1-21-2004

Inequity Measures for Evaluations of Environmental Justice: A Case Study of Close Proximity to Highways in NYC

Jerry O. Jacobson
RAND Graduate School

Nicolas W. Hengartner
Yale University, Department of Statistics

Thomas A. Louis
Johns Hopkins University Bloomberg School of Public Health, tlouis@jhsph.edu

Inequity Measures for Evaluations of Environmental Justice: A Case Study of Close Proximity to Highways in NYC

Jerry O. Jacobson[1], Nicolas W. Hengartner[2], Thomas A. Louis[3]

First submitted: January 28, 2003

Resubmitted: January 21, 2004

[1] RAND Graduate School
Santa Monica, CA
jerryojacobson@runbox.com

[2] Department of Statistics
Yale University
nicolas.hengartner@yale.edu

[3] Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health
tlouis@jhsph.edu

Support provided by grant NSF SBR-9816560 from the U. S. National Science Foundation and the U.S. Environmental Protection Agency. Dr. Louis collaborated on this project while a Senior Statistical Scientist at RAND.

ABSTRACT

Assessments of environmental and territorial justice are similar in that both assess whether empirical relations between the spatial arrangement of undesirable hazards (or desirable public goods and services) and socio-demographic groups are consistent with notions of social justice, evaluating the spatial distribution of benefits and burdens (outcome equity) and the process that produces observed differences (process equity). Using proximity to major highways in NYC as a case study, we review methodological issues pertinent to both fields and discuss choice and computation of exposure measures, but focus primarily on measures of inequity. We present inequity measures computed from the empirically estimated joint distribution of exposure and demographics and compare them to traditional measures such as linear regression, logistic regression and Theil's entropy index. We find that measures computed from the full joint distribution provide more unified, transparent and intuitive operational definitions of inequity and show how the approach can be used to structure siting and decommissioning decisions.

1 INTRODUCTION

Assessments of environmental justice and equity are concerned with the distribution of burden of environmental hazards among socio-demographic and socio-economic groups. In practice, these assessments attempt to discern whether the spatial arrangement of hazardous sites, individuals, and communities are consistent with notions of social justice. The debate distinguishes between evaluation of equity in existing spatial distributions, sometimes called "outcome equity," and in the process that has given rise to them, or "process equity" (Fricker and Hengartner, 2001; Talih and Fricker, 2002). Since the landmark study by the United Church of Christ Commission for Racial Justice (1987), investigations of outcome inequity have employed a broad range of statistical methods and Geographic Information System (GIS) approaches to examine proximity or exposure to Toxic Release Inventory, Petrofund, Superfund, and Land Recycling sites (McMaster et al. 1997; Scott and Cutter, 1997, Chakraborty and Armstrong, 1997; Mitchell et al., 1999; Waller et al., 1997 and 1999; Fricker and Hengartner, 2001; Talih and Fricker, 2001), landfills and incinerators (Been, 1994; Liu, 1997), toxic storage and disposal facilities (TSDFs) (Oakes et al., 1996; Been and Gupta, 1997; Pastor, 2001), and accidental hazardous releases (Margai, 2001). Statistical methods have included bivariate tests (Sexten et al. 1993; Been, 1994), more sophisticated cross-sectional multivariate regressions (Fricker and Hengartner, 2001; Margai, 2001; Pastor, 2001), longitudinal comparisons (Oakes et al., 1996; Liu, 1997; Been and Gupta, 1997; Mitchell et al., 1999; Talih and Fricker, 2002) and Bayesian analysis (Waller et al., 1997). GIS methods for visualizing information relevant to assessments have developed in parallel. For example, Chakraborty and Armstrong (1997) demonstrate the attractiveness of geographic plume analysis compared with circular buffers for estimating exposure from a hazardous site. Scott and Cutter (1997) discuss methods for communicating risk from nearby hazardous facilities to communities. Several years ago, McMaster et al. (1997) presented an important summary of methodological problems with environmental equity assessments, based on considerable evidence at the time that findings were sensitive to the scale, resolution, and choice of outcome measure; they pointed out the importance of reaching a consensus on the most appropriate methodologies given the often dramatic shifts in findings when methods are altered slightly.

To date, however, no consensus regarding standardized analysis or an operational definition of equity is evident in either the GIS or statistical literatures. This paper extends this discussion by examining a range of commonly used inequity measures, and by introducing a new class of measures that attempts to add clarity and transparency to the operational definition of inequity by defining inequity measures explicitly in terms of the empirically estimated joint distribution of exposure to environmental hazards, race, income, and covariates of interest.[1] Among the advantages of this approach is the facility with which the measure of inequity can then be visualized and summarized numerically over all exposure levels—avoiding reliance on a dichotomous, "exposed or not", characterization of exposure. This is particularly useful when parties disagree on the threshold that should be used in analysis or when health impact from proximity to a hazard is unknown. Secondly, the measures presented here mirror conceptual notions of equity, a property lacking from traditional measures. For example, the most common approach in previous studies is to estimate a multivariate regression of a census tract-level indicator of exposure on a measure of the racial composition of the tract; however, this approach does not assess whether the burden of hazardous exposure is shared equally among subpopulations of interest, but rather whether largely minority *census tracts* are more likely to be exposed than others. The question of whether burden is shared equally among subpopulations can be more accurately evaluated by examining empirical adherence to the probability statement that exposure is conditionally independent of race, which is the approach we take here. This approach extends the work of Waller et al. (1997, 1999), who measured inequity as the difference between cumulative distribution functions (CDFs) of exposure for two subpopulations of interest. Of course, even perfect equity may not be desirable if all groups are equally exposed at unacceptably high levels.

A secondary aim of this paper is to demonstrate use of an outcome measure (i.e., a measure of exposure to environmental hazards at each location in the study area) that does not ignore cumulative exposure from multiple hazardous sources. Studies have typically included exposure only to the nearest hazard or to hazards within an individual's census area, though clearly census boundaries are irrelevant to the distance travelled by and diffusion of hazards over an area.

By way of a case study of proximity to major highways in New York City, we provide an example of such an exposure measure, but focus primarily on a comparison of several candidate *inequity* measures for environmental equity assessment, including generalized linear regression, logistic regression, Theil's

[1] We do not address the important and unresolved issues of scale, resolution, and

entropy index, and an inequity measure computed from the joint distribution of exposure and demographics, demonstrating the sensitivity of results to choice of measure. We find that while each has strengths and no single measure will suffice, measures framed in terms of the full joint distribution provide more transparent and intuitive operational definitions of inequity. We conclude by demonstrating that a clearly defined measure of inequity can be used to inform siting and remediation decisions.

We investigate close proximity to major highways, a line source rather than the more commonly studied point source. Since our purpose is an illustration of methods, we do not attempt a definitive assessment. For example, the highway system in NYC consists of a wide variety of road types and road usage. Living close to a cars-only road will produce exposures different from living close to the Cross Bronx Expressway or other roads used by heavy diesel. Wind patterns and other meteorological features also have influence, producing exposures that are a function of distance, direction and other factors. Importantly, for some people proximity to a highway is a convenience, not a drawback. Therefore, our use of a purely distance metric without differentiating road type and without including meteorology must be taken as illustrative. A complete assessment would take the foregoing factors into account and study sensitivity of conclusions to candidate methods of computing an exposure gradient. Finally, it is important to note that our analyses focus on proximity to highways and provide no information on other possible forms of environmental or social inequity or injustice.

---

visualization of risk.

2 RELATIONSHIP TO ASSESSMENTS OF TERRITORIAL JUSTICE

Assessments of environmental justice have much in common with assessments of "territorial justice" and throughout this paper we draw on previous applications found in both literatures to enrich the discussion. Just as environmental justice is concerned with the distribution of hazardous exposures among socio-demographic and socio-economic groups, territorial justice is concerned with the distribution of public goods and services among such groups, given their need (Boyne and Powell, 1991; Davies, 1968). Instead of examining hazardous exposures, studies of territorial justice have examined equity in proximity or access to parks, playgrounds, and recreational facilities (Mladenka, 1989; Talen, 1997; Talen and Anselin, 1998), city streets (Antunes and Plumlee, 1977), primary medical care practitioners (Knox, 1978), and expenditures on transit infrastructure (Boschken, 1998), policing and fire protection (Cingranelli, 1981; Bolotin and Cingranelli, 1983), and health and sanitation (Boyle and Jacobs, 1982), among others. Like environmental equity assessment, territorial justice assessment asks whether empirical relations between the spatial arrangement of goods (or "bads") and socio-demographic groups are consistent with notions of social justice. In both fields, assessments of outcome inequity typically fix a geographic study region within which one attempts to: (1) Assign an exposure (or access) measure to each individual, where the major difference lies in the source of exposure; (2) generate a measure of inequity based on the strength of the relationship between exposure and suspected socio-demographic covariates (e.g., race), adjusted for potential confounders (e.g., property value in the case of objectionable facilities, age structure in the case of public parks); and (3) determine whether the level of inequity is large enough to be important.

3 PROXIMITY TO HIGHWAYS IN NYC

While highways facilitate travel and commerce, they also expose people nearby to ambient risks, including vehicle emissions, noise, and acute obnoxious releases from traffic accidents involving hazardous materials (HAZMAT). The fine particulate air pollution released in emissions has recently been shown to increase the risk of lung cancer and cardiopulmonary mortality (Pope et al., 2000). Chronic exposure to road traffic noise is believed to cause stress and other forms of discomfort (Ouis, 2001). Data from 14 states indicate that 9% to 58% of HAZMAT releases in 1998 occurred in transit,[2] so that proximity to highways increases the risk of exposure to these incidents. Marjai (2001) examined equity in exposure to HAZMAT incidents, but did not focus on transport vs. fixed-facility incidents.

