Who Lives on the Wrong Side of the Environmental Tracks? Evidence from the EPA's Risk-Screening Environmental Indicators Model*

Michael Ash, *University of Massachusetts, Amherst*

T. Robert Fetter, *Science Applications International Corporation (SAIC)*

Objective. We analyze the social and economic correlates of air pollution exposure in U.S. cities. *Methods.* We combine 1990 Census block group data for urbanized areas with 1998 data on toxicity-adjusted exposure to air pollution. Using a unique data set created as a byproduct of the EPA's Risk-Screening Environmental Indicators Model, we improve on previous studies of environmental inequality in three ways. First, where previous studies focus on the proximity to point sources and the total mass of pollutants released, our measure of toxic exposure reflects atmospheric dispersion and chemical toxicity. Second, we analyze the data at a fine level of geographic resolution. Third, we control for substantial regional variations in pollution, allowing us to identify exposure differences both within cities and between cities. *Results.* We find that African Americans tend to live both in more polluted cities in the United States and in more polluted neighborhoods within cities. Hispanics live in less polluted cities on average, but they live in more polluted areas within cities. We find an extremely consistent income-pollution gradient, with lower-income people significantly more exposed to pollution. *Conclusions.* Communities with higher concentrations of lower-income people and people of color experience disproportionate exposure to environmental hazards. Our findings highlight the importance of controlling for interregional variation in pollution levels in studies of the demographic correlates of pollution.

The study of environmental justice examines differential availability of environmental amenities or exposure to environmental disamenities on the basis of socioeconomic, ethnic, or racial difference. Under alternative definitions of environmental justice, inequality may itself constitute injustice, or the cause of the inequality may matter as well. Even on the straightforward

*Direct correspondence to Michael Ash, Department of Economics, University of Massachusetts, Amherst, MA 01003 ⟨mash@econs.umass.edu⟩. Ash will provide all data and coding information to those wishing to replicate the study. Ash and Fetter are equal co-authors. We thank James Boyce and John Stranlund for their invaluable guidance, and we are deeply grateful to Nicolaas W. Bouwes and Steven M. Hassur of the Environmental Protection Agency who painstakingly developed and graciously provided us with access to the Risk-Screening Environmental Indicators data.

SOCIAL SCIENCE QUARTERLY, Volume 85, Number 2, June 2004

question of what social attributes correlate with environmental quality, researchers disagree about methodology and findings. Using different indicators of environmental disamenities, units of spatial analysis, explanatory variables, and theoretical frameworks, researchers have found evidence for widely different conclusions.

With respect to recurrent themes in recent analyses, we contribute three methodological improvements. First, we use a realistic indicator of pollution exposure that is based on the U.S. Environmental Protection Agency's Toxics Release Inventory (TRI) but takes into account relative toxicity and chemical fate and transport. Second, we examine correlations at geographically small units, Census block groups, to avoid the ecological fallacy, that is, reaching conclusions from a large unit of analysis that do not hold at smaller resolution due to spatial heterogeneity. Third, we incorporate regional variations into a national analysis in a novel way.

Across all cities in the contiguous United States, we find that neighborhoods with higher proportions of African Americans tend to experience higher levels of toxicity-adjusted exposure to air pollution from TRI-reporting facilities than do predominantly white neighborhoods, whereas neighborhoods with higher proportions of Hispanics tend to experience lower levels of pollution. However, a model that compares neighborhoods in the same city shows that neighborhoods with more Hispanics and those with more African Americans have more pollution on average. Taken together, these results imply that African Americans tend to live in more polluted cities than do whites, and also tend to live in more polluted neighborhoods within cities. Although Hispanics tend to live in less polluted cities than do whites, they too live in more polluted neighborhoods within cities.

Literature Review

In this section, we survey recent practice in the environmental justice literature, with particular focus on the areas in which we offer methodological improvements. We then briefly address the conventional practices that we follow.

Choice of Pollution Indicator

Many studies of the demographic correlates of pollution use as the dependent variable the presence or absence of a polluting facility, such as a toxic storage and disposal facility (TSDF) (e.g., United Church of Christ, 1987; Mohai and Bryant, 1992; Anderton et al., 1994; Oakes, 1997; Been and Gupta, 1997; Boer et al., 1997; Pastor, 2001). The presence of a polluting facility in a neighborhood has validity as an indicator of environmental quality, especially as perceived by residents, but imprecisely measures exposure to hazard. The mass of pollutants released within community borders better proxies environmental quality, and most studies that use TRI data focus on the mass of toxic releases, for example, Bowen et al. (1995), Kriesel, Centner, and Keeler (1996), and Arora and Cason (1999).

Recognizing that mass of pollutants released within neighborhoods remains a blunt measure of environmental quality, some researchers have adjusted TRI data for toxicity and dispersion. Glickman and Hersh (1995) estimate risks from chronic exposure to industrial facilities in Alleghany County, Pennsylvania. Using TRI and other data, adjusted for toxicity and wind patterns, they find that Census block groups with more African Americans, poor people, and people over age 65 face higher risks compared to the rest of the population. In a high-resolution study of TRI releases in Des Moines, Iowa, Chakraborty and Armstrong (1997) show that plume-based models of dispersion reveal higher exposure of African Americans and poor people. McMaster, Leitner, and Sheppard (1997) demonstrate that finer geographic resolution and toxicity adjusting predict higher exposure of poor people to TRI releases in Minneapolis. Using TRI releases adjusted for chronic health effects and distance from pollution source, Brooks and Sethi (1997) find that zip codes with more African Americans experience greater pollution. The relationship holds even when income, education, urbanization, housing value, manufacturing employment, and population density are held constant. Using an index based on TRI releases, adjusted for toxicity and accounting for chemical fate and transport using EPA-reviewed models and databases, Bouwes, Hassur, and Shapiro (2001) find that densely populated square-kilometer neighborhoods with more African Americans, Hispanics, Asians, and unemployed residents tend to be more polluted than other densely populated neighborhoods.

Unit of Spatial Analysis

The unit of spatial analysis may significantly affect findings. One of the first studies to receive national attention (United Church of Christ, 1987) compares the demographics of zip codes containing commercial TSDFs to those of zip codes without TSDFs. Hird and Reese (1998) explore demographic correlations with 29 indicators of environmental quality at the county level. Brooks and Sethi (1997) use a sophisticated indicator of pollution, but define neighborhoods as zip codes. Such studies may suffer from ecological fallacy: correlations identified at large units of analysis may not hold at finer resolution. Most researchers would probably agree in the abstract that "the area chosen for analysis should correspond to the likely areal distribution of possible harm" (Anderton et al., 1994:128). Uncertainty about the range of possible harm from point sources makes difficult the precise definition of the appropriate area.

