Effective Regulation through Credible Cost–Benefit Analysis: The Opportunity Costs of Superfund

MICHAEL GREENSTONE

Effective Regulation through Credible Cost-Benefit Analysis: The Opportunity Costs of Superfund[1]

MICHAEL GREENSTONE

3M PROFESSOR OF ECONOMICS, MASSACHUSETTS INSTITUTE OF TECHNOLOGY

I. Introduction

American government, at every level, regulates a dizzyingly broad swath of social and economic life. Regulatory policy determines the drugs we can buy, the pollutants in the air we breathe and the water we drink, the speed we can drive, the materials builders use to construct our homes, the cars we buy, and so much more.

In making decisions about regulations, public officials must choose which areas of our lives merit government rules, as well as how stringent those rules should be. For example, the federal government first decided to regulate airborne particulate matter in 1970 and has tightened these regulations twice since then. Simultaneously, the government has had to decide whether and how to regulate hundreds of other air pollutants and other hazards. These choices have been further complicated by the fact the distributional impacts of some pollutants are spread unevenly across the population (e.g., they may differ by region, income, or race). At the same time, policy makers have had to grapple with the economic impacts of proposed environmental rules on manufacturers and other polluters. The essence of regulation is that it requires the regulated to take actions that they would not otherwise take, actions that often increase their costs, reduce their utility, or in some other way harm them.

When faced with this incredible array of complex and often uncertain trade-offs, what is a well-intentioned government to do?

The only humane solution to this enduring dilemma lies with careful and rigorous cost-benefit analysis. This approach's fundamental goal is to analyze regulatory decisions rationally and quantitatively, with the goal of maximizing societal welfare. Specifically, regulators should seek to maximize the expected net benefits of regulation, which is just the difference between the expected benefits (e.g. lives saved, illnesses prevented) and the costs (e.g., investments required to scrub smokestacks, expenditures on monitoring pollution emissions).

Cost-benefit analysis requires that regulators convert both the costs and the benefits of a proposed policy initiative into a common unit, money. Some critics of the practice consider such quantitative translation cold-hearted and impossible (Ackerman and Heinzerling 2004; Kelman 1981), but it is in

[1] I am grateful to Edward Balleisen, David Moss, and Katherine Ozment for insightful comments and encouragement. I thank the Tobin Project for financial support. This paper was prepared for the Tobin Project's Conference, "Toward a New Theory of Regulation."

fact the most humane approach to regulation that we have.[2] By converting all costs and benefits to the same unit, government can avoid setting irrational regulatory policies that harm human welfare. A failure to use cost-benefit analysis could lead to irrational policies across pollutants. For example, it might lead to strict regulation on airborne particulate matter that prevents all deaths due to this pollutant at a tremendous cost to business; while arsenic pollution is regulated very lightly such that hundreds of people die annually, even though tougher rules would require only a minor burden on firms. Cost-benefit analysis is a transparent method to help policy-makers determine which pollutants they should regulate and to what degree.

The current regulatory problem does not involve cost-benefit analysis per se, but rather the poor quality of the evidence underlying many cost-benefit decisions. At its core, regulatory policy aims to alter the world so that the lives of at least some members of society are improved. But in so doing, regulations generally restrict or regulate the behavior of other members of society. The goal of a beneficent government is to implement regulations where the benefits outweigh the costs, accounting for costs and benefits to all members of society.

But without a well developed strategy for evaluating regulation policies, it is impossible to know what would have happened in the absence of the policy. The fundamental issue is that we would like to observe the world with and without a regulation. Then, we could determine whether society gains or is harmed by a regulatory policy. Of course, it's impossible to observe both states of the world simultaneously.

What can be done to improve the quality of evidence on regulations' causal impact on social welfare? The ideal solution is to use the same experimental techniques that are used in "hard sciences" such as chemistry and medicine. Classical experimental design incorporates the random assignment of the population into a treatment group, those affected by a given treatment, and a control group, those who are not. The random assignment of the treatment means that there is no a priori reason to believe that in the absence of the regulation the average behavior of the two groups would have differed. Thus, a comparison of outcomes among the treatment and control populations yields a causal estimate of the treatment.

Although the random assignment of regulation may seem like a radical idea, it should not. Randomized trials are the primary tool to learn about the efficacy of drugs and medical devices. Further, there is increasing acceptance of this approach in many areas of social science, including educational policy and interventions in developing countries.[3]

---

[2] There are several standard criticisms of cost-benefit analysis. They include that it immorally commodifies objects (e.g., human life) that are beyond valuation, gives a false sense of scientific certainty, and unfairly benefit the rich. Several commentators, including Revesz and Livermore (2007) and Sunstein (2004), provide powerful responses to these criticisms. Interestingly, these criticisms have done little to dislodge cost-benefit analysis as the major lens through which government agencies and legislators assess regulations.

[3] See Ioannidis et al. (2001) on the central role of randomized experiments in medicine and epidemiology. Angrist (2004) describes a sea change in research on education policy that has led to a growing consensus that randomized trials are the only way to determine the causal effect of alternative educational interventions. Notably, the US Department of Education has enfranchised this view with the creation of the Institute of Education Science, whose mission is to "provide rigorous evidence on which to ground education practice and policy" (U.S. Department of Education 2006). See Duflo, Glennerster, and Kremer (2007) for a discussion of the increasing use of randomized experiments in assessments of interventions in developing countries.

Despite their evident appeal, randomized evaluations of regulations may not be possible in some instances. In these cases, quasi-experiments and natural experiments provide an appealing alternative approach. In analyzing regulations with quasi-experiments, one measures the differences in outcomes between the treatment and control groups just as in a classical experiment. In these cases, however, "treatment" (or in a regulatory context, "policy") status is determined by politics, public pressure, or some other action beyond the researcher's control. Despite the non-random nature of treatment status, it is still possible to draw valid inferences on the effects of regulation from the differences in outcomes between the treatment and control groups, provided that the quasi-experiment meets certain, potentially testable, assumptions.

Another important feature of evidence-based regulation is to use economic theory to guide the experiment or quasi-experiment. Theory provides the framework for identifying the people and firms that may be affected by a given regulation. Consequently, it is crucial to use economic theory to structure the empirical analysis so that the results can be used to determine the impact of the regulation on societal welfare. Further, economic theory can help to assess the generality of any findings (e.g., whether the findings are likely to apply in other settings).

A government that fails to rely on credible cost-benefit analyses is rolling the dice with its citizens' welfare since implementing regulations whose impacts are unknown is often equivalent to placing bets of tens of billions of dollars and unknown numbers of human lives. Poorly informed regulatory choices can lead to a nation's citizens being exposed to lethal concentrations of pollutants. Or conversely, the imposition of regulations with little benefit can burden citizens and firms with expensive compliance efforts that reduce incomes and the quality of citizens' lives. Indeed, regulations' costs can even shorten individuals' lives as income is an important determinant of longevity. As the costs of regulation in the US amount to many hundreds of billions of dollars, calling the stakes high is an understatement.

The remainder of this essay conducts an abridged cost-benefit analysis based on research by Greenstone and Gallagher (2008) and Currie, Moretti and Greenstone (2007), of the federal Superfund program that clean-ups hazardous waste sites. Through 2005, the federal government has spent approximately $35 billion (2005$) on Superfund clean-ups, and firms have expended considerable additional funds. The program continues to grow, with remediations ongoing at roughly 800 sites and regulators continually adding new sites to the list of those slated for clean up.

This essay focuses on Superfund for several reasons. First, it demonstrates that it is possible to conduct a credible empirical analysis in a setting where it might not have seemed feasible. Second, the cost-benefit analysis is guided by economic theory so the connection between the results and social welfare is immediate (at least in principle). In short, this abridged cost-benefit exercise demonstrates that opportunities for sophisticated, credible cost-benefit analyses are more readily available than is widely believed.

Third, the analysis of the Superfund program suggests that its benefits to the people living near these hazardous waste sites are likely to be smaller than its costs. This finding holds whether one allows for the possibility that the benefits are evident in housing prices and/or infant health. This is an uncomfortable finding, akin to conclusions that people do not like ice cream or sunshine. However, it helps to underscore that, while cost-benefit analyses may not fit our prior expectations, they can lead to an improvement in social welfare by directing resources to the projects that produce the largest social benefits.

The remainder of the paper examines the costs and benefits of the Superfund program. It then concludes with a brief discussion of why the best and most humane path forward for regulation is to implement a culture of experimentation and evaluation and provides some directions on how to jump-start such a culture.

II. An Abridged Cost-Benefit Analysis of the Superfund Program

This section of the paper is divided into five subsections. The first provides a history of the Superfund program and the outline of the research design that, along with my collaborator Justin Gallagher, I have used to test for impacts of clean-ups on house prices (see Greenstone and Gallagher 2008). The second briefly outlines an economic model that guides the empirical analysis. The third derives estimates of the costs of Superfund clean-ups and provides some other statistics about them. The fourth subsection derives estimates of the benefits of Superfund clean-ups as measured in the housing market. In principle, the full benefits (i.e., aesthetic and health) of clean-ups will be capitalized into housing prices. The fifth subsection estimates the impacts of Superfund clean-ups on infant health. While this is an incomplete measure of the potential health benefits, fetuses and infants are a population that is likely to be especially sensitive to exposure to the contaminants found at Superfund sites.

A. The Superfund Program and a New Research Design

1. History and Broad Program Goals

Before the regulation of the disposal of hazardous wastes by the Toxic Substances Control and Resource Conservation and Recovery Acts of 1976, industrial firms frequently disposed of wastes by burying them in the ground. Love Canal, New York offers perhaps the most infamous example of these disposal practices. Throughout the 1940s and 1950s, this area served as a landfill for industrial waste, receiving more than 21,000 tons of chemical wastes. After New York state investigators found high concentrations of dangerous chemicals in the air and soil at Love Canal, concerns about the safety of this area prompted President Carter to declare a state of emergency in 1978, an action that led to the relocation of the area's 900 residents. As David Moss and Mary Oey make clear in their essay on the impact of the Love Canal crisis on regulatory politics (Moss and Oey 2009), this incident helped to galvanize support for addressing the legacy of industrial waste, a movement that culminated in the creation of the Superfund program in 1980.