Examination of whether minorities and the poor are more likely than others to suffer the burden of close proximity to major highways is motivated by Marjai's (2001) finding that minorities and the poor tend to be closer to HAZMAT releases in New York state, Wallace's (1990) contention that NYC has pursued infrastructure policies detrimental to minority groups, and findings from elsewhere in the nation suggesting that the geographic allocation of transit infrastructure, public parks, fire and police protection, and other infrasructure and services often

[2] Based on the U.S. Department of Public Health's Hazardous Substances Emergency Events Surveillance (HSEES) 1998 dataset.

benefit white and wealthy areas more than others (Boshken, 1998; Alesina et al., 1999; Talen, 1997; Talen and Anselin, 1998; Cingranelli, 1981; Boletin and Cingranelli, 1981). It has also been noted that residential segregation provides an easy mechanism for discrimination in facility siting (Massey and Denton, 1993; Cutler and Glaeser, 1997) and that levels of residential segregation in the U.S. remain high (Massey and Denton, 1993, Cutler and Glaeser, 1997; Borjas, 1995).

For the most part, however, land use, zoning, and the location of the more than 187 miles of major highways and 6000 centerlane miles of divided roads in NYC have been established since the first half of the 20th century, when highways and railroads replaced waterways as the primary mode of transportation. The second and current zoning plan for the city has been in place since 1956. Consequently, any differences in close proximity to highways observed today are likely the result of some combination of population mobility and past inequitable siting decisions. Although Talih and Fricker (2002) show that there has indeed been considerable mobility among some minorities, such mobility, just as facility location, can be influenced by discriminatory social as well as economic factors (e.g., Massey and Denton, 1993). In this context, study of exposure to highways in NYC provides an opportunity to evaluate the impact of population mobility in the face of a recognized and fixed environmental hazard.

3.1 Data and Study Area

Our population of interest resides within the bounds of NYC, but we include portions of major highways from a larger surrounding area to guard against "edge effects," which could impact the findings if, for example, wealthy neighborhoods tend to be located in the inner city and poorer neighborhoods on the periphery or vice versa.

Our population demographic data come from the 1990 decennial census and are summarized in the first columns of Table 1. Data on the locations of primary U.S., state, and limited access interstate highways, based on Census Feature Class Codes, were extracted from ArcView 3.2. Highway segments were extracted as line segments of no more than 25m, small in length compared to our critical exposure distance described in the next section, and allowing interpretation of each segment as a point source of exposure located at the segment's midpoint. This implicitly assumes the geometric error of the road data is less than 25m, which we acknowledge may not be the case. All distance calculations are performed in the New York State Plane Long Island Zone coordinate system (Bugayevskiy and Snyder, 1995) .

Table 1 shows that in 1990, over 7.3 million people resided within the NYC five-county area, in which none of the race-ethnic groups could have claimed a majority and in which more than one in four residents who reported to the Census were foreign-born. Income levels and proximity to the nearest highway segment are also shown. Values in the table are computed assuming all individuals reside at the centroid of their block group and share their block group's median family income.[3] Though block groups boundaries are determined in part by population

---

[3] Denote the number of block groups by B, the number of individuals of a race-ethnic group within block group b by n_b and the median family income by I_b. Then, the tabulated mean is the weighted average: $N^{-1}\sum_{b=1}^{B} n_b I_b$, where $N = \sum_{b=1}^{B} n_b$ is the group's total population within the study area.

size, they are connected, geographic regions and so are appropriate units of analysis. Under these assumptions, earnings by non-Hispanic whites were nearly twice those of Hispanics and more than 1.6 times those of African Americans; immigrants earned about 9% less on average than others. The last columns of Table 1 show about one in five residents within 200m of the nearest highway and about 3.3 in 5 residents within 800m. If we compare the mean distance to a highway for the three income groups, we see no evidence that proximity is solely a function of income, as might be expected on purely economic grounds. Further, residents in the top third of the income distribution are closer by 28m than residents in the bottom third on average. The relation persists in the tail of the distribution.

Figure 1 displays the entire study area with NYC block groups shaded and major highways, including those included in the analysis but outside the NYC area, drawn in black. The unshaded block groups are excluded from the analysis either because their population or land area were reported as zero.

*** Table 1: Demographics and proximity to highways, about here ***

*** Figure 1: Study area, about here ***

4 COMPUTING EXPOSURE

Ideally, the measure used to characterize the environmental hazard at each location should account for distance, direction, toxicity and quantity emitted from each source, health impact and other relevant factors. Since our purpose is in comparing inequity measures and not a definitive assessment, for convenience we compute exposure as a function of proximity, but retain the term *exposure* throughout for generalizability of the discussion.

Prior environmental justice studies have generally used one of three approaches to compute exposure. The most common approach characterizes exposure as the count of hazardous facilities in a census area or as a dichotomous indicator of the presence of a facility (Been and Gupta, 1997; Liu, 1997; Oakes et al., 1996; Fricker and Hengartner, 2001). A problem with this approach is that dispersion of pollutants does not honor census boundaries so that edge effects are a real concern. Fricker and Hengartner (2001) smooth demographic attributes over neighboring census tracts to address this problem. A second approach defines the outcome measure as the distance from the census area to the nearest site (Waller et al., 1997 and 1999; Margai, 2001), which obviates edge effects, but ignores the possibility of exposure from multiple nearby sources. A third method employed by Talen (1997), but not in the environmental justice arena, counts the number of sites within a given radius; Talen demonstrated that findings resulting from this measure can be sensitive to the choice of radius.

A more flexible metric has employed in a small number of examinations of equity in access to desirable facilities such as parks and hospitals, relies on a "gravity" model to compute cumulative exposure, which in the environmental context we call the exposure gradient (Knox, 1980; Geertman and Van Eck, 1995; Talen and Anselin, 1998). Compared with the measures described above, the gradient method has the advantage of tallying impacts from multiple sources in a way that decays flexibly with distance; a drawback is that a decay parameter must be specified. But, it can be made to reflect the true decay of the hazard of interest if the information is available.

4.1 The Exposure Gradient

The gradient measured at geographic location i can be expressed as a distance-weighted average over all sites j in the assessment domain:

$$E_i = \sum_j w(d_{ij}) I_j \tag{1}$$

where E is exposure, d is the separation distance, $w(d)$ is a weight function decreasing in distance, and I is the exposure value when d = 0. Summing contributions to exposure from multiple sites in the study area recognizes that the impact of hazards may be cumulative, of particular importance when facilities may be located next to one another. In the present case, a given census area may be proximate to one or multiple highway segments (e.g., near an interchange where multiple highways meet).

Ideally, the gradient should be tuned to a fine geographic resolution, but in practice exposure computed at each census area, such as the tract or block, is used as an exposure surrogate for all individuals within each area. The weights used in the gradient should reflect the diffusion process of the pollutants. For example, if the diffusion is isotropic, the weight will be a function of the Euclidian distance between the point sources and the location at which the gradient is evaluated. More general diffusion processes will give rise to weights that are functions of meteorological and/or percolation distances.

4.2 The NYC Gradient

In the highway application, for purposes of illustration we use (1) with

$$w(d) = L\left(\frac{1}{2}\right)^{\frac{d}{200}}$$

and $I_j = 1$, which decays by 50% at 200m.[4] L is the length of the road segment. This parameterization allows contributions from all sources, but weights segments within close proximity more heavily. Figure 2(a) displays the exposure distribution within race and ethnic groups (normalized by dividing by the maximum over all block-groups) and is similar to those produced by Waller et al. (1997) and Waller et al. (1999). The plot shows the fraction of residents in the study area that suffer exposure greater than E, with E on the horizontal axis. For example, 20% of Hispanics live in block-groups where the exposure gradient is greater than 0.4. Hispanics and Asians are more exposed than others over much of the distribution, while at lower levels African Americans are the least exposed. At any level, Figure 2(b) shows that immigrants suffer greater exposure; but, the summaries by income (Figure 2(c)) are less clear and would produce different rank orderings depending on the critical value chosen.

*** Figure 2: Exposure distributions, about here ***

5 INEQUITY MEASURES

[4] Our choice of 200m is based loosely on Pearson et al. (2000), who identified an association between traffic and cancer at a 750ft separation. However, we are interested in the broader class of factors that make close proximity undesirable (e.g., pollution, noise). Since preferences for this bundle of undesirables are unknown, but certain to vary among the population, we allow contributions to the exposure function from somewhat farther away to avoid excluding road segments that may be too close for some individuals' comfort This parameterization is chosen more to allow us to proceed with illustration of the method than for substantive considerations.

An inequity measure is a quantitative summary that conveys the extent of inequity in the distribution of exposure between subpopulations of interest. The class of inequity measures presented in Section 5.7 equates statistical independence or conditional independence between exposure and demographics as "perfect equity," with degree of inequity computed as degree of departure from independence. Therefore, the joint distribution of exposure and population attributes is the starting point for computing inequity and the full panoply of statistical association measures are available to the analyst.

Our general strategy for measuring inequity is to compute the magnitude of the departure of the joint distribution of exposure, group, and any adjusting factors (denoted by E, G, and W, respectively) from conditional independence of E and G, given W (e.g., independence of exposure and race, given income). Inequity measures derived from the joint distribution have the desirable property that their magnitude does not depend on total population size (though statistical inference will depend on sample size).

Let $f(e,g,w)$ be a joint density or mass function and $\Delta(f)$ a measure—which we have yet to define—of departure from independence. Let us say further that $\Delta(f) = 0$ is equivalent to conditional independence of E and G, given W (i.e., perfect equity). Different numeric and graphical summaries of $\Delta(f)$ will reveal distinct aspects of the geographic configuration and any one summary is unlikely to convey all important aspects of inequity. For example, averages of $\Delta(f)$ over w for each group g reveal the level of inequity among groups without regard to covariates (eg., inequity for each race group without regard to income) and may be appropriate to relate an inequity measure to a known excess incidence of disease in a given group

g_i, while finding the maximum value of $\Delta(f)$ over g and w can identify subpopulations for whom proposed public policies are most inequitable.