Controlling for Regional Characteristics

A third issue highlighted in recent research is what factors should be held constant to control for regional variation. Local development, for example, locations of markets, existing facilities, and transportation networks, influences the siting of polluting facilities. Moreover, local and state governments conduct much industrial location policy and implement environmental regulations; for these policymakers, the relevant social patterns of pollution are local. National analysis may overlook regional variations with regard either to base levels of environmental quality or to the structural relationship between race and other variables and pollution.

Studies have controlled for broad regional variations by allowing the base level of pollution to differ by urban or rural status, by population density (Bouwes, Hassur, and Shapiro, 2001), or by performing separate regressions by region, for example, South versus non-South (Arora and Cason, 1999), South, West, and remainder of United States (Hird and Reese, 1998), or EPA region (Anderton et al., 1994). Other studies have controlled for variations by choosing the comparison population carefully. Anderton et al. (1994) compare demographics between Census tracts containing TSDFs and tracts without TSDFs within the same Metropolitan Statistical Area (MSA) or rural county, positing that only those tracts could serve as alternative sites. Their choice of the comparison population contrasts with that of the United Church of Christ (1987), which compares demographics between tracts containing TSDFs and all tracts in the United States without TSDFs. Mohai (1995) and Been and Gupta (1997) note that the choice of Anderton et al. (1994) reduces the observed differences between the racial and ethnic composition of the host and nonhost tracts.

Kriesel, Centner, and Keeler (1996) report a positive correlation between the total mass of TRI releases within one mile of a block group and the percent of nonwhite residents in the block group in both Georgia and Ohio, holding constant the poverty rate and voter participation. When they add six supposedly "non-discriminatory industrial location factors"—manufacturing employment and wages, presence of an interstate highway, population density, education to proxy labor productivity, and housing values to proxy the cost of living—the positive association with percent nonwhite disappears.

Controlling for industrial location factors is problematic for two reasons. First, it is difficult to identify variables that proxy the intended economic factors but have no independent relationship with pollution. For example, education could influence not only productivity but also the propensity of a community to resist polluting facilities; or housing values could be an effect rather than a cause of facility location. Second, the location factors may reflect discriminatory history with lasting consequences. Examining 12 U.S. cities, Rabin (1989) finds that local planners changed the zoning of residential land occupied mainly by low-income African Americans to industrial or commercial. In many cases, African Americans continued to live side by side with new industrial land uses. If policymakers created conditions in the mid-20th century that encouraged industrial development in African-American neighborhoods, then factors such as the proximity of interstate highways and housing value may not, in fact, be nondiscriminatory.

Additional Control Variables

Many environmental justice studies include income as an explanatory variable. Environmental quality is a normal good: people with higher incomes will choose to live in areas with higher environmental quality, and areas with lower incomes, all else equal, will be more polluted. If correlation between ethnicity or race and pollution disappears when income is held constant, then we have not found environmental racism, *per se*, but have identified environmental inequality and, some would argue, environmental injustice. Recent authors, for example, Been and Gupta (1997), Boer et al. (1997), and Brooks and Sethi (1997), posit a nonlinear relationship between pollution and income. These studies find an inverse-U relationship: in neighborhoods with very low levels of income, pollution reflects additional economic activity and increases with income; but at higher incomes, the relationship becomes negative, as richer neighborhoods exercise economic or political power to obtain high environmental quality.

Population density is another commonly used explanatory variable, but theory gives little guidance about the expected correlation with pollution. A positive correlation might reflect more economic activity and thus more pollution in areas with more people. On the other hand, local officials likely work to reduce pollution in densely populated places.[1] If denser neighborhoods also have more people of color, then a finding of disparate pollution burdens in minority neighborhoods when population density is held constant at least excludes an alternative explanation of the correlation.

Many studies include additional socioeconomic variables either as controls to narrow the possible reasons that racial and ethnic patterns are observed, or because their correlation with pollution is interesting in its own right. Education, voter turnout, the percent of owner-occupied housing units, and the percent of vacant housing units are commonly included. The direction of correlation for these other variables is difficult to predict because they tend to be collinear with race and income in multivariate models.

[1]Boer et al. (1997) find that population density is not an important predictor once they control for the proportion of land devoted to industry, utilities, transportation, and communication, and suggest that in most studies (which do not control for land use), population density stands in for industrial land use. Consistent national data on land use are not available.

Methodology

In this section we develop our multivariate model of the social correlates of pollution exposure. Our measure of exposure to environmental hazards captures great detail about the dispersion and toxicity of pollution. Our pollution measure also has fine geographic resolution, which allows us to use the smallest unit for which all Census socioeconomic data are available—the block group. By analyzing exposure rather than proximity to source and by using Census block groups as the unit of analysis, we address the problem posed by Anderton et al. (1994) of determining the "areal distribution of possible harm," and we avoid the ecological fallacy. We control for regional variation in a novel way by including a fixed effect for each city, which allows for different base levels of pollution and other city-specific patterns of development. At the same time, by restricting the regression coefficients to be identical across cities, we are able to pose the question of whether—at a national level—there are common demographic characteristics of "the wrong side of the environmental tracks."

Although many studies group all racial and ethnic minorities together, we include Hispanics, non-Hispanic blacks, and non-Hispanic Asians and Pacific Islanders as separate categories and find that they have different patterns of exposure. We exclude Native Americans, who have very low representation in U.S. cities, and the residual "other race" Census category.

Based on common practice in the literature, we include the percent of residents with less than a high school education, the percent of vacant housing units, population density, and both median household income and its square. We also include the percentage of households with asset income from interest, dividends, or property rental as a proxy for wealth, although the percentage fails to capture the quantity of unearned income. Lastly, we include the percent of housing units that are owner occupied as a proxy for stability, social cohesion, and, hence, potential effectiveness in resisting the siting of polluting facilities. Although we consider voter participation an important variable for studies of environmental justice, data are available only at the county level; we limit our analysis to variables available for block groups.

We estimate three multivariate specifications of the pollution exposure model. In the equation below, i indexes neighborhoods and j indexes cities. By including fixed effects for 393 cities in some specifications, we control for the component of the error term associated with each city. We thereby identify the demographic correlates of pollution between neighborhoods within cities. The fixed effect captures the base level of all variables for the area; hence, in the fixed-effect specification, the coefficients are identified on the basis of variation within each area.