The centerpiece of the Superfund program, and this paper's focus, is the long-run remediation of hazardous waste sites.[4] These multi-year remediation efforts aim to reduce permanently the serious but not imminently life-threatening dangers caused by hazardous substances. By the end of 2005, the Environmental Protection Agency had placed 1,552 sites on the National Priorities List (NPL), thereby slating them for long-run clean-ups. The next subsection describes the selection process, which forms the basis of our research design.

2. Site Assessment & Superfund Clean-Ups Processes

---

[4] The Superfund program also funds immediate removals, which are short-term responses to environmental emergencies aimed at diminishing an immediate threat. These actions are not intended to remediate the underlying environmental problem and are not exclusive to hazardous waste sites on the NPL.

As of 1996, environmental activities, neighborhood groups, and other interested parties had referred more than 40,000 hazardous waste sites to the EPA for possible inclusion on the NPL. Since there are limited resources available for these clean-ups, the EPA follows a multi-step process to identify the most dangerous sites.

The final step of the assessment process involves the application of a Hazardous Ranking System (HRS), a rating system reserved for the most dangerous sites. The EPA developed the HRS in 1982 as a standardized approach to identify the sites that pose the greatest threat to humans and the environment. The original HRS evaluated the risk for exposure to chemical pollutants along three migration 'pathways': groundwater, surface water, and air. The major determinants of risk along each pathway for a site are the toxicity and concentration of chemicals present, the likelihood of exposure and proximity to humans, and the size of the potentially affected population. EPA officials also consider non-human impacts, but they play a relatively minor role in determining the HRS score.

The HRS produces a score that ranges from 0 to 100, with 100 being the highest level of risk. From 1982-1995, the EPA assigned all hazardous waste sites with a HRS score of 28.5 or greater to the NPL. Only these sites become eligible for Superfund remedial clean-up. The Data Appendix provides further details on the determination of HRS test scores and their role in assignment to the NPL.

Once a site moves onto on the NPL, it generally takes many years until clean-up firms complete their work. The first step is a further study of the extent of the environmental problem and how best to remedy it, an assessment that regulators summarize in the Record of Decision (ROD), which also outlines recommended clean-up actions for the site. After workers finish physical construction of all clean-up remedies, removing immediate threats to health, and putting long-run threats "under control," the EPA gives a site a "construction complete" designation. The final step is the agency's deletion of the site from the NPL.

3. 1982 HRS Scores as the Basis of a New Research Design

This paper's goal is to obtain reliable estimates of the effect of Superfund sponsored clean-ups of hazardous waste sites on housing market outcomes in areas surrounding the sites. The empirical challenge is that NPL sites are the most polluted in the US, so it is likely that there are unobserved factors that covary with both proximity to hazardous waste sites and housing prices. Although this possibility cannot be tested directly, it is notable that proximity to a hazardous waste site is associated with lower population densities, lower household incomes, higher percentages of high school dropouts, and a higher fraction of mobile homes among the housing stock.

Consequently, cross-sectional estimates of the association between housing prices and proximity to a hazardous waste site may be severely biased due to omitted variables.[5] In fact, economists have

[5] Cross-sectional models for housing prices have exhibited signs of misspecification in a number of other settings, including the relationships between land prices and school quality, air pollution, and climate variables (Black 1999; Chay and Greenstone 2005; Deschenes and Greenstone 2006). Incorrect choice of functional form is an alternative source of misspecification (Halvorsen and Pollakowski 1981; Cropper et al. 1988). Other potential sources of biases of published hedonic estimates include measurement error and publication bias (Black and Kneisner 2003; Ashenfelter and Greenstone 2004).

long noted that the possibility of confounding due to unobserved variables is a threat to the validity of the results from efforts to develop reliable estimates of individuals' willingness to pay for environmental amenities (Small 1975). This paper's challenge is to develop a valid counterfactual for the housing market outcomes near Superfund sites in the absence of their placement on the NPL and clean-up.

A feature of the initial NPL assignment process that has not been noted previously by researchers may provide a credible solution to the likely omitted variables problem. In the first year after the legislation's passage, groups and individuals referred 14,697 sites to the EPA, which then investigated them as potential candidates for remedial action. Through an initial assessment process, the EPA winnowed this list to the 690 most dangerous sites. Although the Superfund legislation directed the EPA to develop a NPL of "at least" 400 sites (Section 105(8)(B) of CERCLA), budgetary considerations caused the EPA to set a goal of placing exactly 400 sites on the NPL.

The EPA developed the HRS to provide a scientific basis for determining the 400 out of the 690 sites that posed the greatest risk. Pressured to initiate the clean-ups quickly, the EPA developed the HRS in about a year, applied the test to the 690 worst sites, and ranked their scores from highest to lowest. A score of 28.5 divided numbers 400 and 401, so the initial NPL published in September 1983 was limited to sites with HRS scores exceeding 28.5. See the Data Appendix for further details.

The central role of the HRS score provides a compelling basis for a research design that compares housing market outcomes near sites with initial scores above and below the 28.5 cut-off for at least three reasons. First, it is unlikely that sites' HRS scores were manipulated to affect their placement on the NPL, because the 28.5 threshold was established <u>after</u> the testing of the 690 sites was completed. The HRS scores therefore reflected the EPA's assessment of the risks posed by each site <u>rather</u> than the expected costs or benefits of clean-up.

Second, the HRS scores are noisy measures of risk, so it is possible that true risks are similar above and below the threshold. This noisiness results from the scientific uncertainty about the health consequences of exposure to the tens of thousands of chemicals present at these sites.[6] Further, there was no evidence that sites with HRS scores below 28.5 posed little risk to health. The Federal Register specifically reported that the "EPA has not made a determination that sites scoring less than 28.50 do not present a significant risk to human health, welfare, or the environment" and that a more informative test would require "greater time and funds" (U.S. Environmental Protection Agency 1984).[7]

[6] A recent history of Superfund makes this point. "At the inception of EPA's Superfund program, there was much to be learned about industrial wastes and their potential for causing public health problems. Before this problem could be addressed on the program level, the types of wastes most often found at sites needed to be determined, and their health effects studied. Identifying and quantifying risks to health and the environment for the extremely broad range of conditions, chemicals, and threats at uncontrolled hazardous wastes sites posed formidable problems. Many of these problems stemmed from the lack of information concerning the toxicities of the over 65,000 different industrial chemicals listed as having been in commercial production since 1945" (U.S. Environmental Protection Agency 2000, p. 3-2).

[7] One way to measure the crude nature of the initial HRS test is by the detail of the guidelines used for determining the HRS score. The guidelines used to develop the initial HRS sites were collected in a 30 page manual. Today, the analogous manual is more than 500 pages.

Third, the selection rule that determined placement on the NPL is a highly nonlinear function of the HRS score. This allows for a quasi-experimental regression discontinuity design. Specifically, we will compare outcomes at sites "near" the 28.5 cut-off. If the unobservables are similar or changing smoothly around the regulatory threshold, then one can make causal inferences on the impact of Superfund clean-ups on housing markets.[8]

An additional key feature of this data set is that an initial score above 28.5 is highly correlated with eventual NPL status but is not a perfect predictor of it. The EPA rescored some sites, with the later scores determining whether they ended up on the NPL.[9] The subsequent analysis uses an indicator variable for whether a site's initial (i.e., 1982) HRS score was above 28.5 as an instrumental variable for whether a site was on the NPL to purge the potentially endogenous variation in NPL status.

Finally, it important to emphasize that sites that failed to qualify for the NPL were ineligible for Superfund remediations. My collaborators and I investigated whether these sites were cleaned-up under state or local programs and found that they were frequently left untouched. Among the sites that the EPA targeted through these programs, a typical solution was to put a fence around the site and place signs indicating the presence of health hazards. The point is that the remediation activities at NPL sites dramatically exceeded the clean-up activities at non-NPL sites in scope and cost.

B. Economic Theory as a Guide

As an alternative to the health effects approach, we use the housing market to infer individuals' valuations of clean-ups. Economists have estimated the association between housing prices and environmental amenities at least since Ridker (1967) and Ridker and Henning (1967). However, Rosen (1974) and Freeman (1974) were the first to give this correlation an economic interpretation. In the Rosen formulation, a differentiated good can be described by a vector of its characteristics, $Q = (q_1, q_2,\ldots, q_n)$. In the case of a house, these characteristics may include structural attributes (e.g., number of bedrooms), neighborhood public services (e.g., local school quality), and local environmental amenities (e.g., distance from a hazardous waste site). Thus, the price of the i^{th} house can be written as:

(1) $P_i = P(q_1, q_2,\ldots, q_n)$.

The partial derivative of $P(\bullet)$ with respect to the n^{th} characteristic, $\partial P/\partial q_n$, is referred to as the marginal implicit price. It is the marginal price of the n^{th} characteristic implicit in the overall price of the house.

Locations close to hazardous waste sites must have lower housing prices to attract potential homeowners, so $\partial P/\partial q_n$ reveals the price that allocates individuals across locations. Thus, it can be

[8] The research design of comparing sites with HRS scores "near" the 28.5 is unlikely to be valid for sites that received an initial HRS score after 1982. This is because once the 28.5 cut-off was set, the HRS testers were encouraged to minimize testing costs and simply determine whether a site exceeded the threshold. Consequently, testers generally stop scoring pathways once enough pathways are scored to produce a score above the threshold.