We distinguish between local, group-specific, and global inequity measures. *Local* measures evaluate inequity for fixed e, whereas *group-specific* measures generate a summary inequity value for each group. Graphs of local measures (as a function of exposure) inform on fine-grained aspects of inequity. *Global* measures summarize departures from independence across all values of e, e.g., a weighted average of a local measure. The marginal density of the distribution of exposures in the population and the dose-response function for a health outcome are two natural candidates for weights. Group-specific measures have direct relevance for setting public policy by helping to assess the environmental burden on demographic groups. Global measures are appropriate for testing the null hypothesis of environmental equity.

With respect to controlling for a potential cofounder (e.g., income), most attractive are inequity measures that reduce in magnitude as the set of adjusting factors increases. We call such measures *explicative* in that

$$\Delta(E,G \mid W_1) \geq \Delta(E,G \mid W_1, W_2) .$$

See Efron (1978) for a general class of such measures for binary data. The multiple R^2 for linear regression is explicative; however, as shown below, it does not satisfy the desirable property of always increasing with regressive transfers.

We now discuss and compare four specific inequity measures. The first are regression and logistic regression, both common to assessments of environmental and territorial justice. The third is Theil's entropy index (Theil, 1967; Conceicao, Bradford, and Galbraith, 2000), which has the advantage of examining the entire exposure distribution rather than the mean, as in regression. The final measure is defined explicitly in terms of the joint distribution, drawing on the preceding discussion.[5] Each measure addresses a different aspect of inequity and generally comparison of several measures better informs policy by identifying areas of agreement or disagreement.

5.1 The Regression Approach

The landmark United Church of Christ (1987) study and those that immediately followed relied on bivariate analysis, conducting, for example, t-tests of mean differences in racial composition between geographic areas deemed exposed and others. More recent studies have used multivariate regression models to adjust for confounders (Been and Gupta, 1997; Fricker and Hengartner, 2001; Margai, 2001). For example, Fricker and Hengartner (2001) control for population density, arguing that the factor is likely to influence both siting decisions and the residential clustering of demographic groups. Generally, the regression of exposure on race reports on the statistical significance of race in setting the mean exposure. The dependent variable can be a person-specific exposure, but is typically a measure aggregated over individuals (e.g., at the tract level). Regressors can be a combination of those at the finest level of disaggregation (e.g., the individual) or at a higher level of aggregation (referred to as location-specific).

While the regression approach has attractive properties and is familiar, in its standard form it looks only at departures from zero correlation (rather than

[5] These measures are selected for their popularity in the environmental justice literature. Other possible inequity measures include Moran's I and other measures of spatial association (Anselin, 1995; Ord and Getis, 1995; Getis and Ord, 1992).

statistical independence), only at expected values (rather than the whole distribution) and works best when the dependent variable is approximately Gaussian. Transforms or Generalized Linear Models can deal with departures from normality and percentile regression (Kottas and Gelfand, 2001) can explore dependencies more flexibly, but we do not consider these approaches here.

The coefficient of determination, or R^2, produced by regression models would at first appear to be attractive as a global, summary measure of inequity. Commonly used to summarize how well the covariates account for variation in the dependent variable, R^2 could be used to assess how well demographics explain exposure. R^2 is also explicative. However, R^2 can be deceptive in comparing levels of inequity due to its sensitivity to functional form. Specifically, R^2 is not an appropriate measure for evaluating inequity in existing or alternative configurations of objectionable facilities if any of the configurations results in a non-linear relationship. To see this, consider the five hypothetical block groups in Table 2.

*** Table 2: R^2 example, about here ***

The relation is linear, with intercept=1 and slope=5. Residents in neighborhoods with a greater concentration of African Americans are more exposed and an Ordinary Least Squares fit produces $R^2 = 1$. Now, consider a new site that raises the exposure of the fifth block group from 5.0 to 10.0, but leaves the others unchanged. R^2 reduces to 0.80 because the fit is no longer linear. However, intuitively there is more, rather than less, inequity in the new situation because the additional exposure falls entirely on the group that was more exposed initially.

An alternative regression-based inequity measure is the predicted increase in exposure due to a higher presence of a particular group in an area, e.g.,

$$\frac{E[\text{Exposure} \mid \%\text{African American} > x_A]}{E[\text{Exposure} \mid \%\text{African American} \leq x_A]}$$

relative to the same ratio computed for a reference group (e.g., %white > x_w). This comparison is analogous to a logistic regression and can be estimated as such if we are willing to dichotomize exposure. The threshold x similarly dichotomizes group presence and may differ by group, depending on the suspected mechanisms at work. Ratios for the regression prediction can also be computed at high and low percentiles of the distributions, rather than at the expected value. However, all of these measures are quite ad hoc and thresholds derived from expected harm, as in Waller et al. (1999) would be preferable.

5.2 Regression Results for NYC

Table 3 presents results from regressions computed at the census tract and block group levels. In each case, the natural logarithm of the normalized exposure gradient is predicted by the demographic composition of the population and geographic attributes of the census area. Demographic covariates include the fraction of the population that is immigrant, African American (black), Asian or Pacific Islander, and Hispanic, and median family income. Three potential confounders are also included: log(population), log(housing density), and a binary indicator for whether the census area includes a body of water. The latter is a natural barrier for the placement of highways and could also be related to demographics if, for example, the wealthy seek residence in areas on the coast or with a lake-side view.

To demonstrate the flexibility of the regression approach, we incorporate interaction terms and splines. Interactions with income test the hypothesis that the effect of racial and immigrant concentration on exposure depends on the area's income level. The splines examine whether the effect of demographic composition on exposure is nonlinear. For example, a two-segment, linear spline is constructed to model the effect of the concentration of blacks on exposure, with the "knot" set at 50%; in Table 3, the coefficients corresponding to "black s1" and "black s2" refer the slope of the first segment (0-50% black) and second segment (51-100% black) of the spline, respectively. The spline is specified separately for census areas above and below the median of median family income. Figure 3 helps to visualize the two resulting splines that relate concentration of blacks to exposure. Similar splines are constructed for other groups, and for income, with its knot at the median of median family income among census areas.

*** Figure 3: Splines, about here ***

Ideally, it would be desirable to estimate a model similar to those in Table 3 in which the observations represent individuals rather than census areas. Such a model could be used to test a distinct set of hypotheses related to person-level rather than area-level exposure outcomes; for example, whether whites are more exposed than others. However, there are difficulties in estimating such a model from aggregated Census data: constructing individual-level indicators of race, immigrant status, and income would require counts for all combinations of attributes (e.g., the number of white immigrants with income X per block group). Census data provide some, but not all of the necessary cross-tabulations. Second, without knowledge of each individual's location within the census area, individual-level exposure cannot be computed. One way forward is to allow the dependent variable and some covariates to enter at the area level, with others at the individual level where possible; but this complicates interpretation and violates the assumption that the observations are conditionally independent given the covariates. We do not estimate such a model here.

Table 3 shows that the findings are sensitive to the geographic resolution, replicating findings discussed in McMaster et al. (1997). See Openshaw (1984) for a discussion of this issue, sometimes referred to as the modifiable area unit problem.

Both models support similar conclusions with respect to immigrants, for whom effect sizes are largest: more concentrated immigrant communities are more exposed on average, irrespective of income. However, a more subtle relationship is revealed with the help of the flexible regression model. Specifically, recall that the splines permit estimation of a two-part trend line, which identifies slopes in the relationship between demographics and exposures in two cases: in census areas where immigrants do, and do not, hold a majority. These two-part splines are further interacted with income, so that two-part trend lines are estimated for census areas above, and separately for areas below, median income (Figure 3). The estimates for these spline terms indicate that in lower income areas, exposure increases with immigrant concentration at a higher rate in areas when immigrants hold a majority than when they do not (i.e., Pct immigrant s2 is greater than Pct immigrant s1); but this acceleration reverses at higher income. In contrast, the spline estimates for African Americans reveal that a higher fraction of African Americans is related to increased exposure, but that in census areas where they are the majority, the stronger the majority the less the census area is exposed (i.e., black s1 is positive, but black s2 is negative). In both models the effect of income depends on the demographic group: the main income effect is either not statistically significant or is negligible relative to other effects.

Note that at the tract level, one coefficient is not estimable because the design matrix is not sufficiently informative.

*** Table 3: Regression results, about here ***

5.3 The 2 x 2 Table and Logistic Regression

If being below or above a specific exposure level is of interest, a dichotomous indicator can be used as the dependent variable. The basic approach to assessing departures from independence for a single, dichotomous attribute is to form the 2 x 2 table with columns for group (e.g., Black/White) and rows indicating whether exposure is above and below the threshold *e*. Logistic regression generalizes this approach, allowing the full flexibility of the previously described regression approach.

Statistics such as the log-odds ratio, Kendall's τ, and the uncertainty coefficient can be used to quantify and test for departure from independence. These inequity measures are local in that they are a function of the cut-point e and a variety of threshold cut-points can be assessed. Talen's (1997) finding that a one- versus two-mile radius reversed the finding of white, non-white differences in proximity to parks in two U.S. cities demonstrates that care must be taken in choosing the cut-point.[6] The choice should be guided by knowledge about the harm of various levels of exposure. Where there is ambiguity several cut-points should be investigated.

5.4 Logistic Regression Results for NYC

Table 4 illustrates a similar sensitivity with two logistic regressions on the NYC highway data at the block group level, using the same covariates as in Table 3, but differing in the cut-point of the dependent variable; at the 25th (low exposure) and 75th (high exposure) percentiles of the distribution of the natural logarithm of the normalized exposure gradient.[7] The 75th percentile is distinct in that only socio-economic factors are associated with exposure, suggesting substantial inequity at high exposures. Also note that the rank ordering of effect sizes across demographic groups depends on the cut-point. For example, Asian presence is more strongly associated with exposure than immigrant presence in low income areas at the 75th exposure percentile, but the reverse is true at the 25th percentile.