The full specification of the model is:

$$\begin{aligned}\mathrm{POLLUTION}_{ij} = \beta_0 + \mathrm{MINORITY}_{ij}\,\beta_{\mathrm{MINORITY}} + f(\mathrm{INCOME}_{ij})\\ + X_{ij}\beta_x + \delta_j + \varepsilon_{ij}\end{aligned}$$

where POLLUTION is either the continuous or dichotomous variable described below, MINORITY is a vector with the percent of Hispanic, non-Hispanic black, and non-Hispanic Asian residents, and the polynomial f(INCOME) includes linear and quadratic terms in median household income. The vector X_{ij} designates the additional explanatory variables. The first component of the error term, δ_j, is a fixed effect for the entire city, and the second component, ε_{ij}, is a standard white-noise error for the neighborhood.

In the first specification, we use only ethnicity and race as independent variables. In the second specification, we add income, and for the first two specifications, we explore the importance of the fixed-effect error component. Finally, we estimate the full model. We report the coefficients for the additional variables, but we focus on how they mediate the relationship between the racial/ethnic variables and pollution.

Data

Pollution Data

The pollution data for this article are derived from the EPA's Risk-Screening Environmental Indicators (RSEI) Model (Bouwes and Hassur, 2002). Because it incorporates detailed data on the toxicity and dispersion of chemical releases, the RSEI Model gives more realistic information on potential human health effects from air pollutants than has been available for most previous studies. RSEI provides a unitless measure, intended for relative comparisons, rather than a physically denominated measure of risk or exposure potential.

The pollution sources considered in this article are those facilities that report emissions to the TRI. The 1986 Emergency Planning and Community Right-to-Know Act (EPCRA) requires operations engaged in manufacturing, metal and coal mining, hazardous-waste treatment and disposal, solvent recovery, electrical generation, and chemical and petroleum distribution, as well as federal facilities, to report releases of designated pollutants to air, water, and land if the operations exceed specified thresholds of employment and chemical use (U.S. EPA, 2000). For air releases, TRI guidelines for 1998 required facilities to report emissions of 604 different chemicals and chemical categories (U.S. EPA, 1999). Facilities must estimate and report both intentional ("stack") and unintentional ("fugitive") releases, with some EPA monitoring and oversight. In 1998, 23,396 facilities reported direct releases of 2.1 billion pounds of chemicals to air.

Although many researchers have analyzed the distribution of pollution using TRI data, TRI does not include information on the toxicity of the various chemicals or on their dispersion once released. The RSEI Model adds toxicity and dispersion information to the TRI data. The data used in this article differ from the data available in the public release of the RSEI Model in three main ways. First, the data used here, which were generated as a byproduct of the RSEI modeling procedure, are organized by area, rather than by TRI facility. Second, the data used here consider only exposure to air pollution via inhalation, whereas the published data also consider water and ground pollution and multiple pathways (e.g., ingestion, direct skin contact). Finally, the data in the public release include a population-weighting term used in calculating a risk-related measure; the data used here do not.[2]

According to the databases used in constructing the RSEI Model, the 604 chemicals and chemical categories listed in the TRI vary in toxicity by up to eight orders of magnitude. If a chemical has both cancer and noncancer effects, the higher of the cancer and noncancer weights is used (Bouwes, Hassur, and Shapiro, 2001; Bouwes and Hassur, 1997). Data on chemical toxicity come from EPA's Integrated Risk Information System, Health Effects Assessment Summary Tables, and other sources. All toxicity data have been reviewed by EPA scientists, and most were also peer reviewed by external scientists (Bouwes and Hassur, 2002).

The RSEI Model incorporates facility- and chemical-specific data relevant to potential human exposure. Transport factors include wind speed, direction, and turbulence, and stack heights and exit gas velocities that are either facility specific (where available) or based on median values for the facility's industry (Bouwes and Hassur, 1997, 2002). Chemical-specific factors include rates of decay and deposition. The fate and transport model is the Industrial Source Complex Long-Term (ISCLT3) Model, developed by EPA's Office of Air Quality Planning and Standards (Bouwes and Hassur, 2002; U.S. EPA, 1995).

Based on these data, the RSEI Model estimates ambient concentrations of each TRI pollutant. A concentration is determined for each square kilometer of the 101-km by 101-km grid in which the facility is centered.[3] After calculating ambient concentrations, the model uses standard assumptions about human exposure to derive a surrogate dose—an estimate of the amount of chemical contacted by an individual per kilogram of body weight per day.

The RSEI Model combines chemical-specific toxicity weights with the surrogate dose delivered by each release to obtain a partial score for each square-kilometer cell that represents the relative, toxicity-adjusted potential human health effects from chronic exposure. The partial scores resulting from

[2]The EPA's screening method identifies priorities for cleanup based on overall environmental danger or damage, which increases with the exposed population. For our purposes of identifying the demographic factors that correlate with increased individual exposure to pollution, we do not consider a more populated area, given the same ambient concentrations of pollutants, to be more polluted than a less populated one.

[3]TRI data contain some facility location errors. The RSEI Model development included an EPA Quality Assurance process and a separate geocoding to improve location data for over 9,000 facilities (Bouwes and Hassur, 2002).

releases at different facilities are summed to obtain the score for each cell:

$$Score_g = \sum_f \sum_c Toxicity_c \times Surrogate\ Dose_{cfg}$$

for square-kilometer cell g, where c and f index chemical c released by facility f. We can compare scores across cells to evaluate the relative potential for chronic human health effects. In the published data, the scores are aggregated across cells for each facility, and a single score is reported for each facility. Thus, scores for individual cells are an unpublished building block of the public data. They were made available for this analysis by special arrangement.

Compared to the measures used in most previous studies of the distribution of pollution, the RSEI data have two major advantages. First, the detailed information on chemical toxicity allows a much more realistic measure of the potential human health effects arising from pollution. Second, the data used here are based on a realistic representation of exposure, incorporating chemical fate and transport, as well as stack heights and exit gas velocities. Most previous studies have used a single threshold distance or a simple distance-decay function to approximate the dispersion of pollution, ignoring site-specific characteristics. Moreover, the model achieves fine geographic resolution of pollution risk-related impacts, allowing the use of correspondingly fine units for demographic data without the danger of mismatched areal units.