[9] As an example, 144 sites with initial scores above 28.5 were rescored and this led to 7 sites receiving revised scores below the cut-off. Further, complaints by citizens and others led to rescoring at a number of sites below the cut-off. Although there has been substantial research on the question of which sites on the NPL are cleaned-up first (see, e.g., Sigman 2001), we are unaware of any research on the determinants of a site being rescored.

used to infer the welfare effects of a marginal change in a characteristic.[10] In principle, the price differential reflects both individuals' valuations of the health risk associated with proximity to a site and the site's damage to a neighborhood's aesthetics. In this respect, the use of housing markets to value an amenity provides a fuller examination of the valuation than an exclusive focus on the health risks.[11]

The consistent estimation of (1) is the foundation for accurate welfare calculations of both marginal and non-marginal changes. However, consistent estimation may be difficult since it is likely that there are unobserved factors that covary with, for example, both distance from a hazardous waste site and housing prices.[12] Although this possibility cannot be directly tested, it is notable that proximity to a hazardous waste site is associated with a number of important observable predictors of housing prices. For example, areas with hazardous waste sites tend to have lower population densities and a higher proportion of mobile homes, and are more likely to be in the Northeast.

Consequently, cross-sectional estimates of the association between housing prices and proximity to a hazardous waste site may be severely biased due to omitted variables. In fact, the cross-sectional estimation of equation (1) has exhibited signs of misspecification in a number of other settings, including the relationships between land prices and school quality, total suspended particulates air pollution, and climate variables (Black 1999; Chay and Greenstone 2005; Deschenes and Greenstone 2007).[13] Small (1975) recognized the consequences of the misspecification of equation (1) just one year after publication of the Rosen and Freeman papers:

> "I have entirely avoided…the important question of whether the empirical difficulties, especially correlation between pollution and unmeasured neighborhood characteristics, are so overwhelming as to render the entire method useless. I hope that…future work can proceed to solving these practical problems….The degree of attention devoted to this [problem]…is what will really determine whether the method stands or falls…" [p. 107].

In the intervening years, this problem of misspecification has received little attention from empirical researchers, even though Rosen himself recognized it.[14]

A key assumption underlying the use of housing markets to value proximity to a Superfund site is that the individuals living near a site are aware when clean-ups have occurred. If this assumption is

[10] See Rosen (1974), Freeman (1993), and Palmquist (1991) for fuller explanations of the hedonic method and in particular that P(•) represents the equilibrium interactions of consumers and producers. Further, they describe the necessary conditions to use the hedonic method to recover individuals' demand functions, which allow for the valuation of nonmarginal or large change in the amenity. Also see Ekeland, Heckman, and Nesheim (2004).

[11] Generally, the hedonic approach cannot account for aesthetic benefits that accrue to nonresidents that, for example, engage in recreational activities near the site. The health effects approach has this same limitation.

[12] Additionally, the estimation of equation (1) may be misspecified due to incorrect choice of functional form for observed covariates (Halvorsen and Pollakowski 1981; Cropper et al. 1988).

[13] Similar problems arise when estimating compensating wage differentials for job characteristics, such as the risk of injury or death. The regression-adjusted association between wages and many job amenities is weak and often has a counterintuitive sign (Smith 1979; Black and Kneisner 2003).

[14] Rosen (1986) wrote, "It is clear that nothing can be learned about the structure of preferences in a single cross-section…" (p. 658), and "On the empirical side of these questions, the greatest potential for further progress rests in developing more suitable sources of data on the nature of selection and matching…" (p. 688).

invalid, it is possible that there are substantial benefits to Superfund clean-ups that are not reflected in housing prices. Rather than blindly accept this assumption, this paper will also report on tests of whether Superfund clean-ups led to improvements in measures of infant health. A finding of substantial health benefits but little increase in housing prices might still lead to Superfund's benefits exceeding its costs.

C. Costs and Other Background Information on Superfund Clean-Ups

The housing price analysis that follows emerged from two samples of hazardous waste sites. The first is called the "All NPL Sample" and includes the 1,398 hazardous waste sites in the 50 US states and the District of Columbia that were placed on the NPL by January 1, 2000. The second is the "1982 HRS Sample," comprising the 690 hazardous waste sites tested for inclusion on the initial NPL. As I will explain below, the infant health component of the analysis is based on Superfund sites in Michigan and Pennsylvania.

Table 1 presents summary statistics on the hazardous waste sites in these samples. The entries in column (1) are from the All NPL Sample and are limited to sites in a census tract for which there is non-missing housing price data in 1980, 1990, and 2000. After these sample restrictions, there are 985 sites -- more than 70% of the sites ever placed on the NPL by 2000. Columns (2) and (3) report data from the 1982 HRS Sample. The column (2) entries are based on the 487 sites located in a census tract with complete housing price data. Column (3) reports on the remaining 189 sites located in census tracts with incomplete housing price data (generally due to missing 1980 data). 14 sites are located outside of the continental United States and were dropped from the sample.

Panel A reports on the timing of the sites' placement on the NPL. Column (1) reveals that about 75% of all NPL sites received this designation in the 1980s. Together, columns (2) and (3) demonstrate that the EPA eventually placed 443 of the 676 sites in the 1982 HRS Sample on the NPL. This number exceeds the 400 sites that Congress set as an explicit goal, because, as we have discussed, the agency rescored some sites with initial scores below 28.5, resulting in new scores above the threshold, thus qualifying them for the NPL.

Panel B reports on the size of the hazardous waste sites measured in acres, which is available for NPL sites only. The median site size ranges between 25 and 35 acres across the samples. The means are substantially larger due to a few very large sites. The modest size of most sites suggests that any expected effects on property values might well be confined to relatively small geographic areas around the sites.

Panel C reveals that the clean-up process is slow. We report the median time until the achievement of different milestones, rather than the mean, because many sites have not reached all of the milestones yet. 198 (16) of the NPL sites in column (2) received either the construction complete or deleted designation by 2000 (1990). For this reason, we focus on changes in housing prices, rental rates, and quantities between 1980 and 2000.

Panel D reports the expected costs of clean-up for NPL sites, and E details expected and actual costs among sites that are construction complete or deleted. The EPA estimates the expected costs before any remediation activities have begun, while actual costs are our best estimates of total remediation related expenditures assessed after the site achieves the "construction complete" designation. We believe this is the first time these variables have been reported for the same sites. In the 1982 HRS Sample that we focus on (i.e., column (2)), the mean and median expected costs are $27.5 million and $15.0 million.

Among the construction complete sites in the 1982 HRS Sample, the mean actual costs exceed the expected costs by about 55%. We multiply the overall mean expected cost of $27.5 million by 1.55 to obtain an estimate of the mean actual costs of clean-up in the 1982 HRS Sample of $43 million (the analogous figure in the All NPL sample is $39 million). This estimate of costs understates the true costs, because it does not include the legal costs or deadweight loss associated with the collection of funds from private parties or taxes, nor does it include the site's share of the EPA's costs of administering Superfund. Nevertheless, it is contrasted with the estimated benefits of Superfund clean-ups in the remainder of the paper.[15]

Figure 1 displays the geographic distribution of the 985 hazardous waste sites with complete housing data in the All NPL Sample. There are NPL sites in 45 of the 48 continental states, demonstrating that Superfund is genuinely a national program. The highest proportion of sites is in the Northeast and Midwest (i.e., the "Rust Belt"), reflecting the historical concentration of heavy industry in these regions.

Figure 2 presents a histogram of the initial HRS scores, so it depicts the number of sites with HRS scores in relatively small ranges. The ranges or bins are 4 HRS points wide, because the EPA considered HRS scores within 4 points to be statistically indistinguishable and reflect comparable risks to human health (U.S. Environmental Protection Agency 1991). The distribution looks approximately normal, with the modal bin covering the 36.5-40.5 range. Further, there is no obvious bunching just above or below the threshold, which supports the scientific validity of the HRS scores and suggests that they were not manipulated. Importantly, 227 sites have HRS scores between 16.5 and 40.5. This set is centered on the regulatory threshold of 28.5 that determines placement on the NPL and constitutes the regression discontinuity sample that we exploit in the subsequent analysis.

D. The Impact of Superfund Clean-Ups on Housing Prices

This subsection examines the benefits of Superfund clean-ups as measured through the housing market. It begins by reviewing the econometric or statistical approach, then presents the results, and finally interprets the findings.

1. Econometric Approach

The goal of the empirical exercise is to measure the impact of Superfund clean-ups on the prices of homes located near the remediated hazardous waste sites. There are two key features of the exercise. First, its basis is an examination of the growth of house prices in these areas between 1980 and 2000 using decennial Census price data. The beginning year (1980) is the starting point because it precedes the start of the Superfund program. 2000 is useful as an ending point, because

[15] The similarity of the column (1) sites with the other sites suggests that it may be reasonable to assume that the results from the application of the HRS research design to the 1982 HRS Sample are informative about the effects of the other Superfund clean-ups.

remediation was complete at nearly 65% of the sites originally placed on the NPL by then.

Second, the chief threat is the possibility of unobserved variables that affect the growth of housing prices between 1980 and 2000 near these sites. For example, the fraction of the population living in cities where many Superfund sites are located increased during the 1990s (presumably for reasons unrelated to Superfund). Plus, proximity to a hazardous waste site is associated with lower household incomes, higher percentages of high school dropouts, and a higher fraction of mobile homes among the housing stock. Thus, valid inference on the impact of Superfund requires the identification of an empirical strategy that avoids confounding the direct impact of the clean-ups with these types of unobserved variables.

Greenstone and Gallagher's (2008) potentially valid solution is to restrict the analysis to the neighborhoods around the 690 hazardous waste sites that the EPA deemed to be the most dangerous in the US when deciding which 400 sites to place on the initial NPL. The basic idea is to compare the growth of housing prices near the 400 hazardous waste sites with initial HRS scores exceeding 28.5 that qualified for a Superfund clean-up to housing price growth near the 290 sites with HRS scores below 28.5 that narrowly missed placement on the NPL. The necessary assumption for valid inference is that in the absence of the Superfund clean-ups, the growth in house prices would have been equal in the areas near sites with 1982 HRS scores above and below the 28.5 cutoff for clean-ups.

