Methodology for Distributional Analysis of a National Air Quality Rule

Abt Associates Inc.

Cambridge, MA
Lexington, MA
Hadley, MA
Bethesda, MD
Chicago, IL

November 17, 2008

Prepared for
Kelly Maguire
US EPA / NCEE

Prepared by
Ellen Post
Don McCubbin
Anna Belova
Jin Huang
Nate Frey

Work funded through
Contract EP-W-05-022
Work Assignments 2-39 & 3-73

Methodology for Distributional Analysis of a National Air Quality Rule

DRAFT

Table of Contents

List of Tables ii

List of Figures iii

1 Introduction 4

2 Methods 6
2.1 The Basic Steps in a Distributional Analysis 6
2.2 Conducting a Distributional Analysis in BenMAP 8
2.2.1 Deriving subgroup-specific distributions of baseline pollutant concentrations 10
2.2.2 Estimating subgroup-specific reductions in incidence rates of health effects 11
2.2.3 Characterizing inequality using inequality measures and Lorenz curves 13
2.3 Assessment of Uncertainty 14

3 A Case Study: EPA's Heavy Duty Diesel Rule in 2030 15
3.1 Projected Population Distributions and Baseline $PM_{2.5}$ Concentrations 16
3.2 Population-Weighted Average Reductions in $PM_{2.5}$ and Corresponding Reductions in Health Effects Incidence 19
3.2.1 Subgroup-specific reductions in $PM_{2.5}$ concentrations 23
3.2.2 Subgroup-specific health effects 24
3.3 Other Characterizations of Differences Among Subgroups 24
3.3.1 Comparing other characteristics of subgroup-specific distributions 25
3.3.2 Comparing subgroups using Lorenz curves and inequality measures 28
3.3.3 Comparisons by Region 31

4 Discussion 34

Appendix A. Methodological Details of Distributional Analysis of Benefits of the Heavy Duty Diesel Rule 37
A.1. Overview of a Typical Benefit Analysis 37
A.2. Methodological Details of the Heavy Duty Diesel Distributional Analysis 38
A.2.1 Air Quality 38
A.2.2 Population Forecast for 2030 39
A.2.3 Concentration-response functions 41
A.2.4 Baseline incidence data 42
A.2.4.1 All-cause mortality 42
A.2.4.2 Hospitalization 43
A.2.4.3 Emergency room visits for asthma 47
Appendix A.3: Definitions and Properties of Inequality Measures 49

Appendix B. Relationship between Grid Cell and Census Blocks 51

References 52

List of Tables

Table 1. U.S. Population (in *millions*) by Race and Ethnicity. 15
Table 2. Predicted Racial and Ethnic Group-Specific Population-Weighted Average Baseline $PM_{2.5}$ Concentrations in 2030 19
Table 3. Absolute and Relative Reduction in Mean $PM_{2.5}$ Concentrations and Incidence of *Non-Fatal Acute Myocardial Infarction* (per Million Population) 20
Table 4. Absolute and Relative Reduction in Mean $PM_{2.5}$ Concentrations and Incidence of *Asthma-Related ER Visits* Among Children (0-17) (per Million Population) 21
Table 5. Absolute and Relative Reduction in Mean $PM_{2.5}$ Concentrations and *Hospitalizations* (per Million Population) 21
Table 6. Absolute and Relative Reduction in Mean $PM_{2.5}$ Concentrations and Incidence of *All-Cause Mortality* (per Million Population) 23
Table 7. 2030 Projected Baseline Annual Average $PM_{2.5}$ Concentrations (ug/m^3) by Race and Ethnicity 27
Table 8. 2030 Projected Reductions in Annual Average $PM_{2.5}$ Concentrations (ug/m^3) by Race and Ethnicity 27
Table 9. Generalized Entropy (GE) Indicator of Inequality in Baseline $PM_{2.5}$ Concentrations in the Total Population and in Racial/Ethnic Subgroups 29
Table 10. Atkinson Index of Inequality in Baseline $PM_{2.5}$ Concentrations in Total Population and in Racial/Ethnic Subgroups 30
Table 11. Generalized Entropy (GE) Indicator of Inequality in *Reduction* in $PM_{2.5}$ Concentrations Due to the HDD Rule in the Total Population and in Racial/Ethnic Subgroups 31
Table 12. Atkinson Index of Inequality in *Reduction* in $PM_{2.5}$ Concentrations Due to the HDD Rule in the Total Population and in Racial/Ethnic Subgroups 31
Table 13. Absolute and Relative Reduction in $PM_{2.5}$ Concentrations and Hospitalizations (per Million Population) for All Cardiovascular Illnesses (Except Myocardial Infarctions) Among Adults, Ages 18 – 64: Regional Results 32
Table 14. Air Quality Metrics for HDD Rule 39
Table 15. Demographic Groups and Variables Available in BenMAP 40
Table 16. Epidemiological Studies Used and Estimated Cases of Adverse Health Effects 42
Table 17. Mortality Rates for All-Cause Mortality, by Age Group and Race 43
Table 18. Hospitalization Rates by Endpoint, Race and Age Group 46
Table 19. Emergency Room Visit Rates for Asthma, by Region, Race, Ethnicity and Age Group 48
Table 20. Definitions of Inequality Measures (Reprint from Cowell (2005), p. 137) 49
Table 21. Properties of Inequality Measures (Reprint from Cowell (2005), pp. 66-67) 50

List of Figures

Figure 1. Illustration of an Air Quality Grid over the State of Florida 9
Figure 2. The Lorenz Curve 14
Figure 3. Forecasted 2030 Populations by Race and Ethnicity 17
Figure 4. Forecasted 2030 Baseline Ambient $PM_{2.5}$ Concentrations 18
Figure 5. Forecasted 2030 Reduction in Ambient $PM_{2.5}$ Concentrations (Baseline Minus Control) 18
Figure 6. Racial and Ethnic Group-Specific Distributions of Baseline $PM_{2.5}$ Concentrations 26
Figure 7. Racial and Ethnic Group-Specific Distributions of Reduction in $PM_{2.5}$ Concentrations 26
Figure 8. Subgroup-Specific Lorenz Curves for Baseline $PM_{2.5}$ Concentrations in 2030 29
Figure 9. Grid Cell and Census Blocks 51

1 Introduction

"Environmental Justice" (EJ) has become a pressing social, scientific and political issue over the last decade. The 1994 Executive Order 12898, *Federal Action to Address Environmental Justice in Minority Populations and Low-Income Populations*, requires agencies to perform EJ reviews of their programs, policies, and activities in order to determine their effects on minority and low-income populations. EPA defines "Environmental Justice" as the fair treatment and meaningful involvement of all people regardless of race, color, national origin, or income with respect to the development, implementation, and enforcement of environmental laws, regulations, and policies.[1] EPA further defines "fair treatment" to mean that no group of people should bear a disproportionate share of the negative environmental consequences resulting from industrial, governmental and commercial operations or policies. This definition provides very general guidance on the concept of environmental justice, but does not supply specifics and directions for applying this concept to EPA's programs and activities.