The data also have several important limitations. First, the underlying TRI data are estimated and self-reported with limited EPA oversight; firms may misreport or inaccurately estimate releases (Szasz and Meuser, 1997). Second, the data used here represent only chronic health effects from inhalation of the 604 TRI-listed chemicals released by TRI reporting facilities. Third, the concentration data are based on a dispersion model assuming continuous release rather than on direct measurement.[4]

Demographic Data

We use Census block groups to represent neighborhoods. With the help of local committees, the Census Bureau defines Census block groups to correspond to neighborhoods (U.S. Department of Commerce, 1994). Block groups typically contain 250 to 550 households and fully partition Census tracts, which contain 2,500 to 8,000 residents. The block group is the smallest aggregation for which the Census Bureau publicly releases income data. All demographic and socioeconomic variables come from the 1990 Census of Population and Housing (U.S. Department of Commerce,

[4]See Bouwes and Hassur (1997, 1998, 1999, 2002) for more information about sensitivity analysis and ground-truthing of the RSEI Model.

1992). We use 1990 Census data to ensure that demographic characteristics would be prior to, and hence not caused by, subsequent pollution.

We limit our analysis to Census-designated urbanized areas in the contiguous United States. An "urbanized area" is continuously built up with at least 50,000 people and local density above 391 people per km^2 (U.S. Department of Commerce, 1994). Including suburbs but not rural portions of counties, urbanized areas correspond better than do larger Metropolitan Statistical Areas to the look and feel of a metropolitan area. An appendix, available from the authors, lists the 393 urbanized areas used here with demographics, number of block groups, and average pollution level. In all, 66 percent of the 1990 population of the contiguous United States lived in urbanized areas.[5]

Merging Data Sets

The Census and RSEI data are well matched in geographic precision, but are not in the same geographic format. The RSEI Model divides the continental United States into approximately 8 million square-kilometer cells, of which about 2.2 million have positive scores (for 1998 TRI releases).[6] Census block groups can have irregular boundaries and can be either larger or smaller than one square kilometer. To take full advantage of the geographic resolution of the RSEI data, we merge the pollution and Census data by Census *blocks*, a finer level of resolution than the block group (an average block group contains about 30 blocks), and then aggregate block scores to the block group. We convert Census latitude and longitude to the geography of the grid-score lattice and assign each Census block the score of the square-kilometer cell in the lattice that contains the internal point of the block.[7] Then we compute block-group score as an average of the scores of component blocks.

Descriptive Statistics

Table 1 reports summary statistics for the dependent variable (the RSEI score) and the independent variables.[8] Figure 1 shows a histogram of the RSEI score variable (truncated at 600). With a large mass below one and a

[5]About 2,500 block groups were removed from the analysis because they reported zero population, zero land area, zero median household income, or because zero people in the block group reported their race, ethnicity, education levels, or other characteristics.

[6]We smoothed the data to provide scores, zero or positive, for all cells using an interpolated distance-weighted smoothing routine. An alternative approach of assigning zero to cells without scores gives very similar results, which limits our concern about spatial autocorrelation.

[7]The internal point is the location of the Census landmark (e.g., street intersection) closest to the block centroid and inside the block.

[8]Because block-group characteristics are not weighted by population, the averages may differ from national averages for urban areas.

TABLE 1

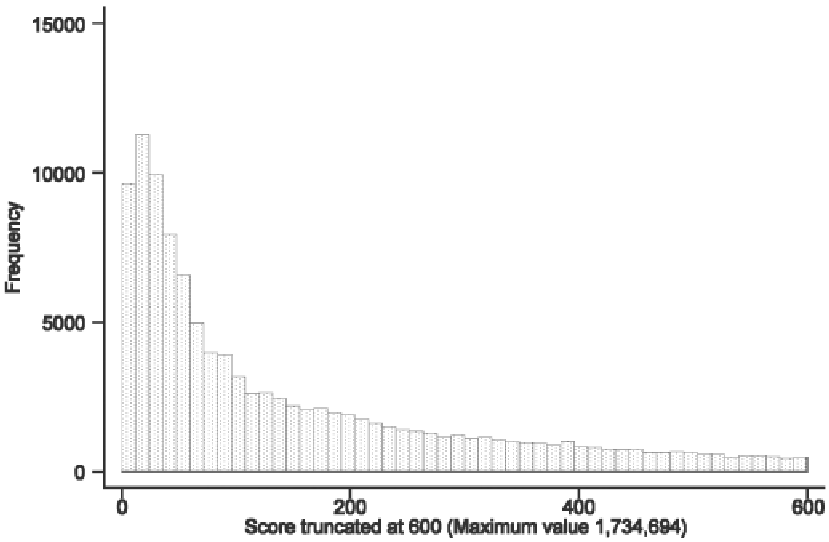
Summary Statistics (*N* = 136,362 Block Groups in Urbanized Areas)

Variable	Mean	Median	*SD*	Minimum	Maximum
RSEI score	716	139	10370	0.00	1734694
% Hispanic	9.9	2.3	18.6	0.0	100
% African American	16.0	2.2	28.7	0.0	100
% Asian/Pacific Islander	3.0	0.2	6.7	0.0	100
% Native American	0.5	0.0	1.7	0.0	94.4
% Other race	0.1	0.0	0.7	0.0	57.1
Median household income (000)	34.1	31.1	18.3	5.0	150
Population density (1,000 persons/km^2)	3.177	1.802	5.827	0.0001	397.556
% Adults without high school diploma	25.0	21.2	17.3	0.0	100
% Households with asset income	40.4	41.1	20.7	0.0	100
% Vacant housing units	7.1	5.1	7.8	0.0	94.4
% Owner-occupied housing units	61.5	66.9	27.9	0.0	100
Block-group area (km^2)	2.59	0.52	22.00	0.001	3202.21
Block-group population	1,196	978	980	1	35,682

NOTE: The 1990 Census topcodes median household income at $150,000.