I now briefly describe the technical details involved in implementing what at its core is simply a comparison of housing price growth near sites with HRS scores above and below the 28.5 cutoff. The basis of the econometric approach is a two-stage least squares (2SLS) strategy that accounts for the possibility that some sites received a second score if it was thought that the initial score was too high or low. Since this decision about rescoring might be related to future house price growth, we rely on the variation in NPL status based on the <u>initial</u> HRS score. Specifically, we fit the following system of equations:

(2) $$y_{c2000} = \theta\,(1NPL_{c2000}) + X_{c1980}'\beta + \varepsilon_{c2000},$$

(3) $$(1NPL_{c2000}) = X_{c1980}'\Pi + \delta\,1(HRS_{c1982} > 28.5) + \eta_{c2000},$$

where c references a census tract. The year (1980, 1982, or 2000) that the variable is measured is also denoted in the subscripts. In practice, the sample is limited to census tracts containing the 487 sites in the 1982 HRS Sample with housing price data in 1980 and 2000.

The indicator variable $1(NPL_{c2000})$ equals 1 if the observation is from a tract that contains a site placed on the NPL by 2000. The vector X_{c1980} includes a wide set of census tract-level variables available in the Census files. These are detailed in the Data Appendix. X_{c1980} also includes the natural log of the mean housing price in 1980. Consequently, the parameter of interest, θ, measures the growth in housing prices in census tracts with a NPL site, relative to census tracts with hazardous waste sites that narrowly avoided placement on the NPL (after adjustment for the X vector).

The indicator variable $1(HRS_{c1982} > 28.5)$ in equation (3) serves as an instrumental variable. It equals 1 for census tracts with a site that has a 1982 HRS score exceeding the 28.5 threshold. We then substitute the predicted value of $1(NPL_{c2000})$ from the estimation of equation (3) in the fitting of (2) to obtain an estimate of θ_{2SLS}. In this 2SLS framework, θ_{2SLS} is identified from the variation in NPL status that is due to a site having a 1982 HRS score exceeding 28.5.

For θ_{2SLS} to provide a consistent estimate of the HPS gradient, the instrumental variable must affect the probability of NPL listing without having a direct effect on housing prices. The next section will demonstrate that the first condition clearly holds. The second condition requires that the unobserved determinants of 2000 housing prices are orthogonal to the portion of the nonlinear function of the 1982 HRS score that is not explained by X_{c1980}. In the simplest case, the 2SLS estimator is consistent if $E[1(HRS_{c82} > 28.5)\ \varepsilon_{c2000}] = 0$.

More informally, the aim of this approach is to compare the growth of housing prices in tracts with NPL sites to tracts with hazardous waste sites that narrowly missed a Superfund clean-up. The instrumental variables strategy purges any bias associated with rescoring sites where clean-ups are expected to have large benefits.

We also exploit the regression discontinuity design implicit in the 1(•) function that determines NPL eligibility in three separate ways to obtain 2SLS estimates that allow for the possibility that $E[HRS_{c82} > 28.5)\ \varepsilon_{c2000}] \neq 0$ over the entire 1982 HRS Sample. This approach focuses the regression so that it compares housing price growth among tracts with NPL sites and with non-NPL hazardous waste sites when the sites have very similar HRS scores. Intuitively, the idea is to compare tracts with HRS scores of 28.6 to those with sites with scores of 28.4. Practically, we actually use a somewhat wider range of HRS scores to avoid small sample problems. But, the intuition is that this approach further refines the comparisons so that apples are being compared to apples (rather than to oranges).

In the first regression discontinuity approach, a quadratic in the 1982 HRS score is included in X_{c1980} to partial out any correlation between residual housing prices and the indicator for a 1982 HRS score exceeding 28.5. This approach relies on the plausible assumption that residual determinants of housing price growth do not change discontinuously at the regulatory threshold. The second regression discontinuity approach involves implementing our 2SLS estimator on the regression discontinuity sample of 227 sites with 1982 HRS scores between 16.5 and 40.5. Here, the identifying assumption is that all else is held equal in the "neighborhood" of the regulatory threshold (or that all tracts are apples in this range of HRS scores). More formally, it is $E[1(HRS_{c82} > 28.5)\ \varepsilon_{c2000} | 16.5 < 1982\ HRS < 40.5] = 0$.

Recall, the HRS score is a nonlinear function of the ground water, surface water, and air migration pathway scores. The third regression discontinuity method exploits knowledge of this function by including the individual pathway scores in the vector X_{c1980}. All three regression discontinuity approaches are demanding of the data, so the resulting estimates are less well determined than is ideal.

2. Results

I now turn to the preferred quasi-experimental approach, assessing the relationship between 1982 HRS scores and NPL status. Figure 3 plots the bivariate relation between the probability that a site was placed on the NPL by 2000 and its initial HRS score among the 487 sites in the 1982 HRS

Sample. The plots are done separately for sites above and below the 28.5 threshold and come from the estimation of nonparametric regressions that use Cleveland's (1979) tricube weighting function and a bandwidth of 0.5.[16] Thus, they represent a moving average of the probability of NPL status across 1982 HRS scores. The data points represent the mean probabilities in the same 4-unit intervals of the HRS score as in Figure 4.

The figure presents dramatic evidence that an initial HRS score above 28.5 is a strong predictor of NPL status. The EPA placed virtually all sites with initial scores greater than 28.5 on the NPL by 2000. Again, rescoring explains the nonzero probability of placement on the NPL by 2000 among sites with an initial score below 28.5. A statistical version of the figure reveals that a HRS score above 28.5 is associated with an 83% increase in the probability of placement on the NPL (Greenstone and Gallagher 2005). It is evident that there is a powerful relationship between HRS scores above 28.5 and NPL status.

Table 2 presents 2SLS estimates of the effect of NPL status on housing prices in 2000. In Panel A, the observations are from the census tracts containing the 487 hazardous waste sites in the 1982 HRS Sample. In Panel B, each observation is comprised of the average of all variables across tracts that share a border with these tracts. In Panels C and D, the sample is comprised of the land area within circles with radii of 2 and 3 miles that are centered at each site's longitude and latitude. The means of the 1980 values of the total housing stock in the four samples are $71, $525, $349, and $796 million, respectively. The exact covariates in each specification are noted in the row headings at the bottom of the table and are described in more detail in the Data Appendix.

The regression discontinuity approach is implemented by altering the column (2) specification in three different ways. In the column (3) specification, the 1982 HRS score and its square are added. In column (4), the separate pathway scores are included. Finally, the column (5) sample is the regression discontinuity sample that is comprised of the 227 sites with 1982 HRS scores between 16.5 and 40.5.

The Panel A results suggest that a site's placement on the NPL has little impact on the growth of property values in its own census tract, relative to tracts with sites that narrowly missed placement on the NPL. The point estimates indicate an increase in prices that ranges from 0.7% to 4.7%, but they all have associated t-statistics less than two. The regression discontinuity specifications in columns (3) through (5) may be the most credible, so it is notable that they produce the smallest point estimates (although they are also the least precise).

Panel B presents the adjacent tract results. The point estimates from the regression discontinuity estimators range between -0.6% and 0.1% and zero cannot be rejected at conventional levels for any of them. Thus, there is little evidence of meaningful gains in housing prices outside the site's own census tract.

Panels C and D summarize the total gain in housing prices associated with a site's placement on the NPL by using the 2- and 3-mile radius circle samples. They also report whether the clean-ups pass cost-benefit tests analogous to those in Table 3. The threshold housing price gains are 13.8% and 5.8%.

[16] The smoothed scatterplots are qualitatively similar with a rectangular weighting function (i.e., equal weighting) and alternative bandwidths.

The circle sample results provide further evidence that the NPL designation has little effect on housing prices. In the columns (3) – (5) specifications, only one of the 6 point estimates is positive. Further in the 8 specifications that adjust for characteristics, the null that the gain in housing prices exceeds the break-even threshold is rejected at conventional significance levels. Overall, these quasi-experimental estimates suggest that Superfund clean-ups fail to pass this cost-benefit test.

Figure 4 provides an opportunity to better understand the source of these regression results. It plots the nonparametric regressions of 2000 residual housing prices (after adjustment for the column (4) covariates) against the 1982 HRS score in the 2-mile radius sample.[17] The nonparametric regression is estimated separately below (dark line) and above (light line) the 28.5 threshold. The graph confirms that there is little association between 2000 residual housing prices and the 1982 HRS score. A comparison of the plots at the regulatory threshold is of especial interest in light of the large jump in the probability of placement on the NPL there. It is apparent that the moving averages from the left and right are virtually equal at the threshold.

3. Interpretation

These results have failed to find evidence that Superfund clean-ups increase social welfare substantially.[18] In light of the significant resources devoted to these clean-ups and the claims of large health benefits, this finding is surprising. This section reviews three possible explanations.

First, the individuals that choose to live near these sites before and after the clean-ups may have a low willingness to pay to avoid exposure to hazardous waste sites. In this case, society provides these individuals a good that they do not value highly. It is possible (and perhaps likely) that there are segments of the population with a high willingness to pay (WTP) to avoid exposure to hazardous waste sites. It may even be the case that the population average WTP is substantial. However, the policy relevant parameter is the WTP of the population that lives near these sites, and this is the parameter that the paper has estimated.[19]

Second, the sites with initial HRS scores less than 28.5 may have also received complete remediations under state or local land reclamation programs. In this case, a zero result is to be expected since both the above and below 28.5 sites would have received the same treatment. Along with Gallagher, I investigated this possibility by conducting an extensive search for information on remediation activities at these sites. From these investigations, we concluded that the clean-up activities were dramatically more ambitious and costly at sites with initial scores exceeding 28.5. For example, we were unable to find evidence of any remediation activities by 2000 at roughly 60%

---

[17] Figure 4 provides a qualitative graphical exploration of the regression results. The relationship between housing prices and 1982 HRS scores cannot be exactly inferred from this graph, because the HRS score has not been adjusted for the column (4) covariates. However, the meaningfulness of this graph is supported by the finding that the covariates are well balanced among sites with 1982 HRS scores above and below the regulatory threshold, especially near the regulatory threshold (Greenstone and Gallagher 2008).