The two most common types of EJ research or distributional analysis are: (1) proximity-to-hazards studies and (2) exposure and health risk analysis. The first category of research evaluates how the distribution and proximity of hazards (e.g., Superfund sites, toxic emissions, and existing waste facilities) relate to community demographics (see Glickman and Hersh, 1995; Stretesky and Lynch, 1999; Davidson and Anderton, 2000; Hite, 2000; Mantaay, 2002). Residential proximity to a waste site or other hazard is often used as a surrogate measure for exposure to contaminants found at those sites. The second category of EJ research, exposure and risk analysis, examines the distributions of exposures and health risks among different socio-demographic groups (see Gwynn, et al., 2000; Apelberg, et al., 2005; Morello-Frosch and Jesdale, 2006; Levy, et al., 2007; Linder, et al., 2008).

In this document, we discuss a method for carrying out the second type of distributional analysis as we analyze the benefits of an air pollution control regulation. There are several potentially interesting EJ questions that our analysis attempts to answer. These questions address potential inequality in (1) *baseline* levels of pollutant exposure, (2) *reductions* in levels of exposure that are expected to result from a pollution rule or regulation, and (3) health benefits associated with the reductions in pollution levels. These types of questions can be summarized as follows:

- Are different socio-demographic subgroups being exposed to significantly different pollution levels?
- When a given rule is implemented, do different socio-demographic subgroups benefit differentially – i.e., do some groups enjoy significantly greater reductions in pollution levels than others?
- Do some groups enjoy significantly greater reductions in health risks as a result of a given rule or regulation?

[1] The definitions of "Environmental Justice" and "Fair Treatment" are from EPA website: http://www.epa.gov/region07/ej/definitions.htm

The answer to the first question may depend on the particular pollutant, or on the sources of the pollutant. For example, different socio-demographic groups may be exposed to significantly different levels of one pollutant but not another. The answer to the second question may also depend on the particular rule or regulation. The answer to the third question will depend, in addition, on the age distributions and baseline incidence rates (for the health effects in question) of the socio-demographic groups being compared.

Below we describe an analytical approach that can address all of these questions by examining and comparing the distributions of individual-specific exposure (or health risk) levels and/or changes in these distributions in different (non-overlapping) subgroups defined by age, sex, race, ethnicity, education and/or income. To illustrate our proposed approach to distributional analysis of air quality rules, we provide a case study where we apply our method using data from an analysis of EPA's Heavy-Duty Engine and Vehicle Standards and Highway Diesel Fuel Sulfur Control Requirements (HDD) Rule (U.S. EPA, 2000).

2 Methods

2.1 The Basic Steps in a Distributional Analysis

The first step in a distributional analysis is to define the subgroups to be compared. Subgroups in distributional analyses have most commonly been defined by race, ethnicity and/or income. Other categorizations (e.g., urban vs. rural, attainment area vs. non-attainment area) are also possible. Since our main objective is developing a method for conducting a distributional analysis and because of limited resources, we focus on race and ethnicity in our case study, with the recognition that the same method could be used with other categorizations

The next step in a distributional analysis is to select a unit of analysis. Distributional analyses have often taken as their unit of analysis some measure of community, such as county (e.g., Perlin, et al., 1995; Bowen, et al., 1995), neighborhood (e.g., Bullard, 1994), radial zone (e.g., Glickman and Hersh, 1995), Census tracts (e.g., Davidson and Anderton, 2000; Linder, et al., 2008; Apelberg, et al., 2005; Morello-Frosch and Jesdale, 2006) and zip codes (e.g., Lejano and Iseki, 2001). There are, however, some substantive problems with this approach.

First, defining a community is a non-trivial and subjective task. As noted above, different units of analysis have been used, varying substantially in size. Several studies have found that units of different size can generate different results (see Taquino, et al., 2002; Williams, 1999; Mennis, 2002). Moreover, because a community is made up of a mixture of people from different subgroups, we may misclassify people when examining whether one subgroup is exposed to more pollution than another. For example, if a community has 60% of its households below the poverty level and is categorized as a "low income" community on that basis, the 40% of households in that community above the poverty level are effectively mischaracterized. A similar problem arises in the categorization of "EJ communities" in terms of race or ethnicity.

We suggest that the ideal analysis would use the individual, rather than the community, as the unit of analysis. This avoids the issues raised above, although it raises an empirical problem in that we do not have air pollution estimates for each individual. Truly individual-specific estimates of air pollution exposure would require personal monitoring, which is not feasible for analyses of large numbers of people. As a result, we simplify and assume that individuals living in close proximity to each other are exposed to the same air pollution levels. We discuss this further below.

Our proposed individual-based method for distributional analysis calculates an empirical distribution of baseline pollutant exposures for each defined subgroup of interest. A subgroup-specific distribution describes the frequency with which each pollutant concentration is experienced by members of the subgroup. Thus, for example, for all pollutant concentrations ($q_1,q_2,\ldots,q_Q$) to which members of the subgroup are exposed, we calculate the following:

$$f_q=n_q/N, \qquad (1)$$

where n_q is the number of individuals who are exposed to pollutant level q and N is the total number of individuals in the subgroup.[2] Correspondingly, f_q is the frequency with which exposure to pollutant level q occurs in the subgroup (i.e., the proportion of the subgroup exposed to pollution level q). A distribution of the changes in pollution exposures is similarly defined.[3] In addition to the proportion of the population, f_q, exposed to a specific pollutant level, q, we may be interested in the proportion of the population that is exposed to pollutant levels that are no more than q. Using the notation above, this is[4]

$$F_q = \sum_{i=0}^{q} \left(\frac{n_i}{N} \right) = f_0 + f_1 + f_2 + \ldots + f_q \,. \qquad (2)$$

The next step in a distributional analysis is selecting measures of comparison and means of presentation. There are several ways to compare the baseline pollutant levels to which individuals in one subgroup are exposed versus those in another subgroup – and similarly several ways to compare reductions in pollutant levels experienced as a result of a rule or regulation. Ultimately, a distributional analysis compares two or more subgroup-specific distributions – of baseline pollutant concentrations and of reductions in pollutant concentrations that result from a rule or regulation. Distributions can be compared using a single summary statistic – e.g., by comparing the means of the distributions. Alternatively (or in addition), they can be compared using several summary statistics – e.g., the means as well as several percentiles of the distributions. These comparisons can be shown via maps, graphically, and/or in tabular form. In addition, several measures of inequality have been developed and applied in distributional analysis such as the Generalized Entropy indicator and the Atkinson index.[5]

In our case study, presented below, we illustrate comparisons of subgroup means (i.e., mean *baseline* pollutant exposures and mean *reduction* in pollutant exposures) as well as more extensive comparisons using, in addition to means and standard deviations, selected percentiles of the distributions. We show comparisons via maps as well as graphically and via tables of results. Finally, we apply two measures of inequality that are particularly well suited to this kind of distributional analysis.

---

[2] As we note below, whereas ideally we would have truly individual-specific pollutant exposure, in practice we must approximate this by assigning grid cell-specific pollutant concentrations to all individuals in a grid cell. This may obscure some intra-grid cell variability and possibly dampen differences among EJ subgroups. This problem could be mitigated, however, to the extent that we can reduce the size of the grid cells used in the analysis.