[6] Talen (1997) uses bivariate Mann-Whitney tests rather than a logistic model, but the caution is equally relevant.

[7] An individual-level logistic regression cannot be estimated using these data because there is insufficient variation in the relation between regressors and the dichotomous outcome.

The models do agree for some assessments. For example, that exposure increases with immigrant presence. However, dichotomizing reduces the information content in the exposure measure, making the immigration and other coefficients nearly inestimable, as we can see from the large coefficients and standard errors in Table 4.

Differences of orders of magnitude in the effect sizes in Table 4 motivate attention to outliers. A map of the location of racial and ethnic majorities in the study area (not shown) revealed two small clusters of Asian or Pacific Islander communities, one comprising 28 block groups in south Manhattan, the other slightly larger and in southwest Brooklyn. An analysis revealed that the median leverage[8] of the Manhattan cluster is more than 13 times that of other block groups in the study area. When the block groups in this cluster (less than 0.1% of the total) are removed, the large odds ratio on "Pct Asian s2" drops from 37.8 to 0.31 and becomes non-significant.

*** Table 4: Logistic regression results, about here ***

5.5 Gini and Theil

The foregoing regression approaches, while informative, are limited in that they do not assess the relation of the full exposure distribution to covariates. However, a rich set of measures of distribution dispersion has developed from the study of income inequality. Economists often measure income inequality by summarizing the fraction of the population that holds a given fraction of income. In

[8] Pregibon's (1981) leverage statistic

1997, for example, 20% of households held 49% of national income. If we observed that 50% of individuals held 50% of income, we would have no reason to suspect income inequity. The analogy to the share of environmental burden carried by subpopulations of interest is clear.

The idea of shares leads to the Gini coefficient, which begins with a plot of the population CDF against the income CDF, supporting statements like the one above at every percentile. The Gini coefficient is computed as the ratio of the area between the 45 degree line representing perfect equity and the Lorenz curve—the curve showing the actual distribution of exposure—to the total area below the Lorenz curve, the maximum deviation from equity possible.

Theil's T (hereafter *TT*) is a less intuitive measure, but has more desirable properties. It is: Lorenz-consistent, so that it produces the same rank ordering as the Gini; mean independent, so that it is insensitive to scalar multiplications of exposure; and satisfies the Pigou-Dalton property, so that it increases with regressive transfers, which as we noted earlier is not a property of regression's R^2. Importantly, the Gini coefficient and *TT* are decomposable (Bhattacharya and Mahalanobis, 1967; Conceicao et al., 2000). The contribution to inequity from each in a mutually exclusive and exhaustive set of groups can be assessed and contributions add.

In assessing environmental equity the decomposability property has other attractive consequences. In addition to assessing the relative magnitude of inequity within and between groups, decomposability allows comparison of the between-group *TT*s under competing grouping frameworks (e.g., race vs. class). This comparison is equivalent to evaluating which partitioning (e.g., by race or class) yields more information (more of the observed inequity). This is analogous to the usual interpretation of R^2 in the regression context, but under TT there is no need to specify a model per se. On the other hand, developing the partitioning scheme requires discretizing continuous covariates, such as income.

With e_i exposure for individual i, $i = 1,2,\ldots, n$, compute

$$\mu = \frac{1}{n}\sum_{i=1}^{n} e_i$$
$$r_i = \frac{e_i}{\mu}$$
$$TT = \sum_{i=1}^{n} r_i \ln r_i$$

Then, TT is the entropy of the share of exposure relative to the uniform distribution and is bounded by $TT \in [0, \ln(n)]$. To decompose TT into contributions from a mutually exclusive and exhaustive set of groups g_j, with $j = 1,2,\ldots,m$, and each with membership n_j, compute $p_j = \frac{n_j}{n}$, $R_j = \frac{\mu_j}{\mu}$, and

$$T_j = \frac{1}{n_j}\sum_{i \in g_j}^{n_j} r_i \ln r_i$$
$$TT = \sum_{j=1}^{m} p_j R_j \ln R_j + \sum_{j=1}^{m} p_j R_j T_j$$

where μ_j is the group's mean exposure. The two terms of TT give the between-group and within-group inequity, respectively (Conceicao and Galbraith, 1998).

5.6 Results for Theil's T

For the NYC highway data, the inequity in exposure between all individuals (i.e., with no partitioning) is TT = 0.45, equivalent to a population in which $1/\exp(TT)$ = 64% of the population suffers all of the exposure and the remaining 36% none (Conceicao et al., 2000).

Figure 4 compares the between-group *TT*s under different partitionings by race-ethnicity (Hispanic, white, African American, Asian or Pacific Islander, and other), family income, and immigrant status. To compute the *TT* inequity between individuals from Census data, we assign individuals the exposure, income, and percentage immigrant values of their block groups. We dichotomize income (above/below the study area median) and immigrant status (above/below the study area median percentage immigrant). The horizontal axes indicate the partitioning scheme (e.g., $r \times r$ partitions by race-ethnicity only; $r \times m$ partitions by race-ethnicity and immigrant status, yielding 10 groups; $r \times i \times m$ is a three-way interaction, which yields 20 groups). The vertical axis shows the between-group *TT*s as a fraction of the total inequity, 0.45. Figure 4 shows that of the three factors, differences in race-ethnicity contribute the most inequity and income alone almost none. Partitioning by income in combination with other attributes yields more inequity, but even a decomposition by all three factors recovers just 4% of the inequity in exposure among individuals.

*** Figure 4: Theil between-group iniquity, about here ***

5.7 Comparing Conditional Distributions

Regressions and Theil's *TT* support only a partial investigation of the conditional independence assumption between race and exposure given controls that lies at the heart of our analysis. One problem is that these measures can be zero even when the conditional independence assumption fails to hold. A more sensitive investigation results from direct comparison of the empirical joint distribution and the one implied by the conditional independence model. One such measure, which is easy to calculate, is

$$\Delta(e, g \mid w) = F(e \mid g, w) - F(e \mid w).$$

Perfect equity is equivalent to $\Delta(e,g,w) = 0$ and departures from 0 measure inequity. Plots of the $\Delta(e,g|w)$ against e for fixed racial subgroups g and covariates w provide a visual tool to assess departure from equity for a particular subpopulation. A negative difference occurs when a larger fraction of the subpopulation than the population at large has exposure greater than e. Thus, the relative exposure burden experienced by group g is decreasing in Δ.

Consider Figure 5(a), which plots $\Delta(e,g|w)$ versus e for the NYC highway data, with g fixed at white, African American, Hispanic, or Asian or Pacific Islander, and w below the median family income. In Figure 5(b), w is above the median family income. The figures show departures of the joint distribution of exposure and race-ethnic group from the independence model, conditional on income. Since the curves for the white and African American subpopulations lie above the y-axis, these groups are less exposed than the average New Yorker, whereas Hispanics and Asians are more exposed than average, as suggested from the bivariate summaries in Table 1. Conditioning on median income does not change the conclusion. However, class differences are apparent by noting that poorer African Americans are generally more exposed than poorer whites, whereas the reverse is true at higher income; similarly, in Figure 5(a) Asians are typically less exposed than Hispanics, while the relation is split at higher income. Figures

5(a) and 5(b) demonstrate once again sensitivity of assessments to e: the inequity curves cross at several points.

Though the plots are revealing, they should be backed up by numerical evaluation of departure from independence. The signed maximal deviation

$$D(g \mid w) = \operatorname{sign}(\max_e \Delta(e, g \mid w)) \cdot \max_e \left|\Delta(e, g \mid w)\right|$$

is one among many reasonable choices, where negative values of $\Delta(g|w)$ imply relatively more exposure. A second candidate is the area under the curve,

$$D(g \mid w) = \int \Delta(e, g \mid w) r(e) de$$

weighted by a function $r(e)$ that measures the health risk of exposure at or less than e. If such an exposure-health response function is available, its conditional expectation can be used to measure differential health outcome among groups (see Waller et al., 1999). Though informative, neither of the above is explicative.

Lacking dose-response information, a more generic computation is needed. Of the many metrics between densities, the total variation,

$$\Delta = \int \left\{ \sum_g \int \left| f(e, g \mid w) - f(e \mid w) f(g \mid w) \right| de \right\} f(w) dw \qquad (2)$$

is easiest to interpret. It corresponds to the smallest fraction of individuals that need to be relabeled (moved) so as to achieve perfect equity. Dividing an individual summand in (2) by the probability, pr(G = g) produces a group-specific measure. Dividing the integrand with respect to e by pr(E = e) (or the related density) provides an exposure-specific inequity measure. Removing the absolute value produces a signed measure, but without the relabeling interpretation.

*** Figure 5: Relative exposure distributions by income, about here ***

5.8 Comparison of Findings

Identifying areas of agreement among measures strengthens support for specific hypotheses and identifying areas of disagreement underlines sensitivity of findings to method. Under all measures examined, differences in income explain little of differences in exposure. Rather, income effects enter through interactions with race, ethnic, and immigrant group. These interactions in the regression context are large, up to 27 times higher than other covariates. The plots comparing the joint distributions show that income affects the rank ordering of groups by exposure, with larger differences among the poor. In contrast, adding income to Theil partitionings by race or immigration yield only 0.1-0.2% more of the total inequity.

Evidence that wealthier individuals tend to be less exposed to major highways is mixed. The joint distributional plots support this hypothesis for African Americans at low levels of exposure, but some of the regressions indicate that wealthier individuals are in fact more exposed. For Hispanics, the individual-level regression supports the hypothesis, but other specifications and the logistic model do not.