[18] Greenstone and Gallagher (2008) also find little impact of the clean-ups on rental prices for rental units, new home construction, or migration into the area surrounding the site. Additionally, Greenstone and Gallagher test for whether the absence of substantial price increases reflects a stigma that remains even after Superfund clean-ups. In particular, they found that a site's placement on the NPL has little immediate impact on housing rental prices in areas near the sites. This suggests that the Superfund designation fails to stigmatize these neighborhoods.

[19] A popular theory is that sites become permanently stigmatized when they are placed on the NPL. As Greenstone and Gallagher (2008) demonstrate, the data contradict this explanation in important ways.

of the sites with scores below 28.5. Further, among the 40% of the sites where there was evidence of clean-up efforts, the average expenditure was roughly $3 million. This is about $40 million less than the estimate of the average cost of a Superfund clean-up. This difference is not surprising, because the state and local clean-ups were often limited to restricting access to the site or containing the toxics, rather than trying to achieve Superfund's goal of returning the site to its "natural state." Nevertheless, some remediation took place at these sites, so it may be appropriate to interpret the results as the impact of the extra $40 million that a Superfund clean-up costs.

Third, there could be substantial health benefits from the clean-ups but the local residents may be unaware of them. Although there is usually substantial newspaper coverage about Superfund clean-ups, this nevertheless remains a possibility, striking at the housing market approach's key assumption that people have perfect information. If this assumption is invalid, then it is necessary to use alternative methods to identify the benefits of Superfund clean-ups.

E. The Impact of Superfund Clean-Ups on Infant Health

This subsection examines the impacts of Superfund clean-ups on infant health. It is part of a larger project on this topic that I am undertaking with my colleagues Janet Currie and Enrico Moretti (Currie, Greenstone, and Moretti 2008).

1. Why Focus on Infants?

It is possible that Superfund clean-ups affect many dimensions of human health. A thorough investigation of this issue would require an individual-level data file with detailed information on the respondents' health status (including mortality) and the location of their residences throughout their lifetimes. Such a data file would allow for an examination of whether exposure to Superfund sites leads to poor health outcomes. For example, it would be possible to test whether living near a Superfund site for an extended period of time increases the probability of being stricken by cancer. This type of data file does not exist in the US. Some available data files report individuals' current residence, but with the high degree of mobility in the US the assumption that individuals have never moved is unappealing.

It is possible to match infants to particular locations through their mother's place of residence on birth certificate forms. In practice, the analysis utilizes a data file constructed by merging data on the location of Superfund sites, the progression of their cleans-ups, and detailed infant health records from the universe of births in Pennsylvania and Michigan for the years 1989-2003. Importantly, the infant health information contains the street address of the mother's residence so it is possible to focus on births that occur within 1 mile of Superfund sites. The infant health data were obtained by reaching agreements with these two states' Departments of Vital Statistics to gain access to the confidential versions of these data files.

A key advantage of focusing on infants is that it seems reasonable to presume that their mother's residence is their place of residence during the entire fetal period and the first year of their lives. Further the fetal and infant periods are especially vulnerable ones, so the results from tests of exposure to Superfund sites on infant health may be informative about the possibility of health impacts in the broader population.

2. Econometric Approach

The goal of this empirical exercise is to measure the impact of Superfund clean-ups on a series of health outcomes for infants born near these sites. Since data is only available for Pennsylvania and Michigan, it is not feasible to implement the research design based on 1982 HRS scores. As an alternative, Currie, Greenstone, and Moretti (2008) implement an econometric approach that at its core compares birth outcomes before clean-ups with those after clean-ups have been completed. Since the most dangerous toxics may be removed during the clean-up process (i.e., not just at the end), the analysis also compares birth outcomes from before the remediation's initiation to those that occurred during the clean-up process. This section briefly summarizes the Currie, Greenstone, and Moretti approach and results for Pennsylvania and Michigan.

I now briefly describe the technical details of the statistical model. Specifically, the analysis is based on the fitting of the following equation for births that occur in Michigan and Pennsylvania:

$$\begin{aligned}
(4)\ \textit{Health Outcome}_{icst} = {} & \alpha_0 + \alpha_1\, 1(\text{Clean-Up Active at Nearest Site})_{st} \\
& + \alpha_2\, 1(\text{Clean-Up Completed at Nearest Site})_{st} \\
& + \alpha_3\, 1(\text{Nearest Site} < 1\text{ Mile})_{st} \\
& + \alpha_{4t} X_{icst} \\
& + \beta_1\, 1(\text{Clean-Up Active at Nearest Site})_{st} * 1(\text{Nearest Site} < 1\text{ Mile})_{st} \\
& + \beta_2\, 1(\text{Clean-Up Completed at Nearest Site})_{st} * 1(\text{Nearest Site} < 1\text{ Mile})_{st} \\
& + \mu_s + \theta_{ct} + \eta_m + \varepsilon_{icst}
\end{aligned}$$

where i represents an infant, c denotes the county of her mother's residence at the time of birth, s indicates the closest hazardous waste site, and t references the year. Additionally, m references her mother. The health outcomes or dependent variables are whether the infant dies within the first year of life, the presence of a congenital abnormality, whether the birth weight was less than 2,500 grams (i.e., classified as low birth weight), birth weight, whether the birth was premature, and the 1-minute APGAR score (a measure of the infant's health immediately after birth).

The indicator variable $1(\text{Clean-Up Active at Nearest Site})_{st}$ equals 1 for births where a clean-up has been initiated but has not been completed at the nearest hazardous waste site, regardless of the distance from the mother's home to the site. The indicator $1(\text{Clean-Up Completed at Nearest Site})_{st}$ equals 1 for births when the clean-up has been completed at the nearest hazardous waste site, again regardless of the distance. The third main effect, $1(\text{Nearest Site} < 1\text{ Mile})_{st}$, equals one for births to mothers that live within 1 mile of the nearest hazardous waste site.

The X vector includes a set of covariates describing the parents that may affect infant health. It includes indicator variables for their age, education, race, and Hispanic origin, all interacted with year indicators. When the mother's or father's information is missing, a new category is created for age, education, race, and Hispanic origin so that these observations are not dropped.

The two parameters of interest are β_1 and β_2. The first captures the variation in outcomes specific to births to mothers that live within 1 mile of a site during the period while the clean-up is ongoing. It measures whether birth outcomes among those living near a site improve during the clean-up, relative to before remediation was initiated. The second parameter of interest, β_2, tests for a mean difference in outcomes among births to mothers living within 1 mile of a site after remediation has been completed (again relative to the period before the clean-up was initiated). Thus, this model allows for the possibility that birth outcomes are differentially affected during the clean-up and after it is completed.

The richness of the data allow for the inclusion of a series of fixed effects that adjust for several forms of unobserved heterogeneity that might otherwise bias the estimates of β_1 and β_2. They include ones for the closest hazardous waste site (μ_s) and county by year (θ_{ct}).

Since the clean-ups can take many years to complete, it is possible that individual mothers gave birth in different stages of the clean-up. For example, clean-up could cause a reshuffling of people in the area near a Superfund site such that the composition of mothers has changed in a way that affects the outcomes. This type of behavioral response would undermine the validity of the analysis. The mother fixed effects, η_m, are an important way to adjust for this unobserved form of heterogeneity, because they ensure that the regression compares the birth outcomes of two children from the same mother where one occurs before the clean-up and the other occurs during remediation or after completion.

3. Results

Figures 5a and 5b graphically depict the locations of the hazardous waste sites in Michigan and Pennsylvania, respectively. Although the sites are concentrated in the most heavily industrialized parts of the state, there is substantial variation in their locations within these states.

Table 3 presents the results of the estimation of equation (4) for six different dependent variables and four separate specifications. The column (1) specification is the most parsimonious as it only includes site fixed effects and year fixed effects. Column (2) adds the mother and father characteristics, while column (3) then adds county by year fixed effects. Finally, the column (4) specification includes site, county by year, and mother fixed effects. This specification is very demanding of the data.

The entries are the estimates of the parameters β_1 and β_2 and below them their estimated standard errors in parentheses. To gain intuition, consider the panel on birth weight, which has a mean of 3,297 grams.[20] The estimates in the first row indicate that the weight of infants was 5.1 to 12.9 grams greater during the clean-up among nearby births, relative to before. These are small effects. Further, the null of a zero effect cannot be rejected at conventional levels for any of the specifications.

The next row reports on β_2, which measures the effect on birth weight in the period after the clean-up is completed, again relative to the period before the clean-up was initiated. The estimates of β_2 range from -5.9 to 38.3. Again, none of these would be judged to be statistically significant by conventional criteria. My conclusion is that there is little evidence that Superfund clean-ups led to increases in the birth weight of infants born within 1-mile of the sites.

The remaining panels provide an opportunity to assess the impact on other measures of infant health, including infant mortality. In general, there is little evidence that Superfund clean-ups have a

[20] Recent research indicates that lower birth weights are associated with negative long run outcomes, including educational attainment and wages (Black, Devereux, and Salvanes 2005).

meaningful impact on infant health outcomes. Some of them suggest improvements in these outcomes, while others indicate declines. However, the null of a zero effect cannot be rejected for any of the 20 estimates for the five remaining outcome variables. It is noteworthy that the mother fixed effect specifications frequently produce the largest point estimates, but they are quite poorly determined so their empirical content is not especially meaningful. Overall, these results fail to provide substantial evidence that Superfund clean-ups cause an improvement in the health of infants born within 1-mile of these sites.

F. Does Superfund Pass a Cost-Benefit Test?

This paper has provided the material necessary to conduct a cost-benefit analysis of Superfund clean-ups. On the cost side, the best estimate of the cost of current Superfund clean-ups is roughly $43 million.