FIGURE 1

Histogram of RSEI Scores by Block Group



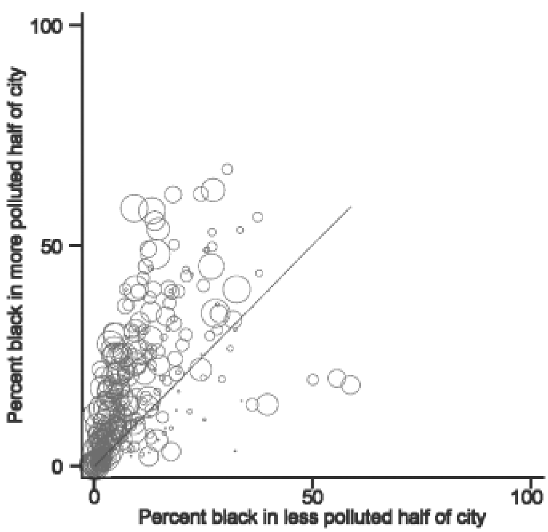
very long tail, the distribution of the dependent variable is clearly nonnormal. The 95th percentile of the dependent variable is around 2,100, and scores range up to 1.7 million. The average block group in this study has about 10 percent Hispanic residents, 16 percent non-Hispanic black residents, 3 percent residents of non-Hispanic Asian heritage, and 70 percent non-Hispanic white residents. Comparing the median and mean values for racial characteristics again shows significant right skew.

Although the mean area of a block group is 2.59 km^2, the median area is 0.52 km^2, and 75 percent of block groups have area less than 1.3 km^2. The average block group has a population of 1,196 with a density of 3,177 persons per km^2. The average median household income of block groups is about \$34,000. On average, 25 percent of the residents over age 18 lack a high school diploma. Forty percent of the households in an average block group receive asset income. On average, 7 percent of housing units in a block group are vacant, and 61 percent are owner occupied.

Figure 2 depicts the racial disparity between the more and less polluted half of each of the 393 urbanized areas. The vertical axis represents the percent African American in the more polluted half of each city (i.e., the block groups containing the more pollution-exposed 50 percent of the population), while the horizontal axis represents the percent African American in the less polluted half of each city. The 45-degree line shows equal representation of African Americans; cities that appear above the line have disproportionate representation of African Americans in the more

FIGURE 2

Percent African-American Population in More and Less Polluted Halves of Cities



NOTE: Size of circle represents pollution disparity between halves.

polluted half of the city, while cities that fall below the line have disproportionate representation of African Americans in the less polluted half of the city. The size of each circle represents the pollution disparity between the more and less polluted halves of the city. Visual inspection suggests that the more polluted halves of most cities are disproportionately African American. Equivalent scatter plots (available from the authors) for percent Hispanic and average income also suggest disproportionate exposure within cities. Multivariate analysis reported below support these findings.

Modeling Techniques

The extreme skew of the dependent variable suggests the use of limited dependent variable estimation rather than OLS. We use the Tobit model and the linear probability model (LPM). To motivate the censored approach of Tobit, we posit the existence of a latent variable representing the incapacity to avoid pollution; negative values of this latent variable would reflect progressively higher levels of capacity to secure environmental quality. But since the observed pollution cannot fall below zero, we observe scores of essentially zero for many block groups. Because several observations with very high scores could overwhelm the results, for the Tobit analysis we impose upper-tail censoring of the dependent variable at 2,300, slightly above the 95th percentile, to limit the influence of several very high scores but not to discard those block groups altogether. That is, the dependent variable for the Tobit is both lower and upper censored, with three ranges, that is, below 1, continuous between 1 and 2,300, and larger than 2,300.[9]

We also apply a dichotomous estimation technique. We estimate the probabilities that a neighborhood is in the more polluted half and most polluted tenth of its city. The quantiles are determined by population; for example, the most polluted tenth of the city is the set of most polluted block groups that contains one-tenth of the population of the urbanized area (not necessarily in contiguous block groups). Although we lose information in moving from a continuous to a dichotomous variable, the results of the dichotomous model are readily interpreted. We apply the LPM despite the well-known problems of induced heteroskedasticity and nonconforming probabilities because it is consistent even in fixed-effect models.

Regression Results

In this section, we first report results based on the Tobit model without and with fixed effects. We then report results based on the dichotomous

[9]Tobit may be inconsistent in models with fixed effects even when other assumptions are satisfied, but our fixed-effect Tobit estimates are reliable because the data include many block groups within cities. We also estimated the model with OLS, truncating the data at the 99th percentile, and the results with and without fixed effects are very similar to the Tobit results.

TABLE 2

Results for Tobit Estimation, With and Without Area Fixed Effects

Area fixed effects	1 No	2 No	3 393 UAs	4 393 UAs	5 393 UAs
% Hispanic	– 2.25**	– 4.68**	3.07**	0.96**	– 0.39**
	(0.09)	(0.09)	(0.08)	(0.08)	(0.10)
% African American	2.87**	0.24**	2.32**	0.81**	0.23**
	(0.06)	(0.06)	(0.04)	(0.048)	(0.05)
% Asian/Pacific Islander	– 10.64**	– 8.26**	– 0.123	– 0.91**	– 0.82**
	(0.25)	(0.24)	(0.188)	(0.19)	(0.19)
Median household income (000)		– 18.4**		– 10.2**	– 6.03**
		(0.3)		(0.2)	(0.27)
Square of income		0.0997**		0.061**	0.038**
		(0.0023)		(0.002)	(0.002)
Population density (1,000/km^2)					– 1.20**
					(0.23)
% Adults without HS diploma					3.18**
					(0.11)
% Households with asset income					– 0.64**
					(0.10)
% Vacant housing					2.34**
					(0.17)
% Owner-occupied housing					0.020
					(0.061)
R^2	0.3%	0.7%	5.5%	5.7%	5.8%

#significant at $p<0.10$; *significant at $p<0.05$; **significant at $p<0.01$.

NOTES: Dependent variable is RSEI score: 1,376 observations left censored at or below 1; 6,209 observations right censored at or above 2,300. All variables are in units indicated in Table 1. Constant is included in Columns 1 and 2. Standard errors are in parentheses.

LPM model. Table 2 reports the coefficient estimates and standard errors for the model specifications estimated with Tobit.

Tobit Model Without Area Fixed Effects

The first two columns of Table 2 report the results of national-level estimation without urbanized-area effects. In Column 1, where the model includes only the race and ethnicity variables, the coefficient on percent African American is positive, indicating that block groups with higher proportions of African Americans tend to have higher RSEI scores, while the coefficients on percent Hispanic and Asian are both negative, meaning that block groups with more of these ethnic groups tend to have lower RSEI scores. When we add income in Column 2, the coefficient on percent African American falls rather sharply but remains positive and materially and statistically significant. This result implies that the positive relationship observed between the percent of African-American residents and the RSEI score observed in Column 1 is in part due to the negative correlation between percent African-American residents and household income. Likewise, the magnitude of the negative relationship for Hispanics increases when income controls are added, which suggests that Hispanic exposure is higher because of lower average Hispanic income. However, the magnitude of the negative relationship between percent Asian and pollution diminishes after income is included, meaning that Asians are more exposed than their incomes alone would suggest. Higher median income is strongly associated with lower pollution. The quadratic and linear terms imply a concave relation with the minimum at $92,000, which means that virtually every block group lies in the domain where increasing income is associated with decreasing pollution exposure.