[3] The frequency, f_q, may alternatively be thought of as a probability – the probability that a (randomly selected) individual in the subpopulation is exposed to pollution level q. Another name for the distribution described by equation (1) is therefore a probability density function (pdf).

[4] Alternatively, this may be thought of as the probability that a (randomly selected) individual in the population is exposed to no more than pollution level q. Another name for the distribution described by equation (2) is a cumulative distribution function (cdf).

[5] We present details about inequality measures in Section 2.2.3 and 3.3.2.

To compare subgroup-specific health benefits resulting from implementation of a rule or regulation, we could estimate individual-specific decreases in health risk – i.e., for each individual, the decrease in probability of incurring a given adverse health effect – and thus derive subgroup-specific distributions of these individual-level risk decreases analogous to the distributions of individual-level reductions in pollutant exposures. However, a more intuitive measure of health benefit for a subgroup is the decrease in incidence rate of an adverse health effect, a population-level measure that is calculated by summing the individual-specific decreases in risk. For example, suppose that implementation of a rule reduces each African-American's risk of a given adverse health effect from 3 in 10,000 to 2 in 10,000. Thus for every 10,000 African-Americans exposed to the pollutant in question, implementation of the rule will result in 1 fewer case of the adverse health effect, or a decrease in the incidence rate of 1 per 10,000 population.[6] To compare the health benefits enjoyed by different subgroups, then, we compare these subgroup-specific reductions in incidence rates instead of comparing distributions of individual-specific health risk reduction.

The discussion above describes the basic steps involved in a distributional analysis – (1) identifying subgroups of interest to compare, (2) selecting a unit of analysis by which to compare the results across subgroups, and (3) selecting measures of comparison and means of presentation of these comparisons. The basis of the comparison is the derivation of the subgroup-specific distributions and the calculation of reductions in health effect incidence rates, which are explained in Section 2.2 below.

2.2 Conducting a Distributional Analysis in BenMAP

An individual-based distributional analysis may be considered a variant on a standard benefit analysis, in which, instead of considering the total population, we consider the populations of each of the delineated subgroups in our analysis separately – i.e., we effectively carry out a benefit analysis separately for each subgroup and then compare the subgroup-specific results.[7]

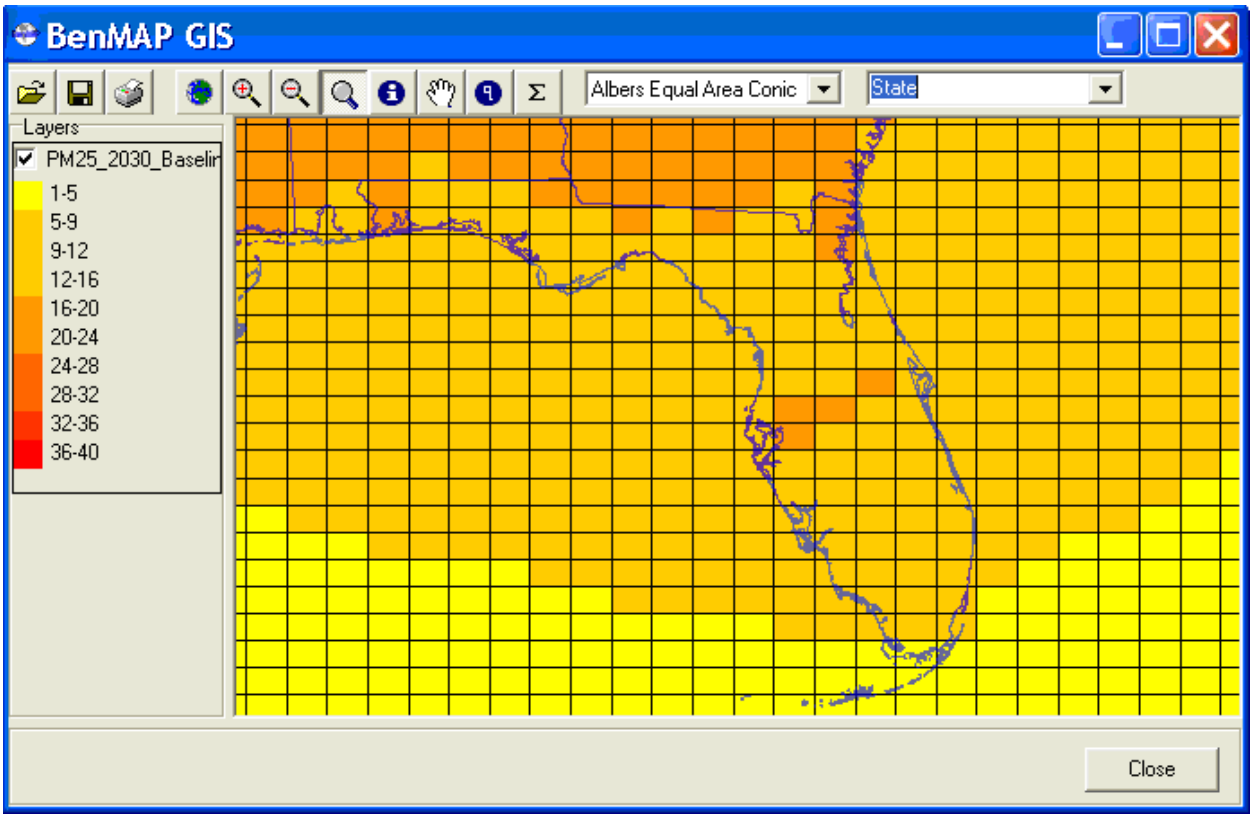
We carry out benefit analyses, and distributional analyses, using the environmental Benefits Mapping and Analysis Program (BenMAP), a tool developed by Abt Associates for EPA for use in estimating the health impacts and economic benefits associated with changes in ambient air pollution. The changes in air pollution are typically calculated with the help of air quality models, which use air pollution emissions data and meteorological data in a complicated series of calculations representing the formation and movement of air pollution in the atmosphere. An air quality model is extremely useful because it can provide estimates of air pollution levels in broad areas of the country, particularly rural areas, where we do not have actual air pollution monitoring data, and because it can provide estimates of air pollution levels for hypothetical scenarios, particularly forecasts for what might happen to air pollution levels in the future.

[6] For ease of explanation, this example gives all individuals in the subgroup the same risk reduction. In general, the change in incidence rate of an adverse health effect within a subgroup (e.g., number of cases avoided per 10,000 population) can be calculated by summing the individual-specific changes in health risk over all individuals in the subgroup and then multiplying by the number in terms of which the rate is defined – e.g., 10,000 if the rate is per 10,000 population.

[7] For basic information about benefit analysis, Appendix A.1 provides an overview of a typical benefit analysis.

The air quality models calculate air pollution levels separately for each cell in a broad air quality grid – like estimating air pollution levels in each cell of a checkerboard. Figure 1 provides an example of what this air quality grid might look like for an analysis of Florida.

