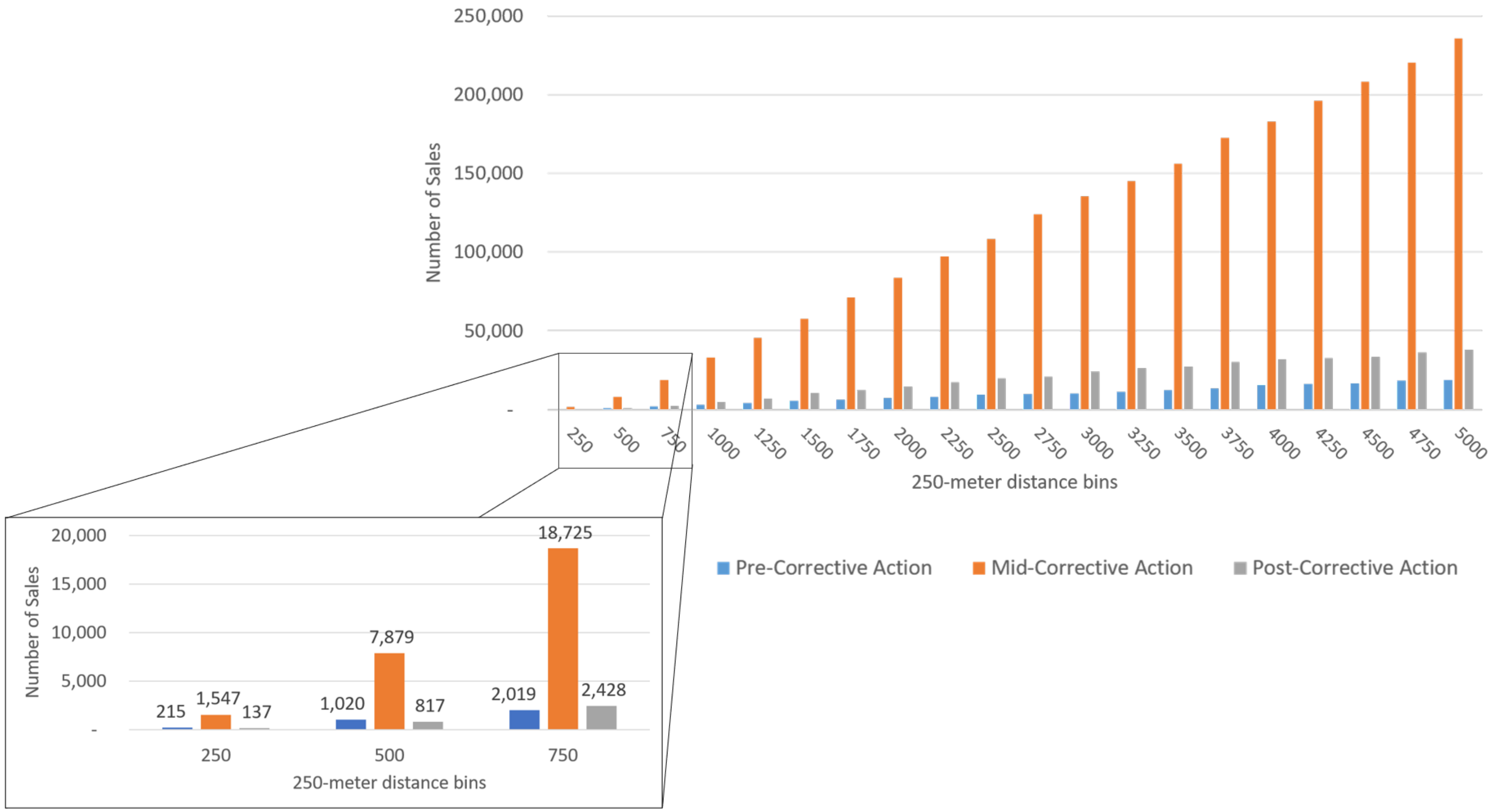


Figure 7. Difference-in-differences Illustration.

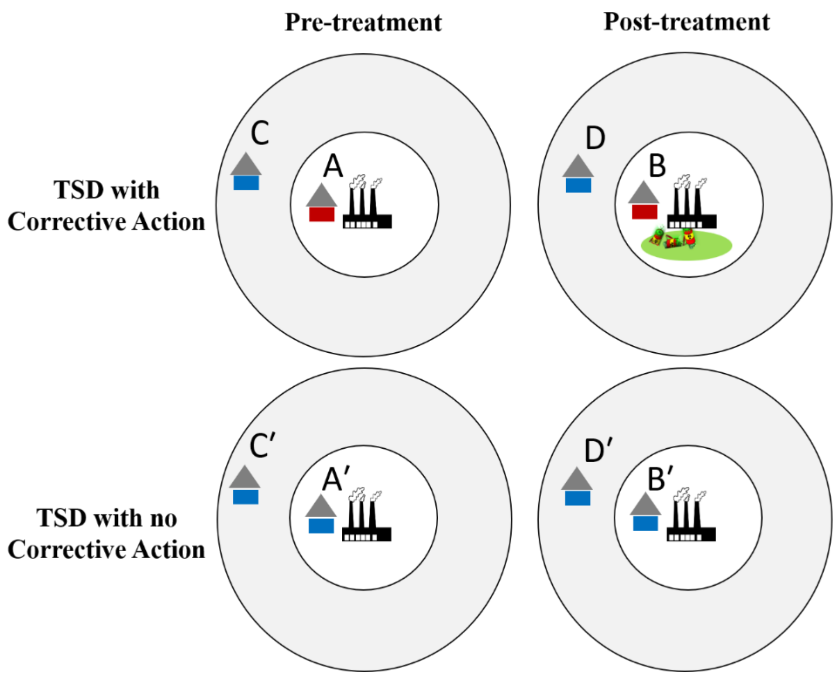
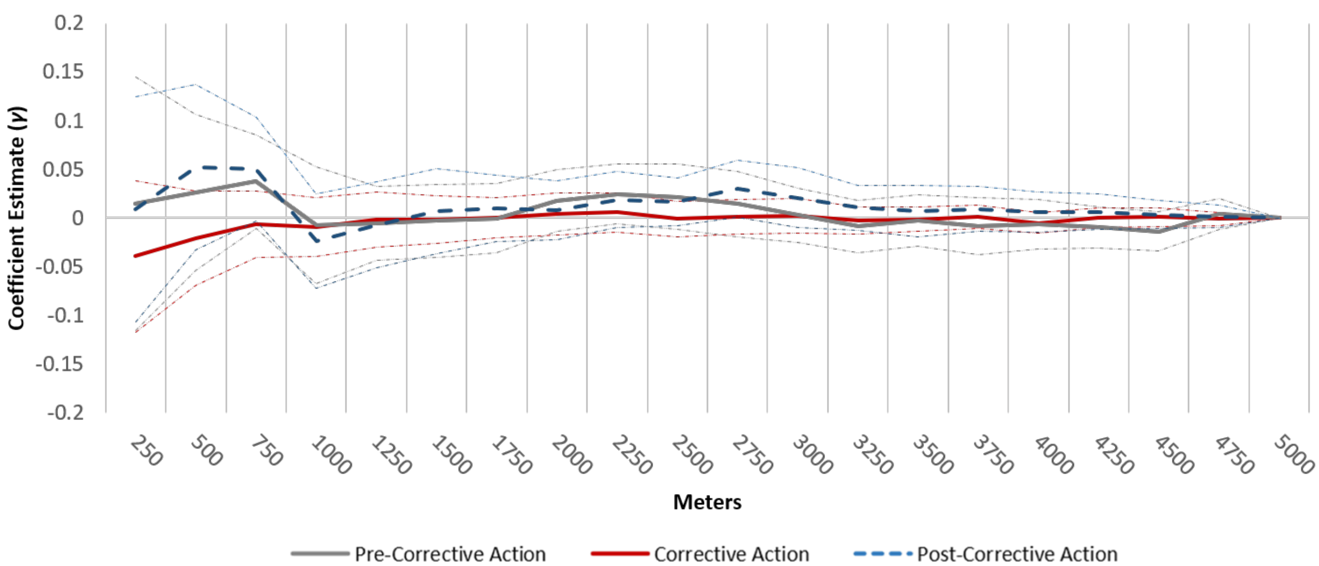