Tobit Model with Area Fixed Effects

The fixed-effect estimates shown in Columns 3 and 4 of Table 2 reveal a striking difference between demographic correlations between cities and correlations within cities. Within urbanized areas, the strong positive relationship between percent African American and pollution score persists. But when base pollution levels are permitted to vary among cities, there is a positive relationship between the percentage of Hispanics in a block group and the RSEI score. Taken together, results from the models with fixed effects (Columns 3 and 4) and the corresponding models without fixed effects (Columns 1 and 2) suggest that Hispanics live in cleaner cities, but that within the cities where they live, they tend to live "on the wrong side of the environmental tracks"—that is, in more polluted block groups.

The coefficient on the percent of African Americans remains positive and significant in the fixed-effects model. Taken together, the results of the models without and with fixed effects indicate that African Americans live both in more polluted cities in the United States and also in the more polluted block groups of the cities in which they live. The magnitude of the intra-city effect revealed by the fixed-effect models is stronger for Hispanics than for African Americans in Columns 3 and 4. In all the Tobit specifications, Asians are found to live in less polluted neighborhoods. The income coefficients remain stable with the addition of the 393 urbanized area fixed effects. The quadratic and linear terms imply that the minimum occurs at around $80,000; more than 95 percent of block groups are in the decreasing part of the function.

Visual inspection of a national RSEI-score map and regression analysis, available from the authors, suggest that the racial and ethnic between-city results are largely due to the concentration of Hispanics and Asians in the West and the Southwest, which generally have lower concentrations of heavy industry and TRI emitters. This analysis also finds that the intercity effect for African Americans—African Americans tend to live in significantly more polluted cities than do whites—is due to the concentration of African Americans in the Rust Belt cities of the Northeast and Midwest.

To elaborate the social characteristics that drive the results, we included an extended list of covariates (Column 5 of Table 2). The coefficient on percent African American is reduced, although the value remains positive and highly significant. The sign on percent Hispanic becomes negative, which suggests that the Hispanic effect is largely explained by other social characteristics, including education, unearned income, and housing vacancy rate. The coefficient on population density is negative, which implies that within urbanized areas, less densely populated block groups have higher RSEI scores. This finding suggests that planners and public health officials work to locate polluting facilities in sparsely populated areas, but is also consistent with the proposition that sparsely populated urban areas have less political influence. The fraction of vacant housing units is also associated with increased pollution; high vacancy is not a correlate of low density ($r = -0.0062$) but may reflect neighborhood disempowerment or distress. Block groups with a greater proportion of adults without high school diplomas, and those with a lower proportion of households receiving asset income, tend to be more polluted.

Despite the high fraction of block groups with pollution scores near zero, which could bias OLS results, simple OLS regression yielded results very similar to those of Tobit, albeit with slight attenuation in the coefficients. The imposition of the upper censoring, however, had substantial effects on the results. If the highest values of the RSEI score are included, they drive the results and generate substantially different regression coefficients.[10]

LPM Results

In Table 3, we report the probability that a block group is in the most polluted fraction of its urbanized area as a function of its social characteristics. The results are based on a fixed-effects linear probability model, and thus are most comparable to Columns 3 through 5 of Table 2. In the first three columns, we examine the probability that a block group is among the *more polluted half* of its city. The estimated coefficients in the LPM can be interpreted as percentage point changes. For example, in Column 1 of Table 3 we find that a block group that is 100 percent Hispanic has a probability 51 percentage points greater of being in the more polluted half of the city compared to an otherwise identical block group that

[10]The censoring value makes little difference. We tested censoring at centiles 90, 95, 99, and 99.5 and found very similar Tobit results for all upper-censoring points. The block groups with the very highest scores, which have median incomes close to the overall median and are more white than is the average block group, exert substantial leverage.

TABLE 3

Results for LPM Estimation with Area Fixed Effects

Most polluted . . . of UA	1 Half	2 Half	3 Half	4 10th	5 10th	6 10th
% Hispanic	0.517**	0.228**	0.061**	0.194**	0.094**	−0.008
	(0.009)	(0.010)	(0.012)	(0.006)	(0.006)	(0.008)
% African American	0.337**	0.138**	0.086**	0.083**	0.014**	−0.033**
	(0.005)	(0.006)	(0.006)	(0.003)	(0.004)	(0.004)
% Asian/Pacific Islander	0.103**	0.000	−0.040#	−0.037**	−0.073**	−0.050**
	(0.022)	(0.022)	(0.022)	(0.015)	(0.015)	(0.015)
Median household income (000)		−0.0121**	−0.0082**		−0.0043**	−0.0011**
		(0.0002)	(0.0003)		(0.0001)	(0.0002)
Square of income		0.000061**	0.000040**		0.0000222**	0.0000052**
		(0.000002)	(0.000002)		(0.0000013)	(0.0000015)
Population density (1,000/km^2)			0.0004**			−0.0003**
			(0.0003)			(0.0002)
% Adults without HS diploma			0.325**			0.245**
			(0.014)			(0.009)
% Households with asset income			−0.027*			−0.083*
			(0.012)			(0.008)
% Vacant housing			0.06**			0.080**
			(0.02)			(0.013)
% Owner-occupied housing			−0.0006			−0.0044
			(0.0072)			(0.0047)
R^2 (within)	5.3%	8.6%	9.2%	1.2%	2.2%	3.3%

#significant at $p<0.10$; *significant at $p<0.05$; **significant at $p<0.01$.

Notes: Dependent variable is dichotomous, with 1 indicating inclusion in the more polluted segment of the city. All variables are in units indicated in Table 1. Coefficients express the percentage point change in probability of exposure per unit change in the explanatory variable. Standard errors are in parentheses.

is 100 percent white non-Hispanic. The results indicate that block groups that are a higher proportion Hispanic, African American, or Asian are all more likely to be in the more polluted half of the urbanized area, although the Asian effect disappears when controls for income are added. We also find that over most of the observed range of incomes, income is negatively correlated with the probability of being in the more polluted half of the city, and the minimum of the income-pollution gradient occurs at a neighborhood income of $99,000. At the median, a $10,000 increase in income is associated with a seven percentage point decrease in the probability of being in the more polluted half of the city. In the elaboration (Column 3), we find, as in the Tobit model, that block groups with a higher fraction of adults who did not graduate from high school are substantially more likely to be in the more exposed half and that asset income is associated with less pollution exposure.