All regressions support a positive link between immigrants and exposure that accelerates as immigrants move past a majority presence. Immigrants are the only group examined for which the result that the wealthy are less exposed than others is strongly supported in some specifications and *not refuted* in others. Finally, the regressions and Theil measure agree that race, income, and immigration explain little of the variation in exposure. The R^2s and pseudo-R^2s are 11-12%, but drop to about 6% when non-demographic covariates are excluded; similarly, the Theil between-group inequity is just 4% of the total inequity between individuals. Although we have not yet developed formal measures to quantify explanatory power for the joint distribution measures, Figures 4(a) and 4(b) suggest deviations from perfect equity are generally 5% or smaller, and at most 10% when comparing poorer Hispanics to others.

Non-linearities in the relationship between exposure and demographics have been revealed by examining majorities and are discussed in previous sections.

6 SITING AND REMEDIATION

In addition to assessing environmental equity, inequity measures can be used to structure a formal approach to siting and decommissioning of exposure sources. Waller et al. (1999) introduce computing spatial "isopleths" to help visualize the inequity that would result from locating a new facility or exposure source at various locations in the study area. In this sense, narrowly defined, an isopleth is a contour that identifies a set of locations where placing a new source would produce a constant value of inequity. Maps of isopleths can be revealing and aid in the siting or decommissioning process, and can be used to incorporate considerations of inequity with other criteria of concern to decision-makers and the public.

When different stakeholders prefer competing measures of inequity, compromise can be facilitated by first normalizing each measure by its optimum value (e.g., percent increase over the optimum). Then, overlay the normalized isopleths produced separately for each measure and note where they intersect; these are regions where the competing measures agree on the inequity outcome. Locations may be excluded from consideration in order to satisfy any other constraints on the decision. Performance of this approach depends on the measures and the method of computing difference from the optimum.

6.1 NYC Isopleths

Figures 6(a) and 6(b) show isopleths for the NYC data computed under two inequity measures: (a) the between-group, fully-decomposed Theil and (b) the R^2 measure produced by the regression in Table 3. Starting from a lattice superimposed over the study area with a grid spacing of 500m, at each lattice point the inequity measure (e.g., Theil or R^2) is recomputed, assuming a single, additional exposure source were located at that point. For purposes of illustration, we assume a 1000m highway segment is to be located with its center at some lattice point in the study area. The 1000m can be interpreted as representing a new road segment or a policy that would increase traffic in the same location by an exposure-equivalent amount. The isopleths, contours of identical inequity values that would result from siting the new segment at each point, are generated from the inequity values using an interpolation function, in this case *filled.contour* in the statistical package “R” (see Cleveland, 1993).

The isopleths for the two inequity measures indicate approximately the same locations for the best (least inequity) and worst (most inequity) candidate sites. However, the measures disagree on the relative inequity that would result from placing the road segment in many locations. Since constraints will usually eliminate the best location from contention, differences in these two measures are sufficiently large to affect siting decisions.

*** Figure 6: Isopleths, about here ***

7 DISCUSSION

Our formal approach to assessment of environmental equity cannot capture the full complexity of the issues; it is an adjunct and not a replacement for proper process and dialogue. However, a formal approach does clarify and communicate assumptions and goals. Basing all assessments on the joint exposure, demographics distribution helps to clarify the notion of inequity that is being examined and unifies assessments, but remains sensitive to choice of exposure and inequity metrics. By combining graphical and quantitative summaries, what may be overly high impact GIS displays can be tempered.

Importantly, the approach allows adjustment for the effects of covariates that may confound the relation between exposure and the attribute of interest. For example, the approach allows evaluation of the share of exposure linked to economic variables, quantifying the exposure cost of being poor.

If the data are available to do so, our approach builds from exposures to individuals rather than exposures to census-delineated aggregates, but the method can be adapted to whatever level of geographic and demographic detail is available.

Despite applying models substantially more flexible than those used in other studies, none of the measures employed explained a large fraction of the relation among exposure, race and income, indicating that relative proximity to highways is the result of complex social processes. However, we do identify immigration and Hispanic ethnicity as strong correlates of close proximity to highways; in a related study, Fricker and Hengartner (2001) found that Hispanics in NYC in 1990 were more exposed to a broad range of environmentally undesirable facilities. We also find that class in NYC is strongly associated with proximity to major highways, but the effect differs by demographic group. For example, we find that in majority immigrant areas, high income is associated with less exposure. Notably, the *main* class effect in our regression models is insignificant; rather, it is the class-demographic interactions that best describe how differences in income relate to observed variation in proximity to highways. This suggests that future environmental justice studies—as they have not done in the past to the best our knowledge—should be careful to test for the presence of interactions, so long as they are justifiable on theoretical grounds or whenever the goal of the study is to best explain variation. Further, by examining majorities, we produced strong evidence of non-linearities in the relation between exposure and demographics and found that small, outlying geographic clusters can exert considerable influence on inequity findings; such sensitivity requires additional investigation.

We should reiterate that while close proximity to highways may be environmentally hazardous and discomforting, being too distant may also be undesirable. We selected a decay parameter to reflect this trade-off, but more attention to the non-monotone relation should be a focus of future assessments.

Further development of the measures presented here and others is needed. Inferential tests need to be adapted to inequity assessment to account for uncertainty in estimating joint distributions and additional case studies and computer simulations can provide a better understanding of the descriptive and inferential performance of the proposed measures.

Results of our formal analyses are dependent on the quality and relevance of inputs, on the computation of exposure and demographic attributes, on the measure of departure from statistical independence (perfect equity), on the geographic scale and resolution of the assessment. These high-leverage choices must be clearly communicated and sensitivity to reasonable modifications of them evaluated.

REFERENCES

Alesina A, Baqir R, Easterly W, 1999, "Public Goods and Ethnic Divisions" *The Quarterly Journal of Economics* **39** 1243 - 1284

Anselin L, 1995, "Local Indicators of Spatial Association-LISA" *Geographic Analysis* **27** 93 - 115

Been V, 1994, “Locally Undesirable Land Uses in Minority Neighborhoods: Disproportionate Siting or Market Dynamics?” *Yale Law Review* **103** 1383 - 1422

Been V, Gupta F, 1997, “Coming to the Nuisance or Going to the Barrios? A Longitudinal Analysis of Environmental Justice Claims” *Ecology Law Quarterly* **24** 1 - 56

Bhattacharya N, Mahalanobis B, 1967, “Regional disparities in household consumption in India” *Journal of the American Statistical Association* **62** 143 - 161

Bolotin F L, Cingranelli D L, 1983, "Equity and urban policy: the underclass hypothesis revisited." *The Journal of Politics* **45** 209 - 219

Borjas G J, 1995, "Ethnicity, Neighborhoods, and Human-Capital Externalities" *The American Economic Review* **85** 365 - 390

Boshken H L, 1998, "Upper-middle class influence on developmental policy outcomes: the case of transit infrastructure" *Urban Studies* **35** 627 - 647

Boyle J, Jacobs D, 1982, "The Intracity Distribution of Services: A Multivariate Analysis" *The American Political Science Review* **76** 371 - 379

Boyne G, Powell M, 1991, "Territorial justice: A review of theory and evidence." *Political Geography Quarterly* **10** 263 - 281

Bugayevskiy L M, Snyder, J P, 1995, *Map Projections: A Reference Manual* (Taylor and Francis, London)

Chakraborty J, Armstrong M P, 1997, “Exploring the Use of Buffer Analysis for the Identification of Impacted Areas in Environmental Equity Assessment” *Cartography and Geographic Information Systems* **24** 145-157

Cingranelli D L, 1981, "Race, politics, and elites: testing alternative models of municipal service distribution." *American Journal of Political Science* **25** 664 - 692

Cleveland, W S, 1993, *Visualizing Data*. (Hobart Press, New Jersey)

Conceicao P, Bradford P, Galbraith J K, 2000, “The Theil Index in Sequences of Nested and Hierarchical Grouping Structures”, University of Texas Inequality Project, Working Paper No. 15

Conceicao P, Galbraith J K, 1998, “Constructing Long and Dense Time-Series of Inequality Using the Theil Index”, University of Texas Inequality Project, Working Paper No. 1

Cutler D M, Glaeser E L, 1997, "Are Ghettos Good or Bad?" *The Quarterly Journal of Economics* **37** 827 - 872

Davies B P, 1968 *Social Needs and Resources in Local Services* (Michael Joseph, London)

Efron B, 1978, “Regression and ANOVA with zero-one data: Measures of residual variation" *Journal of the American Statistical Association* **73** 113 - 121

Fricker R D, Hengartner N W, 2001, “Environmental Equity and the Distribution of Toxic Release Inventory and Other Environmentally Undesirable Sites in Metropolitan NYC” *Environmental and Ecological Statistics* **8** 33 - 52

Geertman S C M, Van Eck J R, 1995, "GIS and models of accessibility potential: an application in planning" *International Journal of Geographical Information Systems* **9** 67 - 80

Getis A, Ord J K, 1992, "The Analysis of Spatial Association by Use of Distance Statistics" *Geographical Analysis* **24** 189 - 206

Knox P L, 1980, "Measures of accessibility as social indicators: a note" *Social Indicators Research* **7** 367 - 377

Kottas A, Gelfand A, 2001, “Bayesian semiparametric median regression modeling" *Journal of the American Statistical Association* **96** 1458 - 1468

Liu F, 1997, “Dynamics and Causation of Environmental Equity, Locally Unwanted Land Uses, and Neighbourhood Changes” *Environmental Management* **21** 643 - 656

Margai F L, 2001, “Health Risks and Environmental Inequity: A Geographical Analysis of Accidental Releases of Hazardous Materials” *Professional Geographer* **53** 422 - 434

Massey D S, Denton N A, 1993 *American Apartheid: Segregation and the Making of the Underclass* (Harvard University Press, Cambridge, MA)

McMaster R B, Leitner H, Sheppard E, 1997, “GIS-based Environmental Equity and Risk Assessment: Methodological Problems and Prospects” *Cartography and Geographic Information Systems* **24** 172-189

Mitchell J T, Thomas D S K, Cutter S L, 1999, “Dumping in Dixie Revisited: The Evolution of Environmental Injustices in South Carolina” *Social Science Quarterly* **80** 229-243.