Figure 8. Price Gradients with respect to Distance from the Corrective Action.



Notes: 95% Confidence intervals displayed by light dotted lines. Distance gradients based on regression coefficient estimates corresponding to 250-meter incremental bins (see section IV.D for details). Coefficients estimated from hedonic regression model following equation (1), and using the sample of n=2,512,354 homes within 5,000 meters of a Corrective Action.

Figure 9. Estimated Price Gradient with respect to Proximity to TSDF (Model 1).

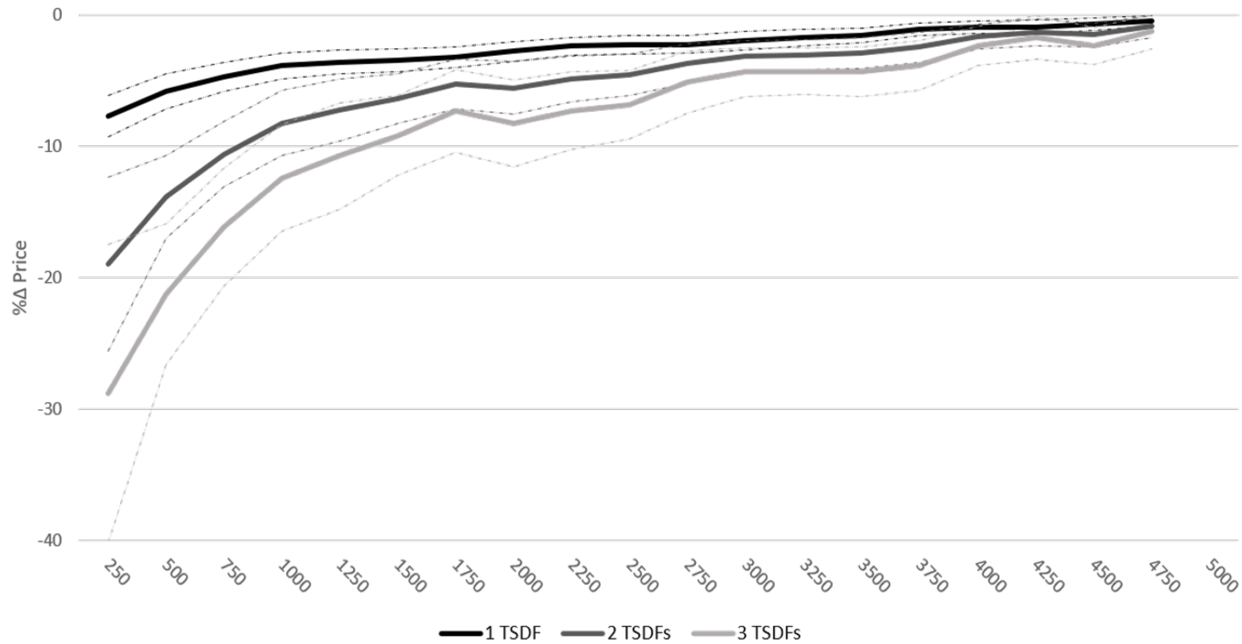
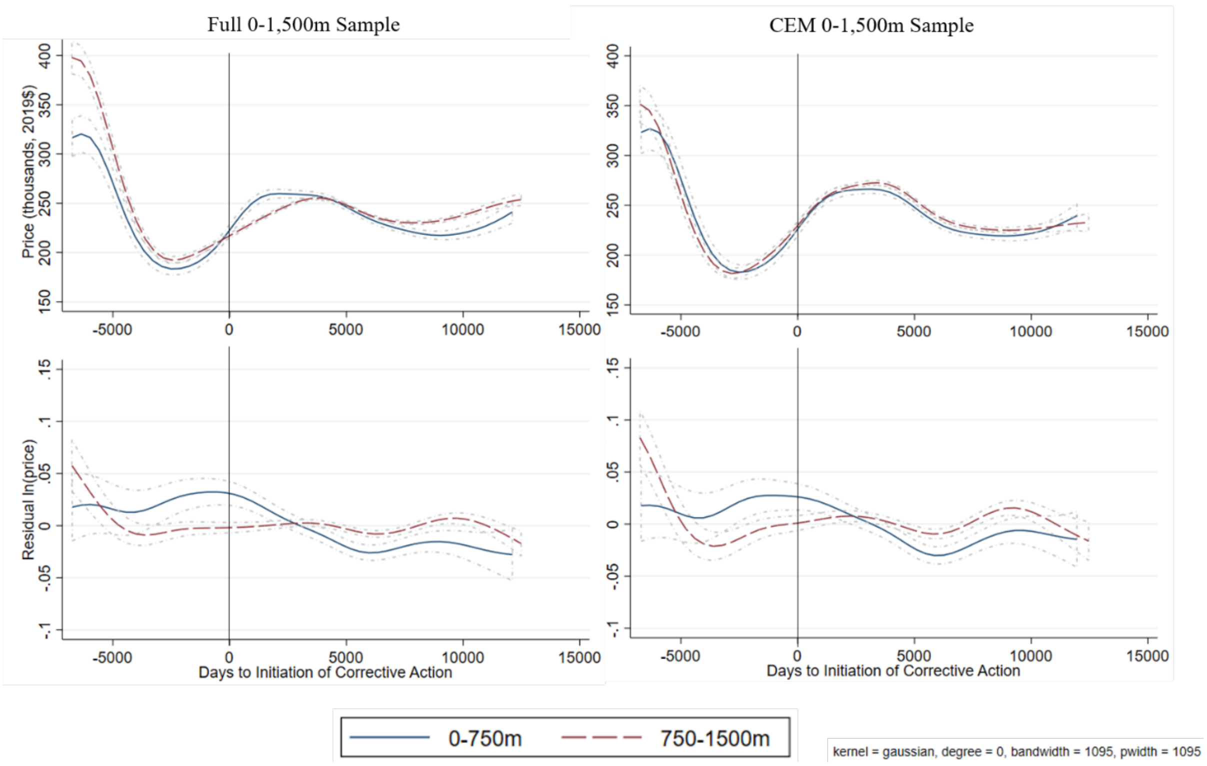
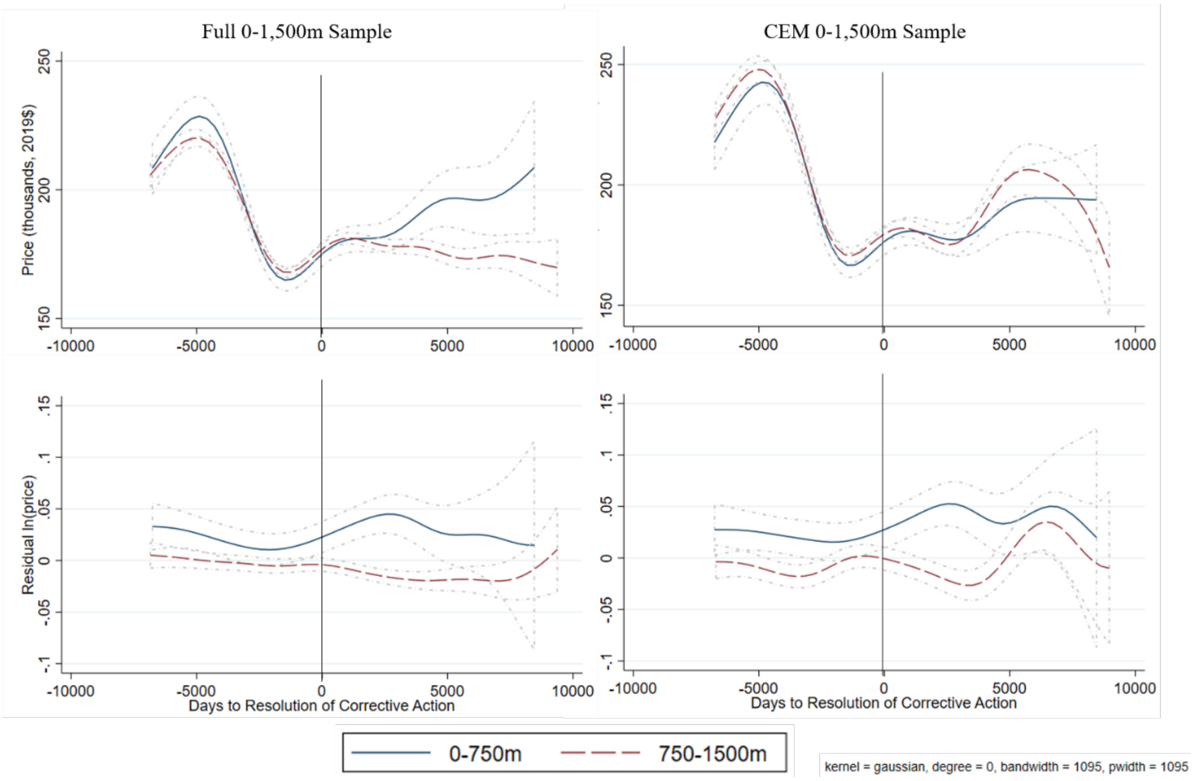