Figure 1. Illustration of an Air Quality Grid over the State of Florida



In this particular case taken from the Nonroad Diesel RIA (U.S. EPA, 2004), the grid cells are roughly 36 kilometers by 36 kilometers. More recent national analyses, such as the Locomotive Marine Rule (U.S. EPA, 2008), are using grid cells that are roughly 12 kilometers by 12 kilometers. In general, national analyses are becoming more refined over time, and our ability to process data is improving. With local analyses, such as for individual metropolitan areas, it is currently possible to have even more refined analyses down to grid cells that are 1 kilometer by 1 kilometer, or smaller.

As we noted earlier, ideally we would want to estimate air pollution for each individual in our analysis. However, this is clearly impossible when analyzing air pollution regulations affecting millions of individuals. The size of the grid cells in the air quality model defines how detailed we can be in estimating exposure for individuals. As a result, we must make the simplifying assumption that everyone in a grid cell has the same level of pollutant concentration. Depending on a variety of factors, including the size of the grid cell and the type of air pollutant, there may

be more or less unobserved variation within a given grid cell, with greater variation having a greater potential effect on the results of our analysis. We discuss this further in Section 4.

Air quality model grid cells typically cross Census and jurisdictional boundaries, so population data – a critical component for a distributional analysis – are not readily available for each grid cell. To calculate the population in each grid cell, we aggregate block data, which is the most geographically detailed data available from the Census Bureau. Each block generally has a few hundred individuals.[8] If the center of a block falls within a grid cell, then we assign the block's population to that grid cell. Figure 9 in Appendix B graphically shows this relationship.

Given air pollution estimates and population data for each grid cell, we can use BenMAP to compare the pollutant exposures of different population subgroups. In addition, BenMAP contains health impact functions and baseline incidence rates of health effects that have been associated with criteria air pollutants. As described more fully below, BenMAP uses these health impact functions together with the appropriate baseline incidence rates to estimate grid cell-specific changes in health effect incidence associated with grid cell-specific changes in ambient pollutant concentrations. The methodological details of benefit analysis carried out in BenMAP, and the specific inputs used for the Heavy Duty Diesel (HDD) distributional benefit analysis, are given in Appendix A.

2.2.1 Deriving subgroup-specific distributions of baseline pollutant concentrations

We used BenMAP to determine the numbers of individuals in each subgroup exposed to each possible concentration of a pollutant in the baseline. In particular, we calculated populations of each racial and ethnic group living within each grid cell in an input air quality grid. We describe this in detail in Appendix A.

Recall that the ambient pollutant concentration for a grid cell is assigned to all individuals living within the grid cell. Because BenMAP calculates the number of individuals in each racial and ethnic group within each grid cell[9], it can assign the grid cell's pollutant concentration to race- and ethnicity-identified individuals. For example, if BenMAP has calculated that there are 3,250 African-Americans and 1,750 whites in a grid cell whose annual average $PM_{2.5}$ concentration is 17 $\mu g/m^3$, then those 3,250 African-Americans and 1,750 whites will each be assigned an annual average $PM_{2.5}$ concentration of 17 $\mu g/m^3$.

We used this approach to estimate the annual average $PM_{2.5}$ concentrations to which all individuals within a subgroup are exposed, for all subgroups, and to derive a frequency distribution for each subgroup. The same procedure can be used to estimate ambient pollutant concentrations to which individuals in each subgroup will be exposed in the baseline (absent the regulation) and in the control scenario (in the presence of the regulation), as well as the

[8] Blocks and blockgroups are defined at: http://www.census.gov/geo/www/geo_defn.html. Blockgroups generally have 600 to 3,000 individuals. Since blocks comprise blockgroups, we estimate blocks generally have a few hundred individuals.

[9] It is possible that some grid cells may not have representation of certain subgroups.

corresponding reductions in pollutant concentrations individuals in the subgroup experience as a result of the regulation.

Note that, because all individuals within a grid cell are assigned the same baseline (and control scenario) pollutant concentration, the mean baseline concentration for a subgroup is referred to as a population-weighted average, because it is calculated as a population-weighted average of grid-cell-specific pollutant concentrations. If x_i denotes the baseline pollutant concentration in the *i*th grid cell, n_{ji} denotes the number of individuals in the *j*th subgroup in the *i*th grid cell, and n_j denotes the total number of individuals in the *j*th subgroup, then the mean baseline pollutant concentration for the *j*th subgroup is

$$\bar{x}_j = \left(\frac{1}{n_j}\right)\sum_i n_{ji} * x_i \quad . \tag{3}$$

The mean reduction in pollutant concentration for a subgroup is similarly calculated as a population-weighted average of grid cell-specific reductions in pollutant concentrations as a result of implementation of a rule or regulation.

2.2.2 Estimating subgroup-specific reductions in incidence rates of health effects

As noted in Section 2.2.1 above, BenMAP enables us to identify each individual in the United States by racial and/or ethnic subgroup and to estimate each individual's baseline pollutant level as well as the reduction in pollutant level that will result from implementation of a rule or regulation. This allows us to effectively carry out a benefit analysis separately for each subgroup – i.e., to estimate subgroup-specific changes in health effect incidence, using the same basic approach used in the typical air pollution benefit analysis.

To estimate the health effect incidence change associated with a specified change in level of a pollutant in a grid cell within BenMAP, we need, in addition to the estimated baseline and control scenario levels of the pollutant:

- Concentration-response (C-R) function(s), which provide an estimate of the relationship between the health endpoint of interest and the concentration of the pollutant; and
- Baseline health effects incidence. The baseline incidence of the health effect in a location is the incidence corresponding to baseline pollutant levels in that location. The baseline incidence is typically calculated as the product of the incidence rate (e.g., number of cases per 10,000 population) and the affected population (divided by 10,000, if the rate is per 10,000 population).

These inputs are combined to estimate the health effect incidence reduction associated with a specified reduction in pollutant levels. Although some epidemiological studies have estimated linear or logistic C-R functions, by far the most common form is the exponential (or log-linear) form:

$$y = Be^{\beta x}, \tag{4}$$

where x is the ambient pollutant level, y is the incidence of the health endpoint of interest at pollutant level x, β is the coefficient of ambient pollutant concentration (describing the extent of change in y with a unit change in x), and parameter B is the incidence when there is no ambient pollutant (i.e., x=0). The relationship between a specified ambient pollutant level, x_0, for example, and the incidence of a given health endpoint associated with that level (denoted as y_0) is then

$$y_0 = Be^{\beta x_0}. \tag{5}$$

Because the log-linear form of C-R function (equation (4)) is by far the most common form, we use this form to illustrate the "health impact function" used to calculate changes in health effect incidence. If we let x_0 denote the baseline (upper) level of the pollutant, x_1 denote the control scenario (lower) level of the pollutant, y_0 denote the baseline incidence and y_1 denote the incidence after the rule is implemented, we can derive the following relationship between the change in x, $\Delta x = (x_0 - x_1)$, and the corresponding change in y, Δy, from equation (4)[10]:

$$\Delta y = (y_0 - y_1) = y_0[1 - e^{-\beta \Delta x}]. \tag{6}$$

Ideally we would use subgroup-specific C-R functions and baseline incidence rates to calculate Δy for each subgroup. In practice, however, subgroup-specific C-R functions are rare, and distributional analyses must typically assume that the C-R relationship is the same across subgroups.