Mladenka K R, 1989, "The distribution of an urban public service: the changing role of race and politics" *Urban Affairs Quarterly* **24** 556 - 583

Oakes J M, Anderton D L, Anderson A B, 1996, “A Longitudinal Analysis of Environmental Equity in Communities with Hazardous Waste Facilities” *Social Science Research* **25** 125 - 148

Openshaw S, 1984, “The Modifiable Area Unit Problem” *Concepts and Techniques in Modern Geography* **38**

Ord J K, Getis A, 1995 "Local Spatial Autocorrelation Statistics: Distributional Issues and an Application" *Geographical Analysis* **27**

Ouis D, 2001, “Annoyance from Road Traffic Noise” *Journal of Environmental Psychology* **21** 101 - 120

Pastor M, 2001, “Racial/Ethnic Inequality in Environmental-Hazard Exposure in Metropolitan Los Angeles” *California Policy Research Center Brief* **13**

Pearson R L, Wachtel H, Ebi K L, 2000, "Distance-weighted traffic density in proximity to home is a risk factor for leukemia and other childhood cancers" *Journal of the Air & Waste Management Association* **50** 175 - 180

Pope C A, Burnett R T, Thun M J, Calle E E, Krewski D, Ito K, Thurston, G D, 2002, "Lung Cancer, Cardiopulmonary Mortality, and Long - term Exposure to Fine Particulate Air Pollution" *Journal of the American Medical Association* **287** 1132 - 1141

Pregibon D, 1981, "Logistic Regression Diagnostics" *Annals of Statistics* **9** 705 - 724

Scott M S, Cutter S L, 1997, "Using Relative Risk Indicators to Disclose Toxic Hazard Information to Communities" *Cartography and Geographic Information Systems* **24** 158-171

Stata Corporation. Intercooled Stata 7.0 for Windows 98/95/NT. Copyright (C) 1985 - 2001

Talen E, 1997, "The social equity of urban service distribution: an exploration of park access in Pueblo, Colorado, and Macon, Georgia" *Urban Geography* **18** 521 - 541

Talen E, Anselin L, 1998, "Assessing spatial equity: an evaluation of measures of accessibility to public playgrounds" *Environment and Planning A* **30** 595 - 613

Talih M, Fricker R D, 2002, “Effects of neighbourhood demographic shifts on findings of environmental injustice: a New York City case-study” *Journal of the Royal Statistical Society A* **165** 375 - 397

Theil H, 1967, *Economics and Information Theory* (Rand McNally and Company, Chicago, IL)

United Church of Christ Commission for Racial Justice, 1987 *Toxic wastes in the United States: A national report on the racial and socioeconomic characteristics of communities with hazardous waste sites* (United Church of Christ, New York)

Wallace, R, 1990, “Urban Desertification, Public Health and Public Order: `Planned Shrinkage', Violent Death, Substance Abuse and AIDS in the Bronx” *Social Science and Medicine* **31** 801 - 813

Waller L A, Louis T A, Carlin B P, 1997, “Bayes methods for combining disease and exposure data in assessing environmental justice” *Environmental and Ecological Statistics* **4** 267 - 281

Waller L A, Carlin B P, Xia H, Gelfand A E, 1998, “Hierarchical spatio-temporal mapping of disease rates” *Journal of the American Statistical Association* **92** 607 - 617

Waller L A, Louis T A, Carlin B P, 1999, “Environmental justice and statistical summaries of differences in exposure distributions” *Journal of Exposure Analysis and Environmental Epidemiology* **9** 56 – 65

Table 1. Demographics and Proximity to Highways in the New York City Study Area

	N*	Fraction of population (%)	Median family income†	Mean distance to a highway (meters)	Fraction of population within … meters of a highway (%)		
					200m	400m	800m
Race & Ethnicity							
All	7,322,291	100.0	34,375	732	20.6	40.8	66.7
White+	3,178,546	43.4	44,602	775	18.7	39.2	65.8
Black	1,874,827	25.6	27,173	779	19.9	36.6	61.1
Hispanic	1,737,885	23.7	23,333	638	24.0	46.3	72.4
Asian	496,287	6.8	34,500	620	23.4	47.1	73.8
Other	34,746	0.5	29,191	709	23.0	41.5	67.2
Immigration							
U.S.-born	5,239,392	71.6	35,571	759	19.6	39.3	65.3
Foreign-born	2,082,899	28.4	32,361	666	23.2	44.4	70.2
Income[a]							
Top third	2,489,702	34.0	n/a	711	20.0	43.0	70.2
Middle third	2,417,740	33.0	n/a	748	19.7	37.9	65.3
Bottom third	2,414,849	33.0	n/a	739	22.2	41.3	64.5

* Number of individuals

† Computed by assigning the median family income of each block group to all individuals within its bounds.

+ White, black, Asian, and "other" include non-Hispanics only.

[a] The tertiles of median block group family incomes (in 1990$) are:

Top third: 0-26,806; middle third: 26,806-41,985; bottom third: 41,985-150,001

Table 3. Multivariate Regression Models of Proximity to Highways in New York City at Different Geographic Resolutions

Dependent Variable is the Log of the Normalized Exposure Gradient (d0=200m)

	Tracts	Block Groups
Log(pop density)	**-0.315 (0.157)**	-0.035 (0.098)
Log(housing unit density)	**0.351 (0.136)**	**0.170 (0.085)**
On water	0.277 (0.249)	**0.685 (0.232)**
Income s1 †	0.000 (0.000)	0.000 (0.000)
Income s2	**0.000 (0.000)**	**0.000 (0.000)**
Pct immigrant s1	**1.208 (0.517)**	**1.098 (0.288)**
Pct immigrant s2	**6.478 (1.337)**	**6.271 (0.760)**
Pct black s1	**2.286 (0.546)**	**1.349 (0.348)**
Pct black s2	**-2.924 (0.844)**	**-2.700 (0.519)**
Pct Asian s1	-0.008 (0.880)	-0.897 (0.535)
Pct Asian s2	**4.679 (1.539)**	**1.816 (0.744)**
Pct Hispanic s1	**2.651 (0.535)**	**2.232 (0.346)**
Pct Hispanic s2	-0.734 (0.904)	-0.495 (0.460)
Pct immigrant s1 x I[>medinc][b]	1.062 (0.672)	**0.862 (0.413)**
Pct immigrant s2 x I[>medinc]	**-9.282 (2.609)**	**-6.607 (1.150)**
Pct black s1 x I[>medinc]	-1.141 (0.785)	0.603 (0.485)
Pct black s2 x I[>medinc]	2.242 (1.174)	0.615 (0.719)
Pct Asian s1 x I[>medinc]	**2.803 (1.213)**	**3.177 (0.702)**
Pct Asian s2 x I[>medinc]	non-est[c]	**-6.970 (1.139)**
Pct Hispanic s1 x I[>medinc]	0.907 (0.796)	0.069 (0.481)
Pct Hispanic s2 x I[>medinc]	-1.057 (3.236)	1.342 (0.995)
Intercept	**-3.828 (0.390)**	**-4.094 (0.236)**
N	2,173	5,666
R^2	11.5%	10.7%

Entries with $p < .05$ are shown in bold; Standard error in parentheses

†The knot in the income spline is set at the median of median family income. For the pct immigrant and pct race splines, the knot is set at 50%.

[a] All 0.000 values are <1E-5

[b] "medinc" is median family income; I[>medinc] =1 if the census area's median family income is greater than the median median family income of all census areas in NYC.

[c] cannot be estimated

Table 2. Five hypothetical block groups with a linear relationship between exposure and demographics.

Block group	Exposure	Percent African American
1	1.0	0.0
2	2.0	0.2
3	3.0	0.4
4	4.0	0.6
5	5.0	0.8

Table 4. Odds ratios for logistic regression models of proximity to highways in New York City at two cut-points of the exposure gradient

Dependent Variable is the 25th or 75th Percentile of the Log of the Normalized Exposure Gradient		
	25th Percentile	75th Percentile
Log(pop density)	0.968 (0.076)	1.046 (0.095)
Log(housing unit density)	**1.181 (0.077)**	1.017 (0.079)
On water	**2.256 (0.534)**	1.417 (0.341)
Income s1	1.000 (0.000)	1.000 (0.000)
Income s2	**1.000 (0.000)**	**1.000 (0.000)**
Pct immigrant s1	0.951 (0.310)	**6.955 (2.401)**
Pct immigrant s2	**1.950E9 (5.570E9)**	**21.677 (22.039)**
Pct black s1	2.281 (0.913)	**3.323 (1.365)**
Pct black s2	**0.133 (0.070)**	**0.061 (0.039)**
Pct Asian s1	0.335 (0.243)	**0.144 (0.101)**
Pct Asian s2	0.214 (1.107)	**37.776 (56.071)**
Pct Hispanic s1	**9.525 (3.886)**	**3.309 (1.404)**
Pct Hispanic s2	1.070 (0.608)	1.334 (0.737)
Pct immigrant s1 x I[>medinc]	1.880 (0.871)	0.581 (0.291)
Pct immigrant s2 x I[>medinc]	**0.000 (0.000)**	0.119 (0.212)
Pct black s1 x I[>medinc]	**5.685 (3.622)**	1.169 (0.684)
Pct black s2 x I[>medinc]	0.374 (0.316)	0.552 (0.539)
Pct Asian s1 x I[>medinc]	**47.861 (48.469)**	**51.904 (49.374)**
Pct Asian s2 x I[>medinc]	0.000 (0.000)	0.000 (0.000)
Pct Hispanic s1 x I[>medinc]	0.447 (0.289)	**3.039 (1.878)**
Pct Hispanic s2 x I[>medinc]	8.058 (16.742)	2.193 (3.239)
N	5,666	5,666
Pseudo R^2	8.34%	5.89%