Figure 10. Treated and Control Group Price Trends Relative to Opening of CA.



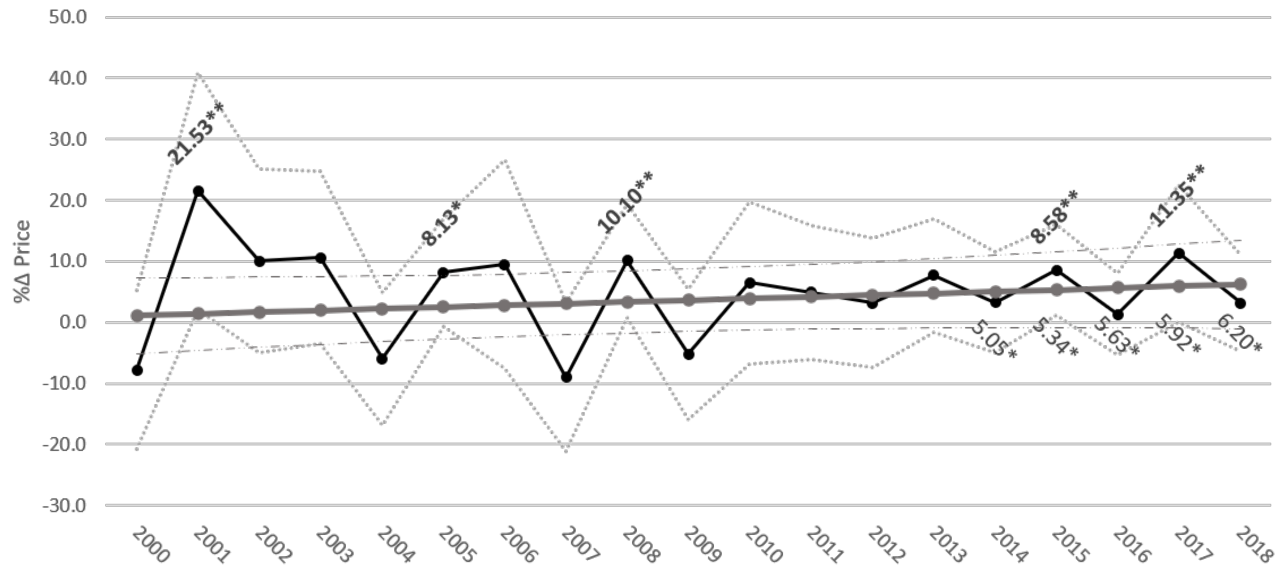
Notes: Trends generated using local polynomial regressions. The left column displays trends estimated using full sample of homes within 1,500m of a CA, and the trends in the right column are estimated using the CEM matched and weighted sample. The top panel shows the graphs of the house prices (2019$), and the lower panel displays the trends of the residuals from a regression model that conditions out otherwise confounding factors (see section VI.C for details). Due to the occasional occurrence of multiple CAs in close proximity to a home, for purposes of these trend graphs the control group is limited to homes within 750-1,500m of a CA in the pre-CA or mid-CA stage at the time of sale, and where there is no CA in any stage within 0-750m (n=145,245). The treated group is limited to homes within 0-750m of a CA in the pre-CA or mid-CA stage (but not both), and where there is no post-CA within 0-750m (n=31,379).