As subsection B highlighted, an appeal of the housing market approach is that in principle it captures <u>all</u> of the benefits of Superfund clean-ups to local residents. For example, the value of any aesthetic improvements, as well as reductions in rates of morbidities and mortality should be reflected in housing prices. The largest estimated gain in housing prices comes from the 2-mile radius sample and the column (3) specification; this estimate implies that the value of the housing stock increased by roughly $7 million between 1980 and 2000. However, this estimate is statistically indistinguishable from zero and the null hypothesis that the gain in housing prices exceeds the costs of the clean-ups is rejected at conventional significance levels. The clear conclusion is that the benefits to local residents of Superfund clean-ups are substantially smaller than the costs.

As I have emphasized, the validity of this conclusion rests on the assumption that local residents are aware of the clean-ups and their health benefits. Rather than let the cost-benefit calculations rest on this unverifiable assumption, I present new evidence on whether Superfund clean-ups affect infant health. This exercise also fails to find that Superfund clean-ups led to meaningful benefits.

There are at least two important caveats to this health analysis. First, it has only been conducted on Superfund sites in the states of Michigan and Pennsylvania. These sites may be unrepresentative of Superfund sites more generally. Further, they are a subset of all Superfund sites and a larger sample is likely to improve the precision of the estimates. The next version of Currie, Greenstone and Moretti's paper will incorporate data from several other states. Second, although it seems plausible that infant health would be especially sensitive to the contamination at Superfund sites, it is possible that there other health outcomes are affected. For example, it is possible that exposure leads to developmental and learning disorders or chronic health conditions, such as cancer. The absence of large panel data sets with precise information on respondents' address poses potentially insurmountable challenges to fully exploring these possibilities.

Overall, the available evidence suggests that the benefits from Superfund clean-ups to the people living near these sites are small, at least relative to the costs of these clean-ups. To the extent that the aim of the policy is to improve the lives of individuals living near these sites, Superfund is not an effective policy; the most optimistic calculation suggests that it costs $1 to get 16 cents in the hands of the targeted group. The current version of Superfund appears to fail a cost-benefit test.

What are the policy implications of this finding? The above analysis cannot reject that there are positive benefits, so one solution is to scale back the extent of Superfund remediations. After all, Superfund remediations aim to return the site to its natural site and, in the US where land is so plentiful, this may not be very valuable. Thus, it may be cost effective to cap the areas where the pollution is concentrated and/or fence off the entire site or even just the part of it that is considered the most dangerous.

III. Implications

This essay's primary message is that humane regulatory policies require the persistent application of credible cost-benefit to assess the regulations that govern a dizzyingly broad swath of social and economic life. This is the best and only hope to develop a system of regulation that maximizes our well being.

There are four concrete steps that can be taken to achieve this goal. The first is that government must adopt a culture of experimentation in assessing regulations. In practice, this means that government should, whenever possible, implement proposed regulations on a small scale and undertake rigorous evaluations. This may mean allowing state or local governments to implement different policies for a specified period of time to infer the impacts of alternative forms of regulation; this has been carried out in several policy arenas already (e.g., welfare policy). In many instances, it is possible to implement genuine randomized trials. In other settings, the structure of the regulations can be used to evaluate them (as the above Superfund example illustrates). In other cases, it is only possible to evaluate a regulation ex post and in these cases the results can be used to strengthen, reform, or even remove the regulation. For example, Greenstone, Oyer, and Vissing-Jorgensen (2006) analyze the consequences of mandatory disclosure laws in US equity markets. The key theme is that quantitative evidence should trump qualitative evidence and rhetorical appeal whenever possible.

Critical to implementing a norm of experimental evaluation is the acceptance that not all studies are of equal value. The most compelling evidence comes from randomized trials. In cases where such evidence is unavailable, quasi-experimental evidence can be a good substitute.

Since in many instances, studies look similar to the untrained eye and results can be manipulated, governments must invest in attracting the best talent available to form review boards. This has been successfully accomplished in the case of drug trials. However, the same commitment to professional assessment is absent in the assessment of regulatory interventions. (My own experiences as an advisor to the EPA has left me certain there is room for improvement in environmental policy.) I find it inhumane to treat economic regulation differently from the regulation of drugs sold to the public.

On a related note, there may be instances where political and other considerations seem to trump the insights from credible cost-benefit analyses. My own personal view is that these instances are rarer than is widely believed. But even in these cases, it is still imperative to engage in credible cost-benefit evaluations. If politics is going to trump cost-benefit analysis, then politics' cost to society should be transparent.

The second step is to fund credible evaluations. Evaluations can be expensive, but their cost is generally small compared to the costs of implementing regulations that harm social welfare. The

Department of Education's recently established "Institute of Education Sciences" and the related "What Works Clearinghouse" demonstrate that scientific analysis can be used to improve our nation's schools. Indeed, education was an area that once seemed to be immune to credible evaluations but this appears to be rapidly changing.

The third step is to be forthright about cases where the evidence is unclear or of an insufficient quality to conduct a credible cost-benefit analysis. Since deciding not to take action is also making a choice, there may still be good reasons to enact a regulation in these cases. However, the absence of credible evidence should be noted and research funded so that future decisions in such areas are on firmer ground.

The fourth step is to recognize and become comfortable with the inevitability that cost-benefit analysis will lead to controversial implications in some instances. Indeed, I chose to focus on Superfund precisely because the results make me, and I suspect many others, uncomfortable. However, the power of credible cost-benefit analysis is that it provides a framework for making these tough judgments. After all, is it more humane to devote our scarce resources to cleaning up Superfund sites instead of devoting them to environmental problems where social payoffs are high? As just one example, my previous research has found large payoffs via higher housing prices and lower infant mortality rates from regulations that reduce total suspended particulate air pollution (Chay and Greenstone 2003a, 2003b, and 2005). The resources that are currently devoted to Superfund could be spent on the regulation of suspended particulate matter or regulations that mitigate other environmental problems.

There will always be decisions between regulations that *must* be beneficial or make us feel good and regulations that improve our lives. Credible cost-benefit analysis helps us make the right choices.

Table 1: Summary Statistics on the Superfund Program

	All NPL Sites w/ non-Missing House Price Data (1)	1982 HRS Sites w/ non-Missing House Price Data (2)	1982 HRS Sites w/ Missing House Price Data (3)
Number of Sites	985	487	189
1982 HRS Score Above 28.5	------	306	95
A. Timing of Placement on NPL			
Total	985	332	111
# 1981-1985	406	312	97
# 1986-1989	340	14	9
# 1990-1994	166	4	3
# 1995-1999	73	2	2
B. Size of Site (in acres)			
Number of sites with size data	920	310	97
Mean (Median)	1,187 (29)	334 (25)	10,507 (35)
Maximum	195,200	42,560	405,760
C. Stages of Clean-Up for NPL Sites			
Median Years from NPL Listing Until:			
ROD Issued	------	4.3	4.3
Clean-Up Initiated	------	5.8	6.8
Construction Complete	------	12.1	11.5
Deleted from NPL	------	12.8	12.5
D. Expected Costs of Remediation (Millions of 2000 $s)			
# Sites with Nonmissing Costs	753	293	95
Mean (Median)	$28.3 ($11.0)	$27.5 ($15.0)	$29.6 ($11.5)
95^{th} Percentile	$89.6	$95.3	$146.0
E. Actual and Expected Costs Conditional on Construction Complete (Millions of 2000 $s)			
Sites w/ Both Costs Nonmissing	477	203	69
Mean (Median) Expected Costs	$15.5 ($7.8)	$20.6 ($9.7)	$17.3 ($7.3)
Mean (Median) Actual Costs	$21.6 ($11.6)	$32.0 ($16.2)	$23.3 ($8.9)

Notes: All dollar figures are in 2000 $s. Column (1) includes information for sites placed on the NPL before 12/31/99. The estimated cost information is calculated as the sum across the first Record of Decisions for each operating unit associated with a site. See the Data Appendix for further details.

Table 2: Two-Stage Least Squares (2SLS) Estimates of the Effect of NPL Status on House Prices

	(1)	(2)	(3)	(4)	(5)
A. Own Census Tract					
1(NPL Status by 2000)	0.037	0.047	0.007	0.022	0.027
	(0.035)	(0.027)	(0.063)	(0.042)	(0.038)
B. Adjacent Census Tract					
1(NPL Status by 2000)	0.066	0.015	-0.006	-0.002	0.001
	(0.035)	(0.022)	(0.056)	(0.035)	(0.035)
C. 2 Mile Radius from Hazardous Waste Sites					
1(NPL Status by 2000)	0.019	0.001	0.023	-0.018	-0.007
	(0.032)	(0.023)	(0.054)	(0.035)	(0.034)
Ho: > 0.138, P-Value	0.000	0.000	0.018	0.000	0.000
D. 3 Mile Radius from Hazardous Waste Site					
1(NPL Status by 2000)	0.055	-0.004	-0.027	-0.024	-0.006
	(0.038)	(0.022)	(0.051)	(0.034)	(0.034)
Ho: > 0.058, P-Value	0.467	0.003	0.048	0.007	0.031
1980 Ln House Price	Yes	Yes	Yes	Yes	Yes
1980 Housing Characteristics	No	Yes	Yes	Yes	Yes
1980 Economic & Demographic Vars	No	Yes	Yes	Yes	Yes
State Fixed Effects	No	Yes	Yes	Yes	Yes
Quadratic in 1982 HRS Score	No	No	Yes	No	No
Control for Pathway Scores	No	No	No	Yes	No
Regression Discontinuity Sample	No	No	No	No	Yes