When we turn to the *most polluted 10th* of cities, the explanatory power of the model drops, but most of the results persist (Columns 4 through 6 of Table 3). In Column 4, which shows results for a model that includes only the race and ethnicity variables, we find that Hispanics and African Americans are substantially more likely than non-Hispanic whites to live in the most polluted 10th of cities, and Asians are less likely to live in the most polluted 10th. When median household income is included in the regression, the race and ethnicity effects decline but remain positive and significant. When we include the full set of neighborhood covariates, the signs on the coefficients for percent Hispanic and percent African American actually reverse, although the size of the coefficients is small. This result suggests that for inclusion in the most polluted portions of the city, the other social indicators explain the correlations between pollution and Hispanic and African American. We find strong negative associations between inclusion in the most polluted 10th and both population density and asset income; we find strong positive relationships between presence in the most polluted 10th and both vacant units and adults with less than high school education. Higher neighborhood income implies a steadily declining probability of membership in the most polluted 10th.[11]

The models explain a small proportion of the variation in the RSEI score. Without fixed effects, the pseudo-R^2 for the Tobit regressions in Table 2 are below 0.01. When we estimate the same models using OLS, the R^2 in the models without fixed effects reach only 0.12 for the models with the full set of explanatory variables. Even when we estimate the models with 393 urbanized area fixed effects, the pseudo-R^2 of the Tobits climbs only to around 0.06. Although R^2 is a poor measure of fit for the LPM, the within-city R^2 for the fixed-effect linear probability models ranges from 0.01 for the

[11]We also examined the correlates of the worst quarter and of centiles 75 through 90, and found results that were generally intermediate between those for the worst 10th and the worse half.

sparest specification of the regression for the most polluted 10th, to 0.09 for the richest specification of the regression for the more polluted half.

In the few studies of environmental inequality that report R^2, the percentage of variation explained exceeds that of our model. We experimented with adding variables used by authors whose models had higher R^2 and found that the additional variables do not substantially raise the R^2 of our model. Aggregated analyses are likely to have upward-biased R^2 because of the ecological fallacy. Our highly disaggregated analysis reduces its effect. Despite a low proportion of explained variation, we find statistically significant correlations between demographic variables and pollution, as evidenced by the high t-statistics in the multivariate results.

Discussion and Conclusions

Previous environmental justice research has generally failed to address the demographics of pollution in terms of toxicity and exposure, focusing instead on proximity to pollution sources or on the mass of pollutants released. Many previous studies use large units of spatial resolution, and most national studies control for regional variation, if at all, only by estimating correlations separately for a small number of regions in the country or separately for densely and sparsely populated regions.

This article addresses these issues by using data from the EPA's RSEI Model. Our results indicate that in the urban United States as a whole, block groups with more African Americans have higher levels of exposure to toxic pollution from TRI facilities, while block groups with more Hispanics and Asians/Pacific Islanders have lower levels. When we control for differences between cities, however, we find that within cities, Hispanics, as well as African Americans, tend to live in more polluted neighborhoods. In national comparisons, this disparity is offset for Hispanics by the fact that they tend to live in cities with relatively low levels of industrial toxics. African Americans, by contrast, tend to live in more polluted cities as well as in the more polluted neighborhoods within cities.

There are several important caveats regarding the data underlying this analysis. First, our dependent variable represents only a subset of pollutants that people face. Mobile and small point sources, such as automobiles and dry cleaners, are excluded. The data also omit exposure via other media, such as water pollution and nonresidential, for example, workplace, exposure. Second, the measure of toxicity does not include possible synergistic effects of multiple pollutants. Third, although the RSEI Model incorporates much site-specific data, the dependent variable relies on modeling and some generalizations.

Our methodology also warrants several caveats. A cross-sectional analysis cannot elucidate causal relationships between demographics and pollution. The correlations between the demographic variables and RSEI scores could be caused by various underlying factors. For instance, even when incomes are similar, African Americans or Hispanics may have lower average wealth than do whites, which would constrain housing choices. African Americans or Hispanics also may tend to have less access to information about the health effects of pollution. Or there may be racism in housing or credit markets, or in the siting of industrial facilities. This study cannot ascertain which processes underlie the results. We note, however, that the strong effect of race and ethnicity, controlling for income, suggest that voluntary move in spurred by low income is not a plausible explanation for differences in exposure.

Two previous national studies account both for relative toxicity and for atmospheric dispersion. Their results are only partially comparable due to differences in methodology but are generally consistent with respect to the coefficients on race and ethnicity variables. Brooks and Sethi (1997) report a positive correlation between the percent of African Americans in a zip code and their index of pollution. Bouwes, Hassur, and Shapiro (2001) report a positive correlation between their index of pollution and both the percent of African Americans and the percent of Hispanics.[12]

This study offers three key insights to inform future research and policy. First, different minority groups should be included separately in econometric analysis rather than lumped together as "all nonwhite" residents. Second, national environmental justice studies should account for variations in base levels of pollution in order to avoid collapsing variation within cities and variation among cities into a single coefficient. Third, for analyzing *exposure* (though not necessarily proximity to point sources), the spatial unit of analysis should be as small as possible because larger units can obscure heterogeneity.

What are the conclusions for environmental justice policy? Our results support past findings that pollution burdens fall disproportionately on African Americans and poor people throughout the United States—and on Hispanics within regions. Our results also suggest that local policymakers bear special responsibility to address disparate exposure. The results for African Americans imply that environmental justice should remain a priority for national as well as regional environmental policy. In addition, the results highlight the value of the EPA's RSEI data for environmental justice analyses. The fine geographic resolution and well-developed dependent variable make these data exceptionally well suited to analyze the demographics of pollution.

[12]Although Bouwes, Hassur, and Shapiro (2001) also use RSEI data to generate a dependent variable, the positive sign on the Hispanic variable differs from our result in the model without fixed effects. Three methodological differences may account for the difference. Their measure of pollution is weighted by population, so that more populated areas have higher values of pollution even if the unweighted pollution level is the same; they use only observations for which the value of the RSEI score is greater than zero; and the unit of their analysis is the square-kilometer cell rather than the Census block group.