Figure 11. Treated and Control Group Price Trends Relative to Completion of CA.



Notes: Trends generated using local polynomial regressions. The left column displays trends estimated using full sample of homes within 1,500m of a CA, and the trends in the right column are estimated using the CEM matched and weighted sample. The top panel shows the graphs of the house prices (2019$), and the lower panel displays the trends of the residuals from a regression model that conditions out otherwise confounding factors (see section VI.C for details). Due to the occasional occurrence of multiple CAs in close proximity to a home, for purposes of these trend graphs the control group is limited to homes within 750-1,500m of a CA in the mid-CA or post-CA stage at the time of sale, and where there is no CA in any stage within 0-750m (n=154,779). The treated group is limited to homes within 0-750m of a CA in the mid-CA or post-CA stage (but not both), and where there is no pre-CA within 0-750m (n=31,508).

Figure 12. Price Impacts of Corrective Action Completion by Year.



Notes: * p<0.10, ** p<0.05, *** p<0.01. The solid black line shows the estimates of the percent changes in price due to completion of a CA, based on equation (8) and the results from Model 4′ (Table 9). The solid grey line shows how the percent change in price estimates vary by year when a linear trend is assumed. The 95% confidence intervals are displayed as dashed or dotted lines. Labels are included for only the statistically significant estimates (p<0.10).

Table 1. Summary of Benefit Analyses for Economically Significant RCRA Regulations by US EPA over Last 10 Years.

	Economically Significant RCRA Regulations	**Qualitative Benefits analysis**	**Avoided costs monetized** (e.g., regulatory, cleanup)	**Other benefits monetized** (e.g., mortality, morbidity)
2014	Revisions to Definition of Solid Waste Recycling Exclusions[1]	X		
2015	Coal Combustion Residuals (CCR) Rule[2]	X	X	X
2016	Hazardous Waste Export-Import Revisions[3]	X	X	
2016	Hazardous Waste Generator Improvements Rule[4]	X	X	
2018	Phase One Amendments to CCR Rule[5]		X	
2018	Safe Management of Recalled Airbags Rule[6]	X	X	
2019	Addition of Aerosol Cans to Universal Waste Rule[7]		X	

Notes: The full references for the Regulatory Impact Analysis (RIA) of each regulation are as follows: (1) US EPA. 2014. Regulatory Impact Analysis: EPA's 2014 Revisions to the Industrial Recycling Exclusions of the RCRA Definition of Solid Waste, Washington, DC, November; (2) US EPA. 2014. Regulatory Impact Analysis: EPA's 2015 RCRA Final Rule Regulating Coal Combustion Residual (CCR) Landfills and Surface Impoundments at Coal-Fired Utility Power Plants, Washington, DC, December; (3) US EPA. 2016. Regulatory Impact Analysis: Hazardous Waste Export-Import Revisions Final Rule, Washington, DC, May; (4) US EPA. 2016. Regulatory Impact Assessment of the Potential Costs, Benefits, and Other Impacts of the Final Hazardous Waste Generator Improvements Rule, Washington, DC, September; (5) US EPA. 2018. Regulatory Impact Analysis: EPA's 2018 RCRA Final Rule Disposal of Coal Combustion Residuals from Electric Utilities; Amendments to the National Minimum Criteria (Phase One), Washington, DC, July; (6) US EPA. 2018. Economic Assessment of the Safe Management of Recalled Airbags Interim Final Rule, Washington, DC, October; (7) US EPA. 2019. Regulatory Impact Analysis of the Final Rule to Add Aerosol Cans to the Universal Waste Rule, Washington, DC, October.

Table 2. Descriptive Statistics of Treatment, Storage, and Disposal Facilities (TSDFs).

Variable[a]	**Mean**	**Std. Dev.**	**Min**	**Max**
Large Quantity Generator (LQG)	0.433	0.496	0	1
Small Quantity Generator (SQG)	0.155	0.362	0	1
Accepts offsite hazardous materials	0.143	0.350	0	1
Recycling Facility	0.096	0.294	0	1
Number of Hazardous Materials	57.4	122.4	1	499

Notes: Descriptive statistics are for all n=2,389 unique TSDFs in the contiguous US. (a) Variables are all indicator variables unless otherwise noted, and the corresponding categories are not necessarily mutually exclusive.

Table 3. Industry Activities at Treatment, Storage, and Disposal Facilities (TSDFs).

Variable[a]	**Mean**	**Std. Dev.**	**Min**	**Max**
Agriculture, Forestry, Fishing and Hunting	0.026	0.159	0	1
Mining	0.020	0.140	0	1
Utilities	0.047	0.211	0	1
Construction	0.033	0.180	0	1
Manufacturing	0.596	0.491	0	1
Trade, Transportation, and Warehousing	0.239	0.427	0	1
Services	0.308	0.462	0	1
Waste Management	0.254	0.436	0	1

Notes: Descriptive statistics are for all n=2,209 (out of 2,389) unique TSDFs in contiguous US where North American Industry Classification System (NAICS) data were available in RCRAInfo. (a) Variables are all indicator variables, and are not necessarily mutually exclusive. Industry categories are based on aggregations of the two-digit NAICS codes, as follows: Agriculture, Forestry, Fishing, and Hunting (NAICS 11); Mining (NAICS 21); Utilities (NAICS 22); Construction (NAICS 23); Manufacturing (NAICS 31-33); Trade, Transportation, and Warehousing (NAICS 42,44-45, and 48-49), and Waste Management (NAICS 56). The "Services" category is an aggregation of the following two-digit NAICS codes: information (NAICS code 51), finance and insurance (NAICS code 52), real estate rental (NAICS code 53), professional/technical services (NAICS code 54), education (NAICS code 61), health care (NAICS code 62), and others (NAICS codes 55, 71,72,81, and 92).

Table 4. Descriptive Statistics of Corrective Actions (CAs).