For some health endpoints, subgroup-specific baseline incidence rates are available, and if they are available they should be used. Indeed, because there appears to be substantial variation in baseline incidence rates between subgroups, we recommend calculating health impacts only when subgroup-specific baseline incidence rates are available. A detailed discussion of the subgroup-specific baseline incidence rates available for a distributional analysis of the HDD rule, their sources, and the methods we used to estimate rates for specific subgroups is given in Appendix A.

Changes in incidence of each health effect for each subgroup were calculated for each grid cell in BenMAP, using equation (6) with subgroup-specific baseline incidence rates where possible. As noted above, the change in pollutant level, Δx, is assumed to be the same for all individuals in the grid cell. The baseline incidence of the health effect for a subgroup was calculated by multiplying the baseline incidence rate (subgroup-specific, if available) by the population for the subgroup in the grid cell. The total change in incidence of a health effect for a subgroup was

[10] If $\Delta x < 0$ – i.e., if $\Delta x = (x_1 - x_0)$ – then the relationship between Δx and Δy can be shown to be $\Delta y = (y_1 - y_0) = y_0[e^{\beta \Delta x} - 1]$. If $\Delta x < 0$, Δy will similarly be negative. However, the *magnitude* of Δy will be the same whether $\Delta x > 0$ or $\Delta x < 0$ – i.e., the absolute value of Δy does not depend on which equation is used.

then calculated by summing the grid cell-specific changes for that subgroup. To compare subgroups, we calculated the corresponding changes in incidence rates – e.g., the changes in incidence per million population.

2.2.3 Characterizing inequality using inequality measures and Lorenz curves

If we observe a variety of exposures occurring with a non-zero frequency, then the distribution is clearly unequal – some individuals are exposed to higher levels of pollution than others. However, inequality is a matter of degree and it is desirable to have an index that would summarize the distribution into one number by quantifying the degree of inequality. There are a number of such inequality measures with a variety of desirable properties.[11]

One property that is particularly desirable in an inequality measure to be used in a distributional analysis comparing subgroups is the property of decomposability. An inequality measure is subgroup-decomposable (or additively separable) if the total inequality can be divided into constituent parts of the distribution. In other words, one can use the measure to assess within- and between- group inequalities (This is analogous to the comparison of within-group variability to between-group variability in an analysis of variance to determine if the group-defining variable significantly affects the dependent variable.). Levy(2006) proposes the Atkinson index as the most appropriate indicator for health risk analysis in part because it has this property.

The Atkinson index is derived from a Social Welfare Function that is of a Constant Elasticity of Substitution (CES) form (see definition in Appendix A.3). It depends on an inequality aversion parameter, $\varepsilon>0$. When $\varepsilon<1$, more weight is placed on the differences (e.g., differences in baseline $PM_{2.5}$ concentrations between individuals) in the top of the distribution. When $\varepsilon>1$, the index is more sensitive to differences in the bottom of the distribution. This index has a maximum of 1, which indicates extreme inequality, and a minimum of 0, which indicates absolute equality.

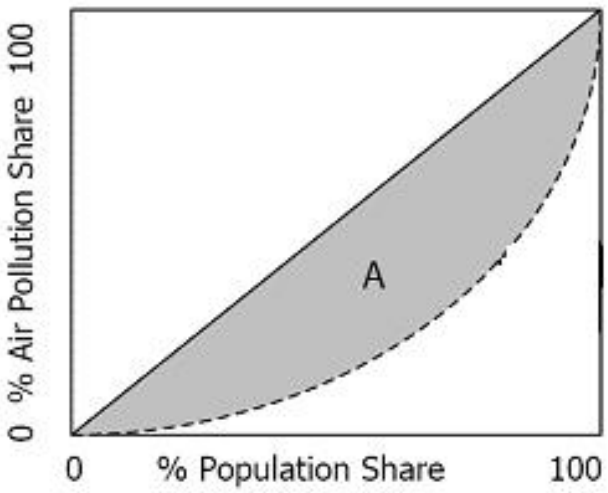
A second inequality measure that has the property of decomposability is the Generalized Entropy indicator, derived from Information Theory (see definition in Appendix A.3). The Generalized Entropy measure depends on a parameter, θ. For $\theta > 0$, the measure is more sensitive to differences in the top of the distribution, while for $\theta < 0$ it is more sensitive to differences in the bottom of the distribution. Unlike the Atkinson measure, the Generalized Entropy indicator does not have a maximum of 1.

Finally, the Lorenz curve is a useful graphical representation of the gap between the distribution in question and a perfectly egalitarian distribution. The Lorenz curve maps cumulative population share to cumulative exposure share – i.e. a point on the Lorenz curve tells us that *x*% of population is exposed to *y*% of total pollution.

[11] Appendix A.3 presents the definitions and properties of various inequality measures. Also see Cowell (2000) for a summary of inequality measures and their properties.

Figure 2 shows the Lorenz curve under different situations. When the distribution is perfectly egalitarian, the Lorenz curve is the 45° line. When the distribution is perfectly unequal (a single individual is exposed to the entire amount of pollution), the Lorenz curve follows the horizontal and then the vertical axis. Other than these two extreme cases, the Lorenz curve would be something like the dotted curve. Area A represents the gap between a completely egalitarian distribution and the distribution being analyzed. The greater the area of A, the greater is the inequality.

Figure 2. The Lorenz Curve



2.3 Assessment of Uncertainty

There are a number of sources of uncertainty, including exposure estimates, population data, incidence rates, and health impact functions. We did not characterize uncertainty in the exposure estimates and simply recognize that there is uncertainty. For some inputs to our analysis, such as air quality modeling, population data, and incidence rates, it is not currently possible to quantify the associated uncertainty. In the case of C-R relationships, we can quantify the uncertainty; however, since our C-R relationships do not vary by population subgroup, it is not immediately obvious that quantifying the uncertainty is informative for this analysis.

3 A Case Study: EPA's Heavy Duty Diesel Rule in 2030

To illustrate the methods described in Section 2, we applied them to EPA's HDD Rule, which is a part of EPA's comprehensive national control program to regulate the heavy-duty vehicle and its fuel as a single system program.

There are two basic parts to the final rule: (1) new exhaust emission standards for heavy duty highway engines and vehicles, and (2) new quality standards for highway diesel fuel (U.S. EPA, 2000). The new emission standards which are applied to heavy-duty highway engines and vehicles took effect in 2007. These emission standards are based on the use of high-efficiency catalytic exhaust emission control devices which are damaged by sulfur. Therefore the emission standards would not be feasible without the new quality standards for diesel fuel, which require a 97% reduction in the sulfur content of diesel fuel. This nationwide program will result in emission levels of particulate matter (PM) and oxides of nitrogen that are 90% and 95%, respectively, below current standards levels (Abt Associates Inc., 2000).

First we define our subgroups. Consistent with the racial categories in the census data, we considered four racial groups – Asian-American, African-American, Native American, and Caucasian (White) – as well as two ethnic groups (Hispanic and non-Hispanic). Table 1 gives the U.S. population by race and ethnicity from Census 2000. Asian Hispanic[12] has the smallest population and White non-Hispanic has the largest.