REFERENCES

Anderton, D. L., A. B. Anderson, P. H. Rossi, J. M. Oakes, M. R. Fraser, E. W. Weber, and E. J. Calabrese. 1994. "Hazardous Waste Facilities: 'Environmental Equity' Issues in Metropolitan Areas." *Evaluation Review* 18(2):123–40.

Arora, S., and T. N. Cason. 1999. "Do Community Characteristics Influence Environmental Outcomes? Evidence from the Toxics Release Inventory." *Southern Economic Journal* 65(4):691–716.

Been, V., and F. Gupta. 1997. "Coming to the Nuisance or Going to the Barrios? A Longitudinal Analysis of Environmental Justice Claims." *Ecology Law Quarterly* 24(1):1–56.

Boer, J. T., M. Pastor, J. L. Sadd, and L. D. Snyder. 1997. "Is There Environmental Racism? The Demographics of Hazardous Waste in Los Angeles County." *Social Science Quarterly* 78(4):793–810.

Bouwes, N. W., and S. M. Hassur. 1997. *Toxics Release Inventory Relative Risk-Based Environmental Indicators Methodology*. Washington, DC: U.S. EPA Office of Pollution Prevention and Toxics. Available at ⟨http://www.epa.gov/oppt/rsei/docs/method97.pdf⟩.

———. 1998. *Groundtruthing of the Air Pathway Component of OPPT's Risk-Screening Environmental Indicators Model (Draft)*. Washington, DC: U.S. EPA Office of Pollution Prevention and Toxics. Available at ⟨http://www.epa.gov/oppt/rsei/docs/ground98.pdf⟩.

———. 1999. *Estimates of Stack Heights and Exit Gas Velocities for TRI Facilities in OPPT's Risk-Screening Environmental Indicators Model*. Washington, DC: U.S. EPA Office of Pollution Prevention and Toxics. Available at ⟨http://www.epa.gov/oppt/rsei/docs/stacks99.pdf⟩.

———. 2002. *User's Manual for RSEI Version 2.0 Beta 2.0*. Washington, DC: U.S. EPA Office of Pollution Prevention and Toxics. Available at ⟨http://www.epa.gov/oppt/rsei/docs/users_manual.pdf⟩.

Bouwes, N. W., S. M. Hassur, and M. D. Shapiro. 2001. *Empowerment Through Risk-Related Information: EPA's Risk-Screening Environmental Indicators Project*. Working Paper DPE-01-06. Amherst, MA: Political Economy Research Institute.

Bowen, W. M., M. J. Salling, K. E. Haynes, and E. J. Cyran. 1995. "Toward Environmental Justice: Spatial Equity in Ohio and Cleveland." *Annals of the Association of American Geographers* 85(4):641–63.

Brooks, N., and R. Sethi. 1997. "The Distribution of Pollution: Community Characteristics and Exposure to Air Toxics." *Journal of Environmental Economics and Management* 32: 233–50.

Chakraborty, J., and M. P. Armstrong. 1997. "Exploring the Use of Buffer Analysis for the Identification of Impacted Areas in Environmental Equity Assessment." *Cartography and Geographic Information Systems* 24(3):145–57.

Glickman, T. S., and R. Hersh. 1995. *Evaluating Environmental Equity: The Impacts of Industrial Hazards on Selected Social Groups in Allegheny County, Pennsylvania*. Discussion Paper 95–13. Washington, DC: Resources for the Future.

Hird, J. A., and M. Reese. 1998. "The Distribution of Environmental Quality: An Empirical Analysis." *Social Science Quarterly* 79(4):693–716.

Kriesel, W., T. J. Centner, and A. G. Keeler. 1996. "Neighborhood Exposure to Toxic Releases: Are There Racial Inequities?" *Growth and Change* 27:479–99.

McMaster, R., H. Leitner, and E. Sheppard. 1997. "GIS-Based Environmental Equity and Risk Assessment: Methodological Problems and Prospects." *Cartography and Geographic Information Systems* 24(3):172–89.

Mohai, P. 1995. "The Demographics of Dumping Revisited: Examining the Impact of Alternate Methodologies in Environmental Justice Research." *Virginia Environmental Law Journal* 14:615–53.

Mohai, P., and B. I. Bryant. 1992. "Environmental Racism: Reviewing the Evidence." In *Race and the Incidence of Environmental Hazards: A Time for Discourse*. Boulder, CO: Westview Press.

Oakes, J. M. 1997. "The Location of Hazardous Waste Facilities." Ph.D. Dissertation. Amherst, MA: University of Massachusetts Amherst.

Pastor, M. 2001. *Building Social Capital to Protect Natural Capital: The Quest for Environmental Justice*. Working Paper DPE-01–02. Amherst, MA: Political Economy Research Institute.

Rabin, Y. 1989. "Expulsive Zoning and the Legacy of Euclid." Pp. 101–21 in C. M. Haar and J. S. Kayden, eds., *Zoning and the American Dream: Promises Still to Keep*. Chicago, IL: Planners Press and American Planning Association in association with Lincoln Institute of Land Policy.

Szasz, A., and M. Meuser. 1997. "Environmental Inequalities: Literature Review and Proposals for New Directions in Research and Theory." *Current Sociology* 45(3): 99–120.

United Church of Christ. Commission for Racial Justice. 1987. *Toxic Wastes and Race in the United States: A National Report on the Racial and Socioeconomic Characteristics of Communities with Hazardous Waste Sites*. New York: United Church of Christ.

U.S. Department of Commerce. 1992. *1990 Census of Population and Housing: STF 3 on CD-ROM*. Washington, DC: Economics and Statistics Administration, Bureau of the Census.

———. 1994. *Geographic Areas Reference Manual*. Washington, DC: Economics and Statistics Administration, Bureau of the Census. Available at ⟨http://www.census.gov/geo/www/garm.html⟩.

U.S. Environmental Protection Agency. 1995. *User's Guide for the Industrial Source Complex (ISC3) Dispersion Models, Volume II—Description of Model Algorithms*. Research Triangle Park, NC: U.S. EPA. Available at ⟨http://www.epa.gov/scram001/userg/regmod/isc3v2.pdf⟩.

———. 1999. *Fact Sheet: Risk-Screening Environmental Indicators*. Washington, DC: Office of Pollution Prevention and Toxics. Available at ⟨http://www.epa.gov/oppt/rsei/docs/factsheet_v2.pdf⟩.

———. 2000. *Summary of 1998 Toxics Release Inventory Data*. Available at ⟨http://www.epa.gov/tri/tridata/tri98/data/1998datasumm.pdf⟩.