Variable[a]	Mean	Std. Dev.	Min	Max
Identified as high priority by ORCR	0.975	0.155	0	1
Risk prioritization (1=high, 2= medium, 3=low)[b]	1.640	0.791	1	3
Risk prioritization missing	0.028	0.164	0	1
Remedy constructed	0.763	0.425	0	1
Physical controls implemented	0.409	0.492	0	1
Institutional controls implemented	0.486	0.500	0	1
CA complete for entire facility	0.293	0.456	0	1
CA duration (years)[c]	15.04	9.12	0.00	30.94

Notes: Descriptive statistics are for all n=689 unique TSDFs in contiguous US where a Corrective Action (CA) involving actual contamination was opened. (a) Variables are all indicator variables unless otherwise noted. (b) The risk prioritization variable has missing values for 19 of the 689 facilities. These missing values are excluded from the risk prioritization descriptive statistics above. (c) CA duration calculated for the 146 (out of 202) TSDFs where the CA was marked as complete in RCRAInfo, and where the opening date was not missing. Duration was coded as missing for three sites where the CA opening date was listed as occurring after the completion date.

Table 5. Residential Transactions Descriptive Statistics.

Variable	Obs	Mean	Std. Dev.	Min	Max
Price (2019$)	9,655,158	273,031	202,269	20,001	999,982
Acres	9,402,685	0.252	0.236	0.050	2.000
Missing: Acres†	9,655,158	0.026	0.160	0	1
Stories	7,822,515	1.410	0.503	1.0	3.0
Missing: Stories†	9,655,158	0.190	0.392	0	1
Bathrooms	7,428,256	1.858	0.759	1	6
Missing: Bathrooms†	9,655,158	0.231	0.421	0	1
Interior square footage	9,162,177	2,958	1,902	750	12,000
Missing: Interior square footage†	9,655,158	0.051	0.220	0	1
Age (years)	9,002,285	45.946	29.517	0	150
Missing: Age (years)†	9,655,158	0.068	0.251	0	1
% Land Developed w/in 0-200m	9,655,158	54.549	28.394	0	100
% Land Developed w/in 0-500m	9,655,158	50.404	23.462	0	100
Highway w/in 500m†	9,655,158	0.389	0.487	0	1
# TSDFs w/in 0-750m	9,655,158	0.03	0.17	0	3
# TSDFs w/in 0-1500m	9,655,158	0.14	0.40	0	6
# TSDFs w/in 0-5000m	9,655,158	1.73	1.28	1	18

Note: The final sample includes n=9,655,158 single-family home transactions. Descriptive statistics for some variables are for a smaller sample due to missing values, as reflected by the corresponding missing value indicators. Variables denoted with † are binary indicators.

Table 6. Percent Change in Price by 250m Treatment Bin: Broader 5km Sample.

	Model 1		**Model 2**
	TSDF 5km Sample		**CA 5km Sample**
	Inter-area DID	**Triple Differences**	**Intra-area DID**
Corrective Action Opened			
0 - 250m	-6.0488	-6.0271	-5.4310
	(5.2949)	(5.4566)	(5.5639)
250 - 500m	-5.5737*	-5.5519*	-4.8546
	(3.0746)	(3.0597)	(3.2209)
500 - 750m	-4.6508**	-4.6288**	-4.4798**
	(2.2282)	(2.0776)	(1.9920)
Corrective Action Completed			
0 - 250m	1.2203	3.0748	5.4544
	(5.9036)	(5.5915)	(6.2958)
250 - 500m	5.4998	7.4327*	7.9086*
	(4.4316)	(3.9951)	(4.0436)
500 - 750m	5.3839*	7.3146***	6.0865***
	(2.8552)	(2.4172)	(2.3520)
Tract Fixed Effects	Yes		Yes
House and Location Attributes	County × Year		County × Year
Year Fixed Effects	County × Year		County × Year
Quarter Fixed Effects	County × Quarter		County × Quarter
Observations	9,653,889		2,512,354
Adjusted R-squared	0.784		0.784

Notes: * p<0.10, ** p<0.05, *** p<0.01. Standard errors in parentheses, clustered at the county level. Estimates calculated following equations (2)-(5) using the "nlcom" command in Stata 17/MP, and are based on the coefficient estimates from the hedonic regression models 1 and 2. Full regression coefficient results presented in Table A.1 of the Appendix.

Table 7. Percent Change in Price for 0-750m Bin.

	Model 3		**Model 4**	**Model 5**	**Model 6**
	TSDF 5km Sample		**CA 5km Sample**	**CA 1500m Sample**	**CA 1500m CEM Sample**
	Inter-area DID	**Triple Differences**	**Intra-area DID**	**Intra-area DID**	**Intra-area DID**
Corrective Action Opened					
0 - 750m	-5.0128**	-4.9815**	-4.6267**	-2.5431	-2.7196
	(2.2552)	(2.1763)	(2.1603)	(2.2201)	(2.3786)
Corrective Action Completed					
0 - 750m	5.3000*	7.2290***	6.5213***	6.5489**	7.4924***
	(3.0158)	(2.5140)	(2.4709)	(2.6099)	(2.7229)
Tract Fixed Effects	Yes		Yes	Yes	Yes
House and Location Attributes	County × Year		County × Year	County × Year	County × Year
Year Fixed Effects	County × Year		County × Year	County × Year	County × Year
Quarter Fixed Effects	County × Quarter		County × Quarter	County × Quarter	County × Quarter
Observations	9,653,889		2,512,354	201,495	99,310
Adjusted R-squared	0.784		0.784	0.788	0.811

Notes: * p<0.10, ** p<0.05, *** p<0.01. Standard errors in parentheses, clustered at the county level. Estimates calculated following equations (2)-(5) using the "nlcom" command in Stata 17/MP, and are based on the coefficient estimates from the hedonic regression models 1 and 2. Full regression coefficient results presented in Table A.1 of the Appendix.

Table 8. Regression-based statistical tests of pre-treatment parallel trends.