Standard errors in parentheses; notation as in Table 3

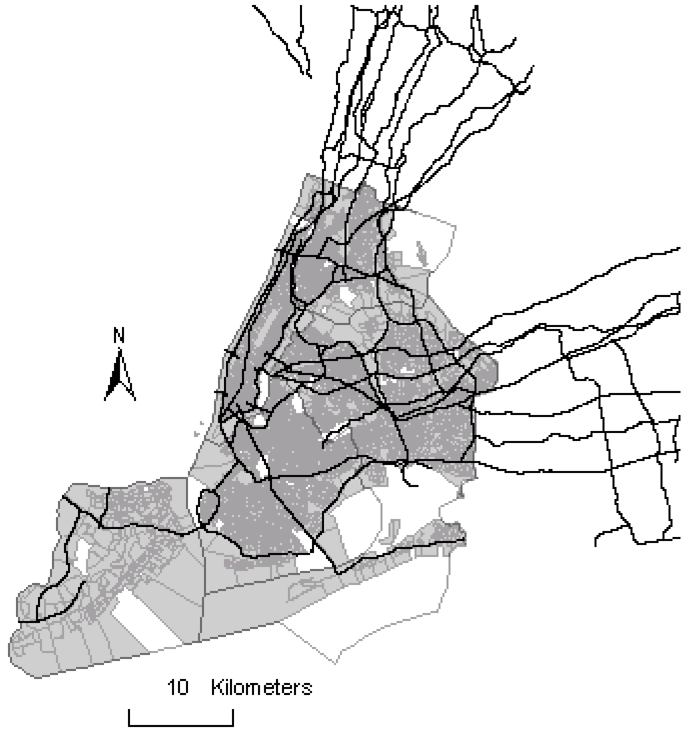
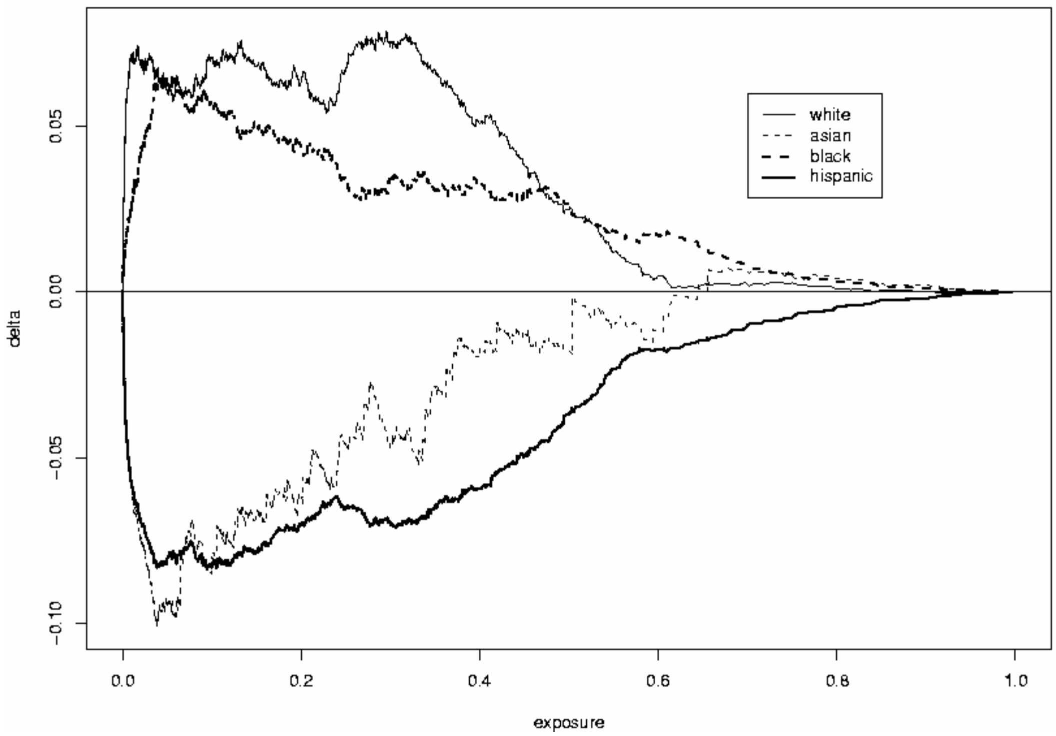


Figure 1. The New York City study area and major surrounding highways. Block groups included in the analysis are shaded. Unshaded block groups reported zero population or land area.

(a) Below median family income population



(b) Above median family income population

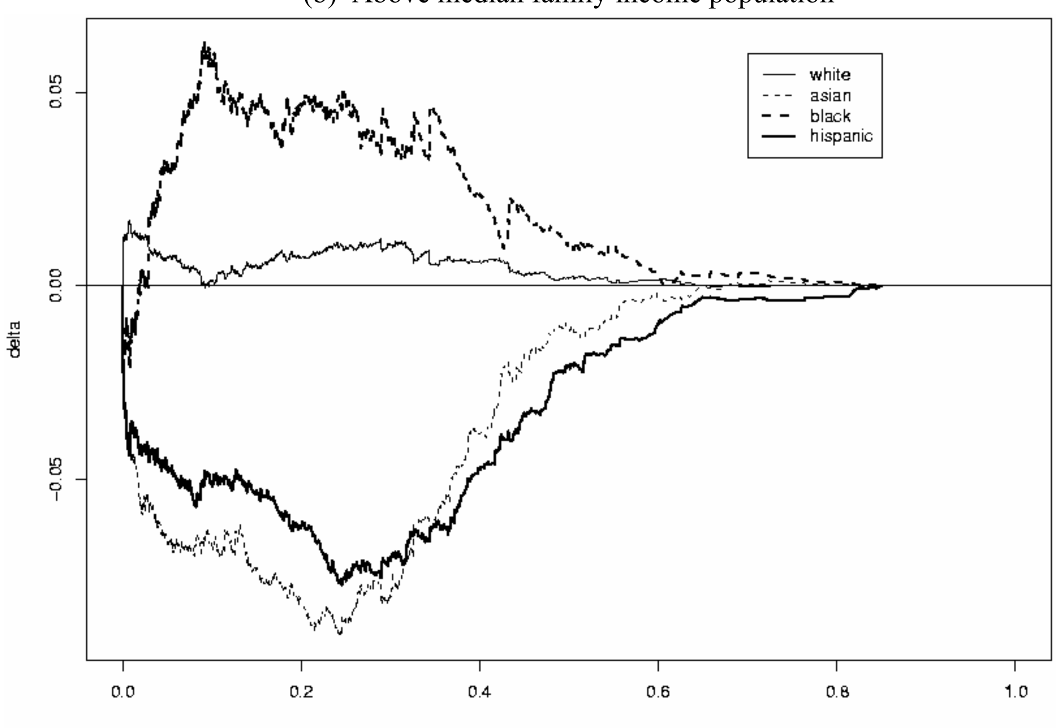


Figure 5. Departures of the joint distribution of exposure and race-ethnic group from the independence model, conditional on income. Delta is the fraction of individuals in each race-ethnic group who have exposure less than *e* (i.e., the exposure CDF) minus the same fraction for all individuals in the same income class.

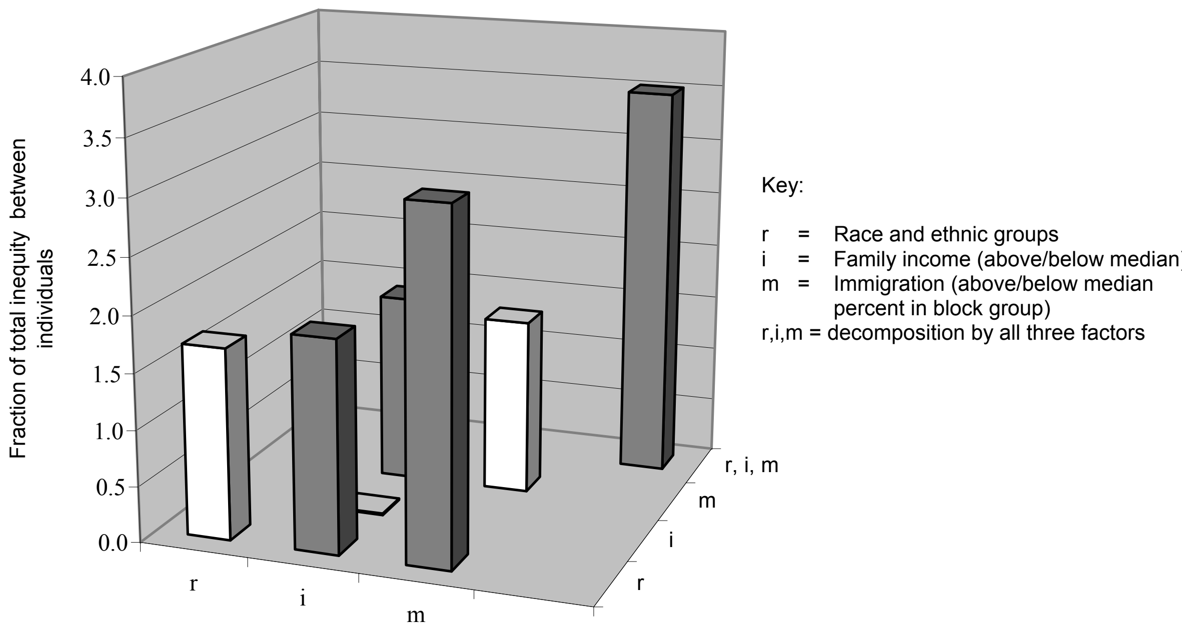
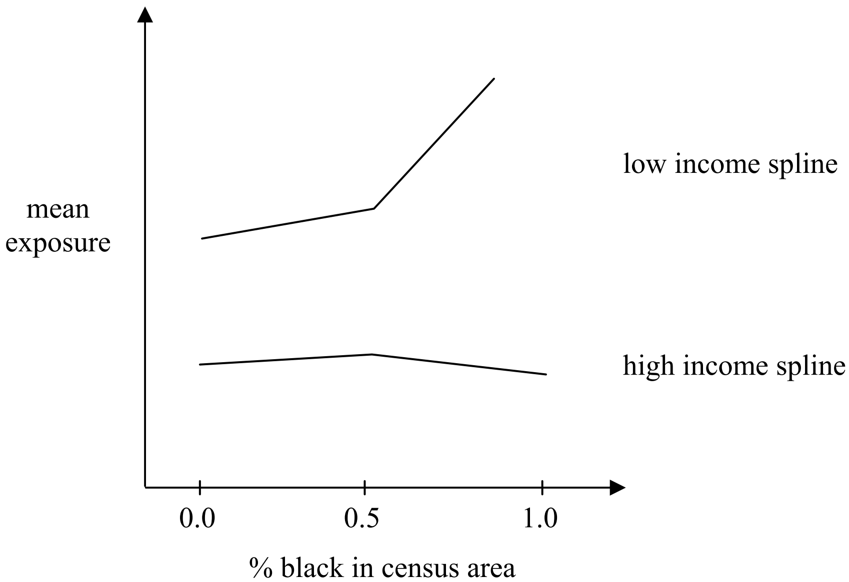