Table 1. U.S. Population (in *millions*) by Race and Ethnicity.

Race	Ethnicity	Population
Asian-American	Hispanic	0.2
	Non-Hispanic	10.4
African-American	Hispanic	1.4
	Non-Hispanic	34.3
Native American	Hispanic	0.6
	Non-Hispanic	2.1
White	Hispanic	34.8
	Non-Hispanic	195.6

Source: U.S. Census Bureau (2002, Table 1)

For each of the racial subgroups, as well as for some combinations of racial, ethnic, and age groups, we examined:

[12] Asian Hispanic is a term for Hispanic Americans having Asian blood and for those Hispanics who consider themselves or were officially classified by the United States Census Bureau, Office of Management and Budget, and other U.S. government agencies as Asian-Americans. (http://en.wikipedia.org/wiki/Asian_Hispanic)

- the ambient $PM_{2.5}$ concentrations to which they will be exposed in 2030 in the baseline (in the absence of the rule)[13];
- the reduction in ambient $PM_{2.5}$ concentrations they will experience in 2030 as a result of the rule; and the corresponding reduction in each of several health effects they are expected to experience as a result of the rule.

The specific inputs to the analysis (air quality data, concentration-response functions, population data, and baseline incidence rates) are described in detail in Appendix A.2. Section 3.1 presents baseline population-weighted average $PM_{2.5}$ concentration and population distribution, projected to 2030, separately for each subgroup. Section 3.2 presents the projected population-weighted average reduction in ambient $PM_{2.5}$ concentrations as well as the corresponding reduction in health effects predicted to result from the HDD rule in 2030 for each subgroup.

3.1 Projected Population Distributions and Baseline $PM_{2.5}$ Concentrations

Figure 3 shows subgroup-specific maps of population distribution projected to 2030[14]. Figure 4 show a map of projected baseline ambient $PM_{2.5}$ concentrations. A comparison of the subgroup-specific projected population maps with the map of projected baseline ambient $PM_{2.5}$ concentrations gives a "broad brush" picture of which groups are projected to live in areas of high and low $PM_{2.5}$ concentrations.

$PM_{2.5}$ can be considered as a regional pollutant compared with some other pollutants (e.g. CO). Figure 4 shows that the high $PM_{2.5}$ concentrations are mostly in the Eastern half of the United States and, to a lesser extent, in southern California – areas in which most African-Americans and Asian-Americans are projected to live. In contrast, Native Americans are projected to be concentrated largely in the Southwest and to some extent in California. Because the Southwest is projected to have relatively low $PM_{2.5}$ concentrations, Native Americans have the lowest baseline $PM_{2.5}$ concentrations of any of the racial subgroups.

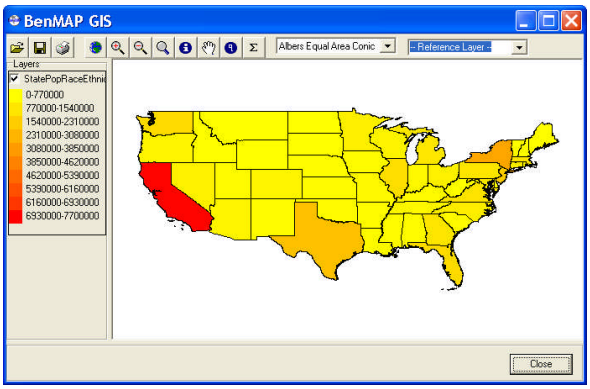
Figure 5 presents the forecasted reductions in ambient $PM_{2.5}$ concentrations across the U.S. For regions that have high baseline $PM_{2.5}$ concentrations, the forecasted reductions are also large, as expected.

[13] The proposed method can apply to other pollutants as well. The original analysis examined O_3 and $PM_{2.5}$. In this final report, we focus on $PM_{2.5}$ for simplicity.

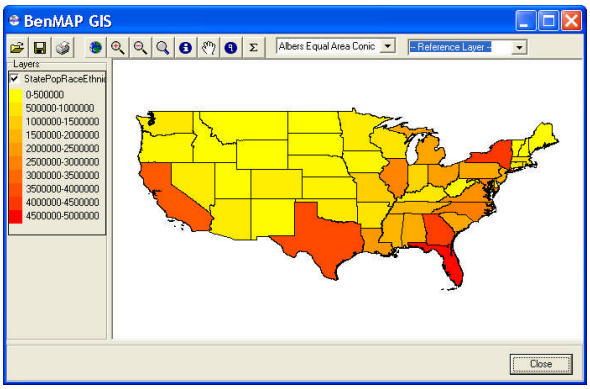
[14] For details about how BenMAP forecasts the population to 2030, see Appendix A.2.2.