	(1) CA 1500m Sample[a]	(2) CA 1500m CEM Sample[a]	(3) CA 1500m Sample[b]	(4) CA 1500m CEM Sample[b]
0-750m of Corrective Action Site	0.0966	0.1348	0.0111	0.1655
	(0.1133)	(0.1172)	(0.0852)	(0.1029)
Corrective Action Opened				
Years before	0.0950	0.0679		
	(0.0813)	(0.1052)		
Years before × 0-750m	0.0024	0.0003		
	(0.0048)	(0.0049)		
Corrective Action Completed				
Years before			0.1292**	0.1520***
			(0.0505)	(0.0549)
Years before × 0-750m			-0.0022	0.0018
			(0.0039)	(0.0062)
Tract Fixed Effects	Yes	Yes	Yes	Yes
House and Location Attributes	County × Year	County × Year	County × Year	County × Year
Year Fixed Effects	County × Year	County × Year	County × Year	County × Year
Quarter Fixed Effects	County × Quarter	County × Quarter	County × Quarter	County × Quarter
Observations	15,832	8,055	23,608	10,890
Adjusted R-squared	0.811	0.813	0.715	0.754

Notes: * p<0.10, ** p<0.05, *** p<0.01. Standard errors in parentheses, clustered at the county level. Regression models estimated using "reghdfe" command in Stata 17/MP. (a) Regressions in columns (1) and (2) are estimated using the subsample of transactions that occur prior to the opening of the first Corrective Action within 1500m. (b) Regressions in columns (3) and (4) are estimated using the subsample of transactions in the treated group (0-750m) that occur after the opening, but before the completion of, the first Corrective Action within 750m; as well as the control group (750-1,500m) transactions that occur prior to completion of the first Corrective Action within 750-1,500m and that had no pre-, mid-, or post-Corrective Action sites within the 0-750m treatment zone.

Table 9. Policy Illustration of Capitalization Effects versus Ex Post Lower Bound Welfare Estimates.

	Model 4 Time Invariant CA Effects	**Model 4′ Time variant CA Effects**				**Model 4″ Time variant CA Effects**
Assumed Ex Post Price Surface	**NA**	**2015**	**2016**	**2017**	**2018**	**2015-2018**
Percent Increase	6.5213***	8.5838**	1.3415	11.3531**	3.1579	5.3344*
	[1.6783, 11.3642]	[1.1200 - 16.0476]	[-5.2571 - 7.9400]	[0.1278 - 22.5784]	[-4.7679 - 11.0837]	[-0.6441 - 11.3128]
Mean Price Increase (2019$)	$10,149***	$13,105**	$2,194	$16,902**	$5,075	$8,395*
	[3,073 - 17,225]	[2,611 - 23,600]	[-8,457 - 12,846]	[1,894 - 31,910]	[-7,272 - 17,422]	[-537 - 17,328]
Total Increase (millions, 2019$)	$184.5***	$238.2**	$39.9	$307.3**	$92.3	$152.6*
	[55.9 - 313.1]	[47.5 - 429.0]	[-153.7 - 233.5]	[34.4 - 580.1]	[-132.2 - 316.7]	[-9.8 - 315.0]

Note: * p<0.10, ** p<0.05, *** p<0.01. 95% confidence intervals displayed in brackets. First column displays the estimated capitalization effects based on model 4 (in table 7). Subsequent columns display the estimated price changes along a fixed ex post price surface, and are based on a variant of model 4 where all coefficients with respect to CA stages are allowed to vary by year. Additionally, models 4′ and 4″ include separate interaction terms to allow the coefficients corresponding to the presence of a TSDF and number of TSDFs in each incremental 250-meter bin to vary by year and by county. Model 4″ constraints the CA effects to be the same across all ex post years, 2015-2018.

APPENDIX

Table A1. Full Hedonic Property Value Regression Results: Broader 5km Sample.

	Model 1 TSDF 5km Sample	Model 2 CA 5km Sample
TSDF Present†		
0-250m	0.0493	0.1167**
	(0.0400)	(0.0592)
250-500m	0.0294	0.1173***
	(0.0179)	(0.0248)
500-750m	0.0154	0.0048
	(0.0153)	(0.0291)
750-1000m	0.0068	0.0247*
	(0.0085)	(0.0135)
1000-1250m	0.0019	0.0147
	(0.0089)	(0.0141)
1250-1500m	-0.0043	-0.0178**
	(0.0064)	(0.0088)
1500-1750m	-0.0106*	-0.0138*
	(0.0057)	(0.0071)
1750-2000m	0.0008	0.0048
	(0.0063)	(0.0096)
2000-2250m	0.0016	0.0024
	(0.0062)	(0.0085)
2250-2500m	0.0009	-0.0053
	(0.0049)	(0.0065)
2500-2750m	-0.0082**	-0.0038
	(0.0034)	(0.0051)
2750-3000m	-0.0076**	-0.0070
	(0.0038)	(0.0045)
3000-3250m	-0.0045	-0.0100**
	(0.0029)	(0.0044)
3250-3500m	-0.0012	-0.0031
	(0.0033)	(0.0064)
3500-3750m	0.0031	-0.0043
	(0.0034)	(0.0049)
3750-4000m	-0.0025	-0.0111***
	(0.0032)	(0.0041)
4000-4250m	-0.0049**	-0.0073**
	(0.0024)	(0.0036)
4250-4500m	0.0016	0.0020
	(0.0022)	(0.0040)
4500-4750m	-0.0005	-0.0033
	(0.0021)	(0.0036)

Number of TSDFs		
0-250m	-0.1297***	-0.1990***
	(0.0399)	(0.0602)
250-500m	-0.0894***	-0.1716***
	(0.0168)	(0.0186)
500-750m	-0.0638***	-0.0603**
	(0.0136)	(0.0259)
750-1000m	-0.0465***	-0.0687***
	(0.0102)	(0.0170)
1000-1250m	-0.0385***	-0.0515***
	(0.0102)	(0.0170)
1250-1500m	-0.0307***	-0.0201**
	(0.0072)	(0.0094)
1500-1750m	-0.0218***	-0.0210**
	(0.0072)	(0.0086)
1750-2000m	-0.0290***	-0.0343***
	(0.0079)	(0.0125)
2000-2250m	-0.0257***	-0.0281***
	(0.0071)	(0.0105)
2250-2500m	-0.0240***	-0.0189**
	(0.0058)	(0.0084)
2500-2750m	-0.0147***	-0.0187**
	(0.0049)	(0.0077)
2750-3000m	-0.0122***	-0.0141**
	(0.0039)	(0.0060)
3000-3250m	-0.0131***	-0.0085**
	(0.0035)	(0.0042)
3250-3500m	-0.0143***	-0.0119*
	(0.0041)	(0.0072)
3500-3750m	-0.0140***	-0.0085
	(0.0042)	(0.0053)
3750-4000m	-0.0071**	-0.0007
	(0.0032)	(0.0041)
4000-4250m	-0.0042	-0.0017
	(0.0032)	(0.0038)
4250-4500m	-0.0085***	-0.0103**
	(0.0028)	(0.0042)
4500-4750m	-0.0041	-0.0031
	(0.0027)	(0.0040)
4750-5000m	-0.0031*	-0.0030
	(0.0019)	(0.0027)
Pre-Corrective Action†		
0-250m	0.0136	0.0110
	(0.0616)	(0.0646)