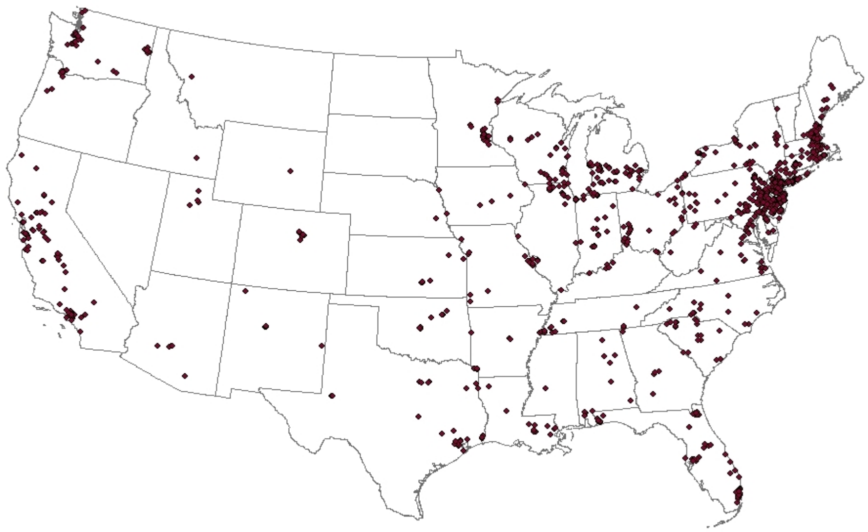
Notes: The entries report the results from 24 separate instrumental variables regressions. The ln (2000 median house price) is the dependent variable throughout the table. The units of observation are the census tract that contains the site (Panel A), tracts that share a border with the site (Panel B), the areas within a circle of 2 mile radius from the site (Panel C), and the areas within a circle of 3 mile radius from the site (Panel D). In Panels B-D where the unit of observation is comprised of multiple census tracts, the dependent and independent variables are calculated as weighted means across the relevant census tracts where the weight is the fraction of the tract that fits the Panel's sample selection rule multiplied by the tract's 1980 population. The variable of interest is an indicator for NPL status and this variable is instrumented with an indicator for whether the tract had a hazardous waste site with a 1982 HRS score exceeding 28.5. The entries are the regression coefficients and heteroskedastic consistent standard errors (in parentheses) associated with the NPL indicator. In Panel A (B-D) the samples sizes are 487 (483) in columns (1) through (4) and 227 (226) in column (5). Panels C and D also report p-values from tests of whether the NPL parameters multiplied by the value of the housing stock in 1980 exceeds $43 million, which is our best estimate of the cost of the average clean-up. The values of the housing stocks in 1980 in the four panels are roughly $75, $552, $311, and $736 million (2000 $s), respectively. See Greenstone and Gallagher (2008) for further details.

Table 3: Association between Superfund Clean-Ups and Birth Outcomes Births Within 1 Mile of NPL Site

	(1)	(2)	(3)	(4)
Probability of Infant Mortality (Mean = 0.0067)				
1(< 1 mile) * 1(After Clean-up Starts)	.0001	.0001	-.0004	-.0007
	(.0009)	(.0009)	(.0009)	(.0119)
1(< 1 mile) * 1(After Clean-up Ends)	-.0006	-.0006	-.0011	.0073
	(.0010)	(.0010)	(.0009)	(.0176)
Probability of Congenital Abnormality (Mean =0.0196)				
1(< 1 mile) * 1(After Clean-up Starts)	-.0024	-.0007	-.0001	-.0072
	(.0021)	(.0020)	(.0017)	(.0138)
1(< 1 mile) * 1(After Clean-up Ends)	-.0017	-.0015	.0001	.0038
	(.0015)	(.0015)	(.0019)	(.0196)
Probability of Low Birth Weight (Mean = 0.089)				
1(< 1 mile) * 1(After Clean-up Starts)	-.0030	-.0044	-.0001	-.0063
	(.0033)	(.0033)	(.0027)	(.0302)
1(< 1 mile) * 1(After Clean-up Ends)	.0026	.0012	.0063	.0030
	(.0029)	(.0027)	(.0027)	(.0384)
Birth Weight (Mean = 3,297)				
1(< 1 mile) * 1(After Clean-up Starts)	8.38	11.59	2.10	38.08
	(8.47)	(7.63)	(6.73)	(76.25)
1(< 1 mile) * 1(After Clean-up Ends)	-14.13	-13.68	-26.63	-41.11
	(8.47)	(8.03)	(7.99)	(74.36)
Probability of Premature Birth (Mean = 0.073)				
1(< 1 mile) * 1(After Clean-up Starts)	-.0008	-.0018	.0092	.0107
	(.0032)	(.0037)	(.0028)	(.0271)
1(< 1 mile) * 1(After Clean-up Ends)	-.0047	-.0041	.0056	.0132
	(.0038)	(.0039)	(.0040)	(.0398)
Apgar 1-Minute Score (8.4)				
1(< 1 mile) * 1(After Clean-up Starts)	.0462	.0610	-.0176	-.2344
	(.0547)	(.0553)	(.0546)	(.6812)
1(< 1 mile) * 1(After Clean-up Ends)	-.1139	-.1161	-.1553	.1205
	(.0885)	(.0880)	(.0698)	(.7403)
Site Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	No	No
Mother, Father Characteristics x Year	No	Yes	Yes	No
County by Year Fixed Effects	No	No	Yes	Yes
Mother Fixed Effects	No	No	No	Yes

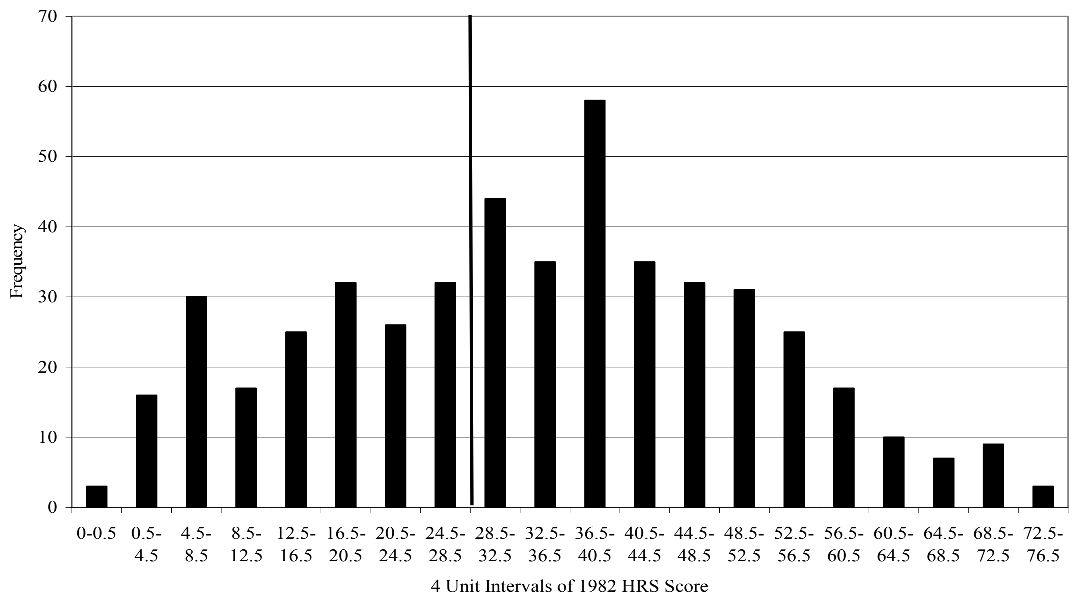
Notes: Standard errors clustered at the county-year level in parenthesis. The sample includes mothers 15 to 45 in PA and MI. Mother and father characteristics include dummies for age, education, race, Hispanic origin, all interacted with year. Observations for which the father information is missing are included in the analysis. When father is missing, a new category is created for father age, education, race and Hispanic origin. In columns 1 to 4, sample sizes for rows 1 to 6 are 122,471, 122,471, 122,063, 122,485, 122,485, and 121,793.

Figure 1: Geographic Distribution of NPL Hazardous Waste Sites in the All NPL Sample



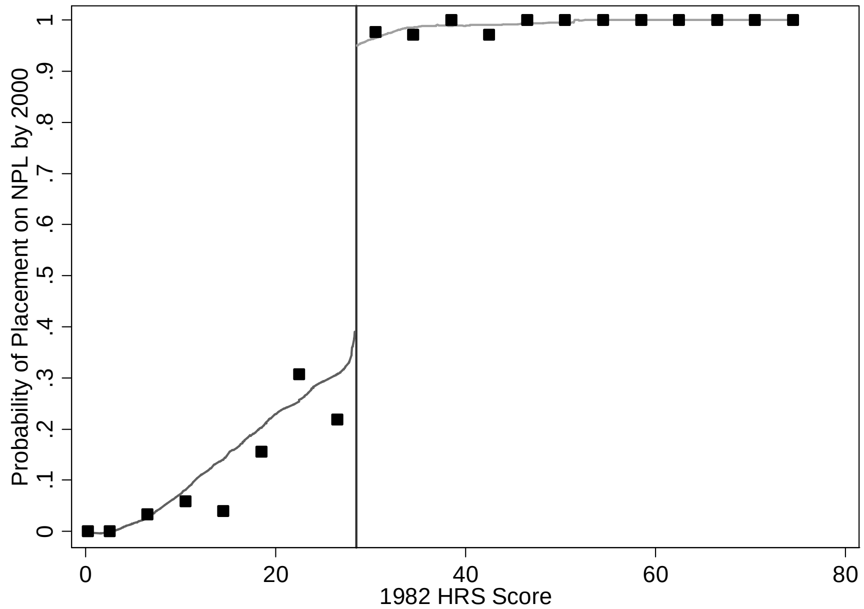
Notes: The All NPL sample is comprised of the 985 hazardous waste sites assigned to the NPL by January 1, 2000 that we placed in a census tract with nonmissing housing price data in 1980, 1990, and 2000.