Figure 3. Forecasted 2030 Populations by Race and Ethnicity

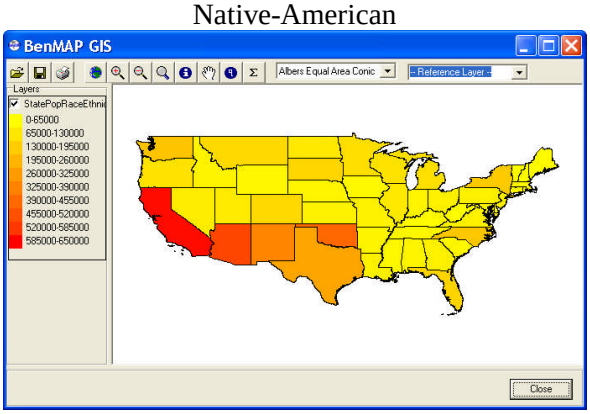
Asian-American



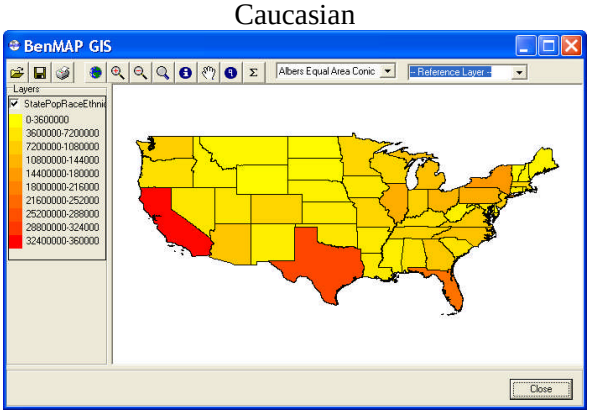
African-American



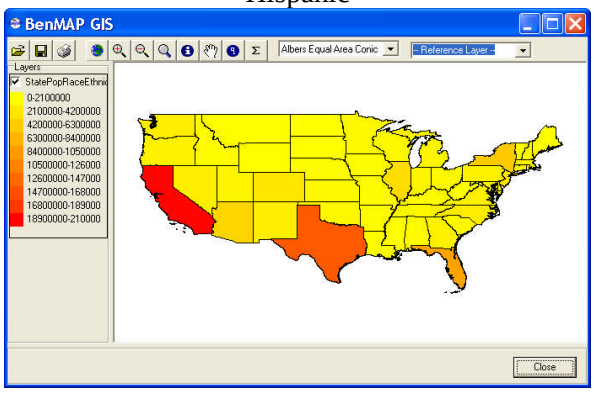
Native-American



Caucasian



Hispanic



Non-Hispanic

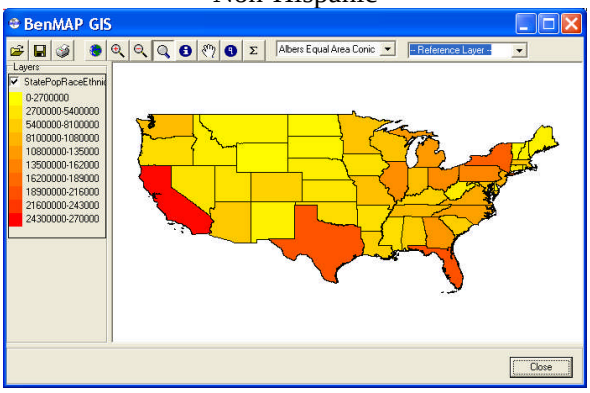


Figure 4. Forecasted 2030 *Baseline* Ambient $PM_{2.5}$ Concentrations

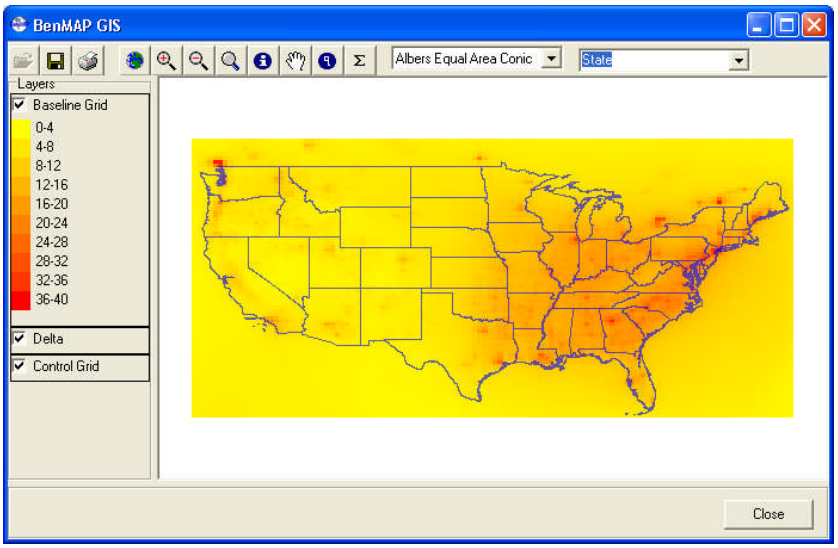
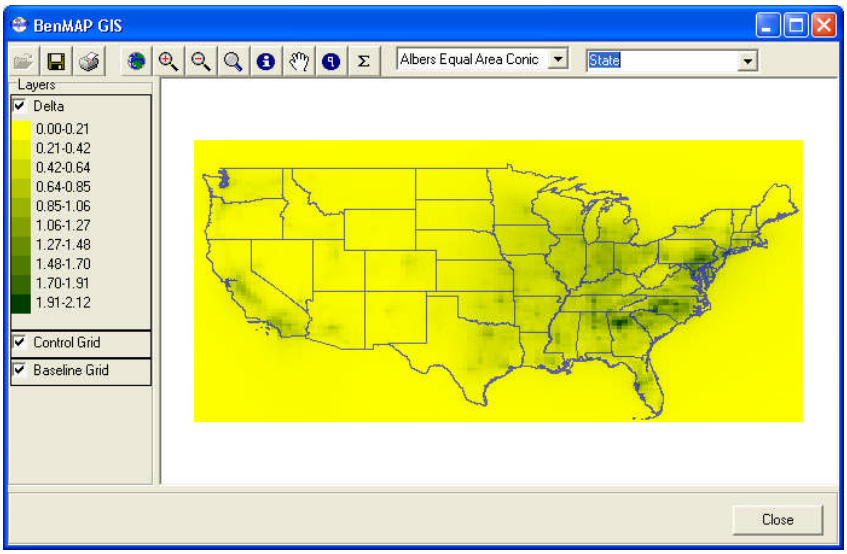


Figure 5. Forecasted 2030 *Reduction* in Ambient $PM_{2.5}$ Concentrations (Baseline Minus Control)



The baseline $PM_{2.5}$ concentrations that subgroups are predicted to experience in 2030 can also be characterized quantitatively by calculating population-weighted averages of grid cell-specific concentrations, as shown in equation (3) above. Racial and ethnic group-specific population-weighted average baseline $PM_{2.5}$ concentrations in 2030 are shown in Table 2.

Table 2. Predicted Racial and Ethnic Group-Specific Population-Weighted Average *Baseline* $PM_{2.5}$ Concentrations in 2030

Racial/Ethnic Subgroup	Mean Baseline $PM_{2.5}$ Level in 2030 (ug/m^3)
Asian-American	16.7
African-American	18.1
Native American	10.2
White Hispanic	13.4
White non-Hispanic	14.1
Total Population	14.7

As can be seen from Table 2, on average, African-Americans and Asian-Americans are predicted to experience the highest baseline $PM_{2.5}$ concentrations (i.e., in the absence of the HDD rule), and Native Americans are predicted to experience the lowest. As noted above, this is largely a result of the relative proportions of subgroups that are projected to live in those areas of the United States where particulate matter air pollution is highest —in particular, Asian-Americans and African-Americans are projected to be disproportionately located in those portions of the country with the highest projected baseline $PM_{2.5}$ concentrations.

As noted in Section 2.3, our results are associated with various uncertainties. However there is no uncertainty due to sampling error surrounding the estimates in Table 2, because these means are not based on samples but on a complete census of the population. Thus the usual tests to determine whether estimated means are statistically significantly different from each other do not apply.

3.2 Population-Weighted Average Reductions in $PM_{2.5}$ and Corresponding Reductions in Health Effects Incidence

We present the reduction in health effects incidence rates alongside the reduction in $PM_{2.5}$ concentrations each subgroup is predicted to experience as a result of the HDD rule, separately for each of several health effects for which there is epidemiological evidence of an association with $PM_{2.5}$. This juxtaposition makes it easier to see the correspondence or lack of correspondence between the two. For each subgroup, we show (1) the *absolute* reduction in $PM_{2.5}$ concentration the subgroup is predicted to experience and the *absolute* reduction in the health effect (cases per million population), as well as (2) the *relative* reduction in $PM_{2.5}$ concentration and the *relative* reduction in the health effect per million population – relative to the total population. The relative reduction allows us to see at a glance how one subgroup is expected to fare relative to others, both in terms of the reduction in $PM_{2.5}$ concentration they will experience and in terms of the reduction in health effects expected to result.