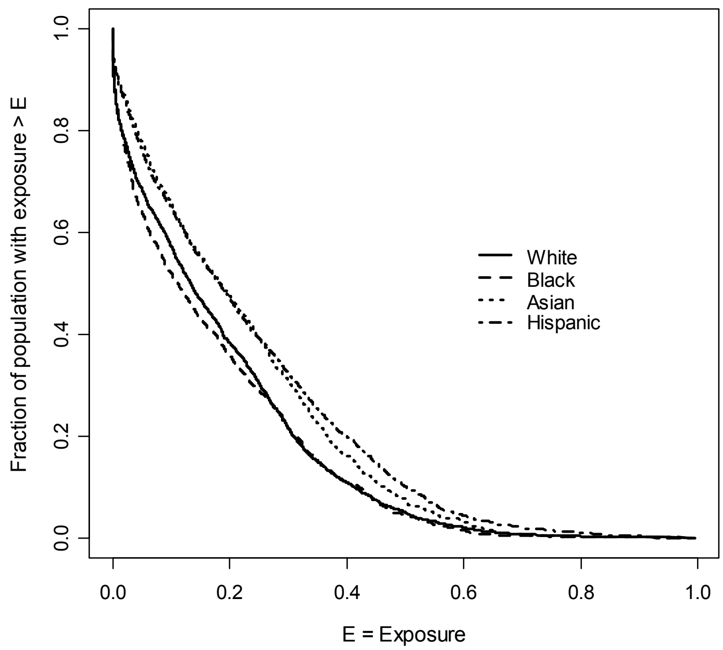
Figure 4. Theil's between-group *T* for exposure to highways, as a fraction of the total Theil inequity between all individuals (*TT*=0.45), under different partitioning schemes by race-ethnicity (r), family income (i), and/or immigrant status (m). For example, the second bar in the front row, *i x r*, shows that a decomposition by income, race, ethnicity yields 1.9% of the total inequity, while a decomposition by race and ethnicity alone (the bar to the left, *r x r*) yields 1.7% of the total inequity. Shading distinguishes single-factor decompositions (white) from decompositions based upon multiple factors (gray).

Figure 3. Example of a spline interacted with income estimated in the regression model*

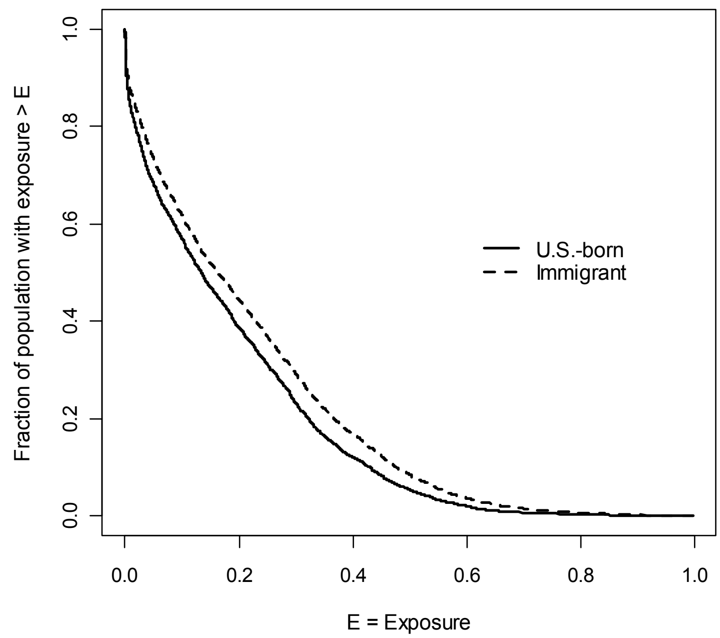


* Relationships shown are for illustration. They do not represent the true estimates.

(a)



(b)



(c)

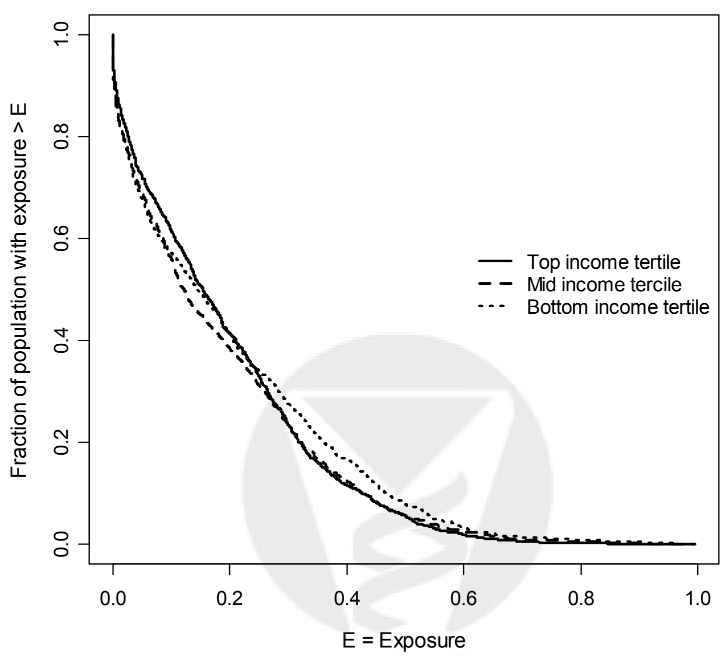


Figure 2. Distribution (1 – CDF) of the normalized exposure gradient by race and ethnicity (a), immigrant status (b), and family income (c).

(a) Theil

(b) Regression R^2

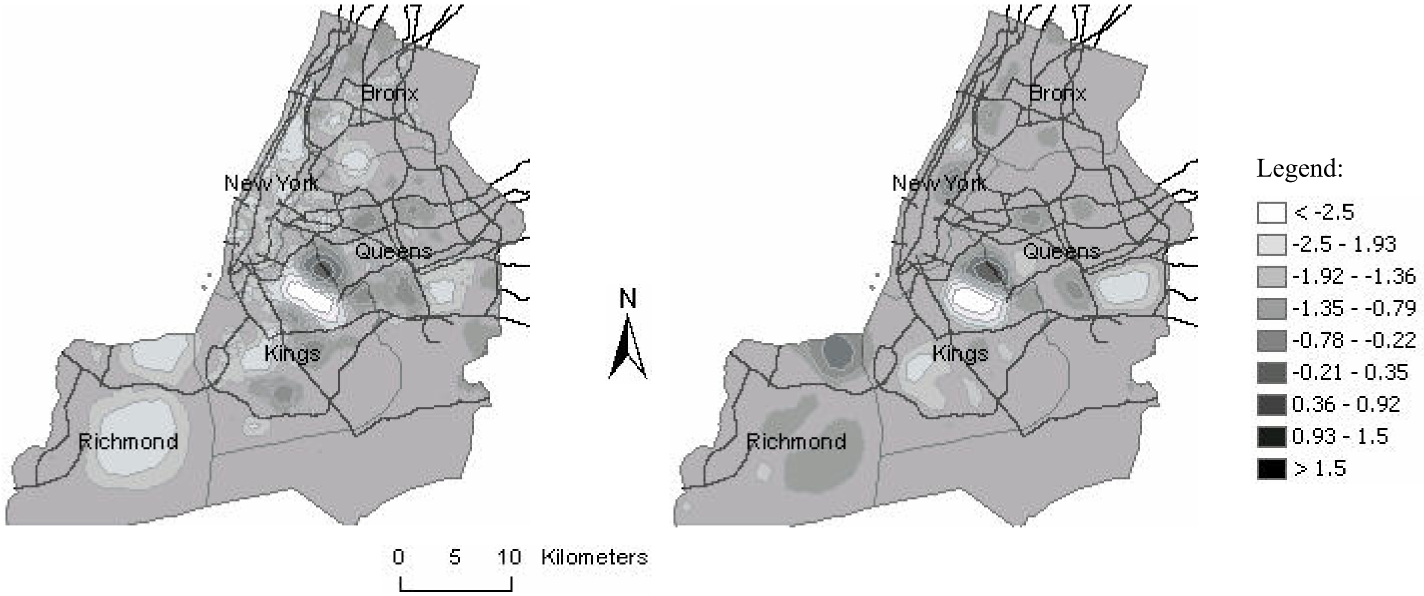


Figure 6. Isopleths of inequity in proximity to major highways using: (a) Theil's measure and (b) regression-based R^2 as measures of inequity. The values in the legend are standardized units of the respective inequity measure. The solid black lines are major highways.