250-500m	0.0351	0.0234
	(0.0326)	(0.0367)
500-750m	0.0309	0.0340*
	(0.0215)	(0.0205)
750-1500m	-0.0079	-0.0096
	(0.0182)	(0.0137)
1500-5000m	-0.0000	
	(0.0093)	
Mid-Corrective Action†		
0-250m	-0.0488*	-0.0406
	(0.0296)	(0.0358)
250-500m	-0.0223	-0.0222
	(0.0200)	(0.0213)
500-750m	-0.0167	-0.0076
	(0.0139)	(0.0137)
750-1500m	-0.0081	-0.0053
	(0.0078)	(0.0074)
1500-5000m	-0.0027	
	(0.0057)	
Post-Corrective Action†		
0-250m	-0.0367	0.0025
	(0.0515)	(0.0582)
250-500m	0.0313	0.0439
	(0.0397)	(0.0393)
500-750m	0.0357	0.0414*
	(0.0238)	(0.0217)
750-1500m	-0.0263	-0.0154
	(0.0196)	(0.0140)
1500-5000m	-0.0092	
	(0.0152)	
Constant	12.2474***	12.1692***
	(0.0044)	(0.0079)
Tract Fixed Effects	Yes	Yes
House and Location Attributes	County × Year	County × Year
Year Fixed Effects	County × Year	County × Year
Quarter Fixed Effects	County × Quarter	County × Quarter
Observations	9,653,889	2,512,354
Adjusted R-squared	0.784	0.784

Notes: Dependent variable is ln(price). * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Standard errors in parentheses, clustered at the county level. Regression models estimated using "reghdfe" command in Stata 17/MP. Variables denoted with † are binary indicators.

Table A2. Full Hedonic Property Value Regression Results with Pooled 0-750m Treatment Bin.

	Model 3 TSDF 5km Sample	**Model 4 CA 5km Sample**	**Model 5 CA 1500m Sample**	**Model 6 CA 1500m CEM Sample**
TSDF Present[†]				
0-250m	0.0535	0.1179*	0.1705***	0.1290***
	(0.0412)	(0.0602)	(0.0553)	(0.0382)
250-500m	0.0293	0.1142***	0.0798*	0.0054
	(0.0182)	(0.0246)	(0.0411)	(0.0480)
500-750m	0.0158	0.0073	0.0039	-0.0078
	(0.0153)	(0.0291)	(0.0420)	(0.0379)
750-1000m	0.0068	0.0246*	0.0294	-0.0004
	(0.0085)	(0.0135)	(0.0195)	(0.0337)
1000-1250m	0.0019	0.0146	0.0177	0.0393
	(0.0089)	(0.0141)	(0.0223)	(0.0263)
1250-1500m	-0.0043	-0.0178**		
	(0.0064)	(0.0088)		
1500-1750m	-0.0106*	-0.0139*		
	(0.0057)	(0.0071)		
1750-2000m	0.0008	0.0047		
	(0.0063)	(0.0097)		
2000-2250m	0.0016	0.0024		
	(0.0062)	(0.0085)		
2250-2500m	0.0009	-0.0053		
	(0.0049)	(0.0066)		
2500-2750m	-0.0082**	-0.0038		
	(0.0034)	(0.0051)		
2750-3000m	-0.0076**	-0.0071		
	(0.0038)	(0.0045)		
3000-3250m	-0.0046	-0.0100**		
	(0.0029)	(0.0044)		
3250-3500m	-0.0012	-0.0031		
	(0.0033)	(0.0064)		
3500-3750m	0.0031	-0.0043		
	(0.0034)	(0.0049)		
3750-4000m	-0.0025	-0.0111***		
	(0.0032)	(0.0041)		
4000-4250m	-0.0049**	-0.0072**		
	(0.0024)	(0.0036)		
4250-4500m	0.0016	0.0020		
	(0.0022)	(0.0040)		
4500-4750m	-0.0005	-0.0034		
	(0.0021)	(0.0036)		
Number of TSDFs				
0-250m	-0.1370***	-0.2111***	-0.2580***	-0.2515***
	(0.0407)	(0.0556)	(0.0391)	(0.0333)

250-500m	-0.0896***	-0.1722***	-0.1267***	-0.0714***
	(0.0169)	(0.0192)	(0.0305)	(0.0187)
500-750m	-0.0639***	-0.0604**	-0.0396	-0.0533
	(0.0136)	(0.0255)	(0.0344)	(0.0352)
750-1000m	-0.0465***	-0.0687***	-0.0478**	-0.0098
	(0.0102)	(0.0170)	(0.0194)	(0.0326)
1000-1250m	-0.0385***	-0.0514***	-0.0273	-0.0438*
	(0.0102)	(0.0170)	(0.0189)	(0.0235)
1250-1500m	-0.0307***	-0.0201**	-0.0039	-0.0005
	(0.0072)	(0.0094)	(0.0072)	(0.0104)
1500-1750m	-0.0218***	-0.0209**		
	(0.0072)	(0.0085)		
1750-2000m	-0.0290***	-0.0343***		
	(0.0079)	(0.0126)		
2000-2250m	-0.0257***	-0.0281***		
	(0.0071)	(0.0105)		
2250-2500m	-0.0239***	-0.0189**		
	(0.0059)	(0.0084)		
2500-2750m	-0.0147***	-0.0187**		
	(0.0050)	(0.0078)		
2750-3000m	-0.0122***	-0.0141**		
	(0.0039)	(0.0060)		
3000-3250m	-0.0131***	-0.0085**		
	(0.0035)	(0.0042)		
3250-3500m	-0.0143***	-0.0120*		
	(0.0041)	(0.0071)		
3500-3750m	-0.0141***	-0.0086		
	(0.0042)	(0.0053)		
3750-4000m	-0.0071**	-0.0008		
	(0.0032)	(0.0041)		
4000-4250m	-0.0042	-0.0017		
	(0.0032)	(0.0038)		
4250-4500m	-0.0085***	-0.0103**		
	(0.0028)	(0.0042)		
4500-4750m	-0.0041	-0.0031		
	(0.0027)	(0.0040)		
4750-5000m	-0.0031*	-0.0030		
	(0.0019)	(0.0027)		
Pre-Corrective Action†				
0-750m	0.0316	0.0301	0.0293	0.0599
	(0.0229)	(0.0234)	(0.0334)	(0.0492)
750-1500m	-0.0078	-0.0095		
	(0.0182)	(0.0137)		