Figure 2: Distribution of 1982 HRS Scores



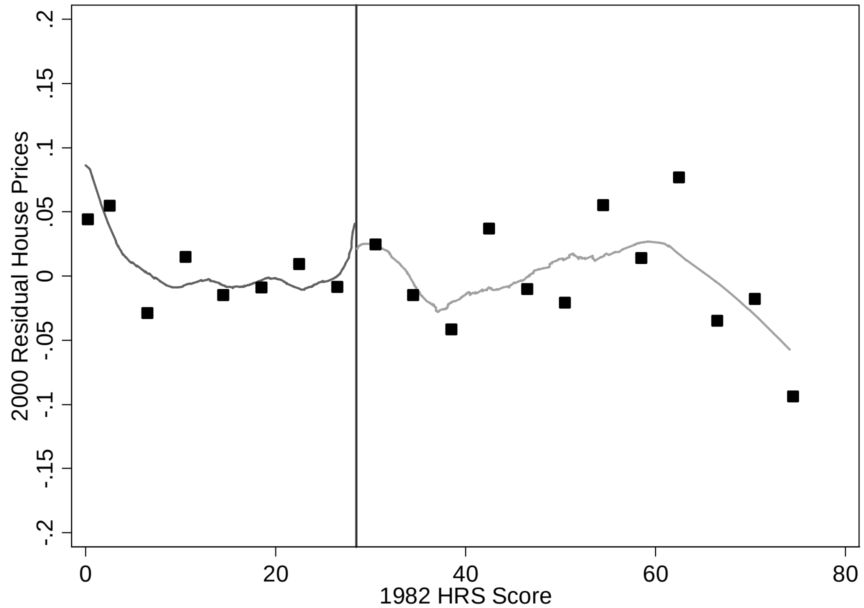
Notes: The figure displays the distribution of 1982 HRS scores among the 487 hazardous waste sites that were tested for placement on the NPL after the passage of the Superfund legislation but before the announcement of the first NPL in 1983. The 188 sites with missing housing data in 1980, 1990, or 2000 are not included in the subsequent analysis and hence are excluded from this figure. The vertical line at 28.5 represents the cut-off that determined eligibility for placement on the NPL.

Figure 3: Probability of Placement on the NPL by 1982 HRS Score



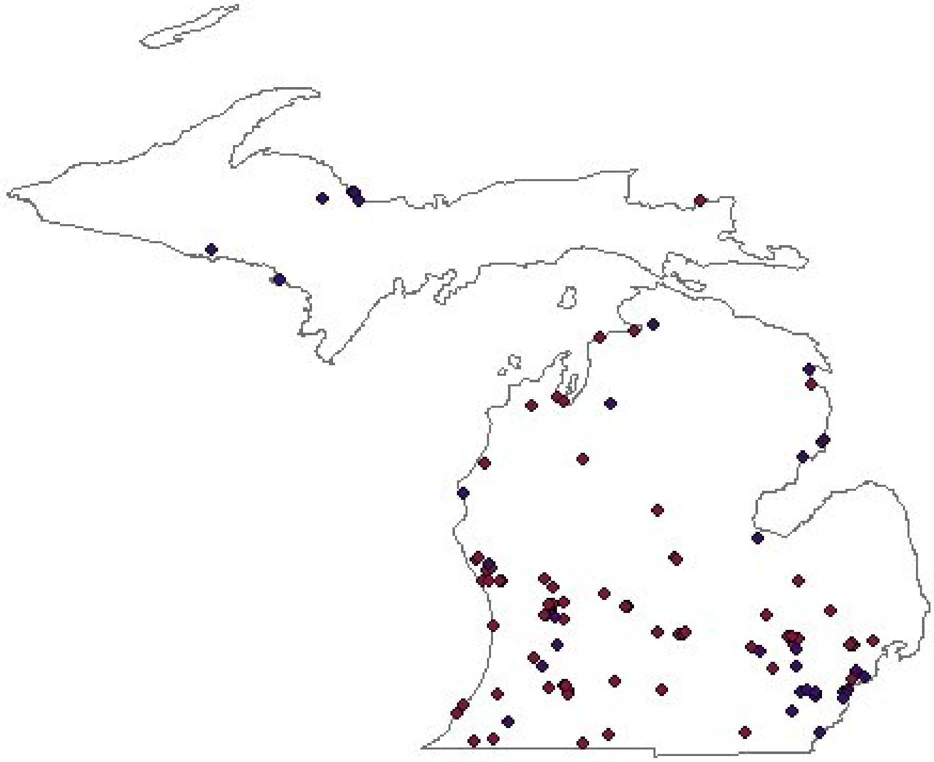
Notes: The figure plots the bivariate relation between the probability of 2000 NPL status and the 1982 HRS score among the 487 sites in the 1982 HRS sample. These plots are done separately for sites below (dark colored line) and above (light colored line) the 28.5 threshold. They come from the estimation of nonparametric regressions that use Cleveland's (1979) tricube weighting function and a bandwidth of 0.5. The data points present the mean probabilities in the same 4-unit intervals of the HRS score as in Figure 2. See the text for further details.

Figure 4: 2000 Residual House Prices after Adjustment for Column 4 Covariates, 2-Mile Radius Sample



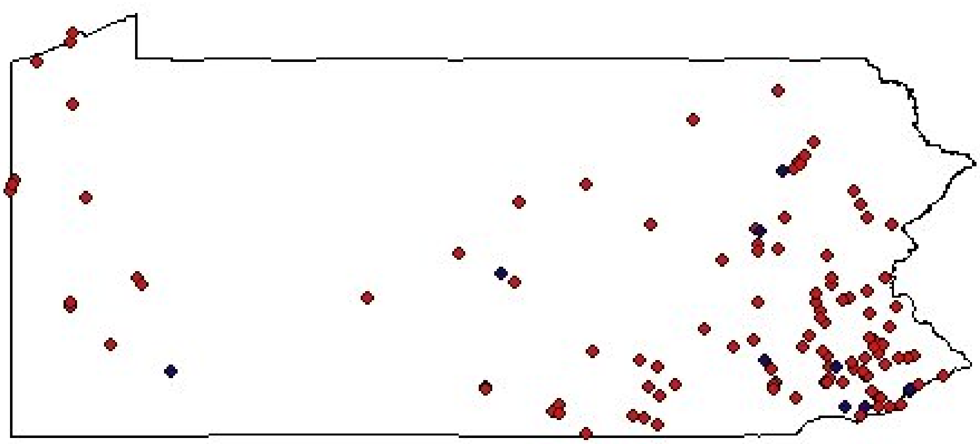
Notes: The figure plots the results from nonparametric regressions between 2000 residual housing prices from the 2 mile radius sample after adjustment for the covariates in the column (2) specification of Table 4 (except the indicator for a HRS score above 28.5) and the 1982 HRS scores. The nonparametric regressions use Cleveland's (1979) tricube weighting function and a bandwidth of 0.5. These plots are done separately for sites below (dark colored line) and above (light colored line) the 28.5 regulatory threshold. The data points are based on the same 4-unit intervals of the HRS score as in Figures 2 and 3. See the text for further details.

Figure 5a: Geographic Distribution of Hazardous Waste Sites in Michigan



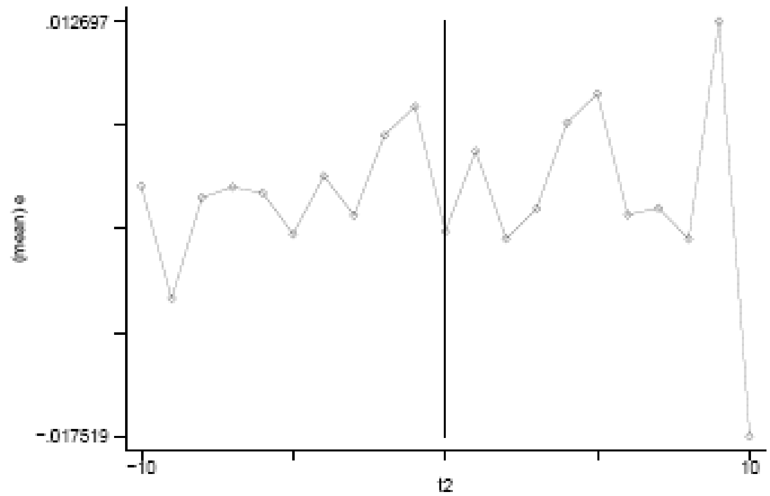
Notes: National Priorities List sites are in red and non-National Priorities List sites are in blue.

Figure 5b: Geographic Distribution of Hazardous Waste Sites in Pennsylvania

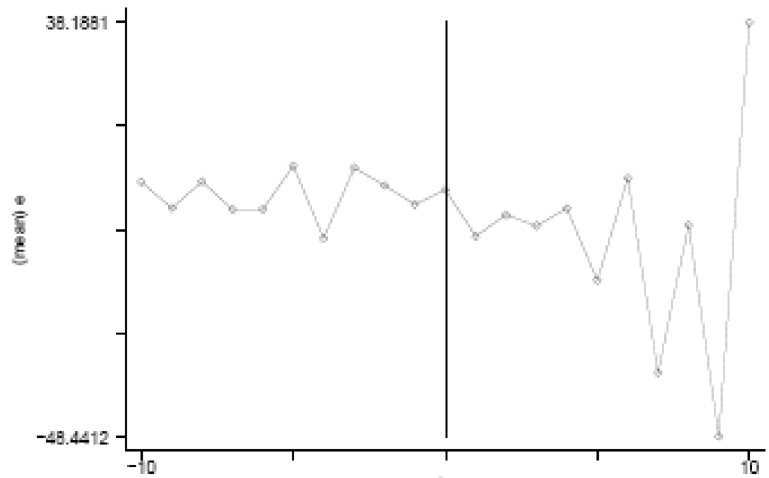


Notes: National Priorities List sites are in red and non-National Priorities List sites are in blue.

Figure 6: Health Outcomes Before and After a Site Cleanup: Low Birth Weight and Birth Weight



Low Birth Weight



Birth Weight

Note: The panels show the conditional average outcome for the 10 years before the clean up process is completed and the 10 years after the clean up process is completed. The vertical line represents the year when the clean up process is completed. The sample includes all births that occurred within 1 mile of one of the sites in PA and MI. Controls include site fixed effects, county by year effects, mother and father age, education race and Hispanic origin.