1500-5000m	-0.0000			
	(0.0093)			
Mid-Corrective Action†				
0-750m	-0.0198	-0.0131	0.0229	0.0514
	(0.0147)	(0.0149)	(0.0332)	(0.0466)
750-1500m	-0.0082	-0.0053	0.0194	0.0191
	(0.0078)	(0.0074)	(0.0185)	(0.0211)
1500-5000m	-0.0027			
	(0.0057)			
Post-Corrective Action†				
0-750m	0.0318	0.0400*	0.0709*	0.0216
	(0.0257)	(0.0237)	(0.0382)	(0.0493)
750-1500m	-0.0263	-0.0153	0.0039	-0.0830**
	(0.0196)	(0.0140)	(0.0237)	(0.0396)
1500-5000m	-0.0092			
	(0.0152)			
Constant	12.2474***	12.1693***	12.0086***	12.0375***
	(0.0044)	(0.0079)	(0.0177)	(0.0186)
Tract Fixed Effects	Yes	Yes	Yes	Yes
House and Location Attributes	County × Year	County × Year	County × Year	County × Year
Year Fixed Effects	County × Year	County × Year	County × Year	County × Year
Quarter Fixed Effects	County × Quarter	County × Quarter	County × Quarter	County × Quarter
Observations	9,653,889	2,512,354	201,495	99,310
Adjusted R-squared	0.784	0.784	0.788	0.811

Notes: Dependent variable is ln(price). * p<0.10, ** p<0.05, *** p<0.01. Standard errors in parentheses, clustered at the county level. Regression models estimated using "reghdfe" command in Stata 17/MP. Variables denoted with † are binary indicators.

Table A3. Comparison of Sample Means Across Treated and Control Groups.

	Unweighted 0-1,500m Sample				CEM-Weighted 0-1,500m Sample			
Variable	Control: 750-1500m CA	Treated: 0-750m CA	t-stat		Control: 750-1500m CA	Treated: 0-750m CA	t-stat	
Acres	0.245	0.236	6.54	***	0.227	0.222	-3.48	***
Stories	1.42	1.44	-6.96	***	1.42	1.42	2.12	**
Bathrooms	1.74	1.72	3.65	***	1.66	1.65	-0.51	
Interior Square Footage	3,019	3,070	-4.59	***	3,036	3,019	-1.27	
Age	48.0	50.2	-10.90	***	50.3	50.7	1.82	*
% developed w/in 200m	55.29	49.70	32.88	***	49.33	49.44	0.54	
% developed w/in 500m	49.89	43.16	50.84	***	44.47	43.32	-6.88	***
Highway w/in 500m†	0.47	0.58	-37.97	***	0.51	0.58	20.03	***
# TSDs 0-1500m	1.21	1.23	-5.63	***	1.22	1.23	3.70	***
# TSDs 0-5000m	2.58	2.71	-9.91	***	2.90	2.88	-1.32	

Note: Variables denoted with † are binary indicators. Sample sizes vary across variables because of excluded observations due to missing values, but in the full 0-1,500m sample there are n=167,238 and n=34,761 observations in the 750-1500m control group and 0–750m treated group, respectively. The sample sizes are less in the CEM matched sample, with n=72,554 in the control group and n=26,898 in the treated group.

TableA4. Differences in Univariate Distributions between Treated and Control Groups: Unweighted 1,500m of a CA Sample.

	L1	mean	min	25%	50%	75%	max
Acres	0.0677	-0.0093	0	-0.002	-0.005	0	0
Missing: Acres†	0.0018	0.0018	0	0	0	0	0
Stories	0.0269	0.0273	0	0	0	0	0
Missing: Stories†	0.0053	-0.0053	0	0	0	0	0
Bathrooms	0.0310	-0.0141	0	0	0	0	0
Missing: Bathrooms†	0.0004	0.0004	0	0	0	0	0
Interior square footage	0.0501	63.3690	0	67	80	65	-16
Missing: Interior square footage†	0.0048	-0.0048	0	0	0	0	0
Age (years)	0.0444	2.4315	0	0	3	3	0
Missing: Age (years)†	0.0072	-0.0072	0	0	0	0	0
% Land Developed w/in 0-200m	0.0983	-5.5883	0	-8	-8.1	-6.3	0
% Land Developed w/in 0-500m	0.1502	-6.7327	0	-8.5	-8.4	-8.3	0.2
Highway w/in 500m†	0.1106	0.1106	0	0	1	0	0
# TSDFs w/in 0-5000m	0.0558	0.1306	0	0	0	1	-1

Notes: Variables denoted with † are binary indicators. This table displays the differences in the mean, min, max, and quartiles across the treated (0-750m of a CA) and control (750-1,500m of a CA) groups for each univariate distribution. The *L1* statistics comparing the full distributions are also provided. Table generated using "imb" command in Stata 17/MP, using the Scott algorithm to coarsen the distributions when calculating the *L1* distances (for details see Blackwell et al., 2009; Iacus et al., 2012). The multivariate *L1* distance is 0.9986.

Table A5. Differences in Univariate Distributions between Treated and Control Groups: CEM-Weighted 1,500m of a CA Sample.

	L1	mean	min	25%	50%	75%	max
Acres	0.0541	-0.0055	0	0	0	0	0
Missing: Acres†	2.9E-15	9.0E-16	0	0	0	0	0
Stories	0.0248	-0.0070	0	0	0	0	0
Missing: Stories†	0.0101	0.0101	0	0	0	0	0
Bathrooms	0.0076	-0.0021	0	0	0	0	0
Missing: Bathrooms†	1.6E-14	2.1E-14	0	0	0	0	0
Interior square footage	0.0394	-16.2670	0	-21	-9	-37	-25
Missing: Interior square footage†	3.2E-15	4.3E-15	0	0	0	0	0
Age (years)	0.0482	0.4106	0	-1	0	0	0
Missing: Age (years)†	4.7E-15	5.8E-15	0	0	0	0	0
% Land Developed w/in 0-200m	0.0644	0.1139	0	1.4	-1.1	-1.1	0
% Land Developed w/in 0-500m	0.1310	-1.1493	0	-1	-2.5	-2.7	1
Highway w/in 500m†	0.0713	0.0713	0	0	0	0	0
# TSDFs w/in 0-5000m	0.0598	-0.0224	0	0	0	0	-1

Notes: Variables denoted with † are binary indicators. This table displays the differences in the mean, min, max, and quartiles across the treated (0-750m of a CA) and control (750-1,500m of a CA) groups for each univariate distribution. The *L1* statistics comparing the full distributions are also provided. Table generated using "imb" command in Stata 17/MP, using the Scott algorithm to coarsen the distributions when calculating the *L1* distances (for details see Blackwell et al., 2009; Iacus et al., 2012). The multivariate *L1* distance is 0.9957. Although still rather large, the smaller multivariate *L1* here compared to Table A4 further demonstrates that the CEM algorithm improved the comparability of the treated and control groups.