We calculate the relative reduction by dividing the absolute reduction for each racial-age group by the absolute reduction for the total population for that age group. For example, if Asian-American adults (ages 18-64) have an absolute reduction in mean $PM_{2.5}$ concentration of 0.69 ug/m^3 as a result of HDD rule and the absolute reduction for all adults (ages 18-64) is 0.59

ug/m^3, then the relative reduction is approximately 1.2 (≈ 0.69 / 0.59). This means that Asian-American adults experience 20 percent more $PM_{2.5}$ reduction compared with the total population, so they benefit more from the HDD rule than the population as a whole. In contrast, while the *absolute* reduction in mean $PM_{2.5}$ concentration is positive (0.39 ug/m^3) for Native-American adults, the *relative* reduction is 0.7 (<1). This indicates that, although Native-American adults do benefit from the HDD rule, they benefit less than the population as a whole.

Tables 2, 3, 4, and 5 present results for non-fatal acute myocardial infarction, emergency room (ER) visits for asthma, cause-specific hospital admissions, and all-cause mortality, respectively.[15]

Each table includes:

- Subgroup information: While the subgroups are defined by age, race and ethnicity for asthma ER visits, they are defined only by age and race for other health endpoints. This is because incidence data by ethnicity are not available for the other health endpoints.
- Absolute and relative reduction in $PM_{2.5}$ concentrations as a result of the HDD rule, as explained at the beginning of Section 3.2.
- Baseline Incidence: The subgroup-specific baseline incidence rates for the health effect(s) contribute to the calculation of reduction in incidence associated with reduction in $PM_{2.5}$ concentrations, as explained in Section 3.2.2 below.
- Absolute and relative reduction in incidence for each health effect as a result of the HDD rule (see discussion in Section 3.2.2 below).

Table 3. Absolute and Relative Reduction in Mean $PM_{2.5}$ Concentrations and Incidence of *Non-Fatal Acute Myocardial Infarction* (per Million Population)

Age / Race	Baseline Incidence per 100,000 Pop.	Absolute Reduction in $PM_{2.5}$ Level (ug/m^3)	Relative Reduction in $PM_{2.5}$ Level	Absolute Reduction in Incidence per Million Pop.	Relative Reduction in Incidence per Million Pop.
Adults (18-64)					
Asian-American	107.3	0.69	1.2	12.9	0.5
African-American	166.6	0.74	1.2	28.7	1.1
Native American	91.8	0.39	0.7	7.2	0.3
White	202.5	0.56	0.9	27.5	1.0
Total Population	*188.9*	*0.59*	--	*26.4*	--
Elderly (65+)					
Asian-American	858.3	0.67	1.2	114.6	0.7
African-American	1,134.1	0.74	1.3	198.5	1.1
Native American	2,176.7	0.39	0.7	135.9	0.8
White	1,358.3	0.54	1.0	177.9	1.0
Total Population	*1309.4*	*0.57*	--	*175.4*	--

[15] The health endpoints we presented here do not represent the complete list of endpoints attributable to the HDD rule. They are chosen to fully represent the range of ages, races, and ethnicities of interest in the analysis.

Table 4. Absolute and Relative Reduction in Mean $PM_{2.5}$ Concentrations and Incidence of *Asthma-Related ER Visits* Among Children (0-17) (per Million Population)

Race	Ethnicity	Baseline Incidence / 100,000 Pop.	Absolute Reduction in $PM_{2.5}$ Level (ug/m^3)	Relative Reduction in $PM_{2.5}$ Level	Absolute Reduction in Incidence / Million Pop.	Relative Reduction in Incidence / Million Pop.
Asian-American	Hispanic	244.4	0.58	1.0	15.3	0.2
	Non-Hispanic	170.7	0.71	1.2	12.0	0.1
African-American	Hispanic	1,548.3	0.65	1.1	180.9	1.9
	Non-Hispanic	2,099.2	0.74	1.2	253.7	2.7
Native American	Hispanic	265.4	0.51	0.8	15.0	0.2
	Non-Hispanic	258.6	0.30	0.5	6.3	0.1
White	Hispanic	614.1	0.54	0.9	59.9	0.6
	Non-Hispanic	768.1	0.58	1.0	76.1	0.8
Total Population	--	*889.9*	*0.60*	--	*94.5*	--

Table 5. Absolute and Relative Reduction in Mean $PM_{2.5}$ Concentrations and *Hospitalizations* (per Million Population)

Effect / Race	Baseline Incidence per 100,000 Pop.	Absolute Reduction in $PM_{2.5}$ Level (ug/m^3)	Relative Reduction in $PM_{2.5}$ Level (ug/m^3)	Absolute Reduction in Incidence per Million Pop.	Relative Reduction in Incidence per Million Pop.
All Cardiovascular (less Myocardial Infarctions): Adults (18-64)					
Asian-American	608.1	0.69	1.2	5.2	0.7
African-American	1,487.0	0.74	1.2	15.2	1.9
Native American	980.6	0.39	0.7	5.5	0.7
White	841.9	0.56	0.9	6.8	0.9
Total Population	*918.6*	*0.59*	--	*7.9*	--
All Cardiovascular (less Myocardial Infarctions): Elderly (65+)					
Asian-American	5,459.7	0.67	1.2	49.2	0.8
African-American	8,393.8	0.74	1.3	97.1	1.5
Native American	12,718.9	0.39	0.7	59.5	0.9
White	6,944.8	0.54	1.0	60.9	1.0
Total Population	*7037.2*	*0.57*	--	*63.7*	--
Congestive Heart Failure: Elderly (65+)					
Asian-American	1,651.2	0.67	1.2	27.6	0.7
African-American	2,746.2	0.74	1.3	62.6	1.6
Native American	2,735.0	0.39	0.7	26.4	0.7
White	2,105.1	0.54	1.0	36.3	0.9
Total Population	*2143.9*	*0.57*	--	*38.2*	--
Dysrhythmia: Elderly (65+)					
Asian-American	867.3	0.67	1.2	5.6	0.6
African-American	1,044.4	0.74	1.3	9.9	1.0
Native American	2,409.9	0.39	0.7	7.4	0.8
White	1,413.3	0.54	1.0	9.9	1.0
Total Population	*1348.4*	*0.57*	--	*9.6*	--
Ischemic Heart Disease (less Myocardial Infarctions): Elderly (65+)					
Asian-American	1,967.7	0.67	1.2	16.9	0.9
African-American	1,980.5	0.74	1.3	20.5	1.1
Native American	5,553.9	0.39	0.7	26.4	1.4
White	2,269.9	0.54	1.0	18.1	1.0
Total Population	*2247.5*	*0.57*	--	*18.3*	--

Effect / Race	Baseline Incidence per 100,000 Pop.	Absolute Reduction in $PM_{2.5}$ Level (ug/m^3)	Relative Reduction in $PM_{2.5}$ Level (ug/m^3)	Absolute Reduction in Incidence per Million Pop.	Relative Reduction in Incidence per Million Pop.
Asthma: Children (0-17)					
Asian-American	189.8	0.70	1.2	4.1	0.7
African-American	719.3	0.73	1.2	18.6	3.1
Native American	294.2	0.37	0.6	4.2	0.7
White	196.0	0.57	0.9	3.6	0.6
Total Population	*280.0*	*0.60*	--	*6.1*	--
Asthma: Adults (18-64)					
Asian-American	52.6	0.69	1.2	1.0	0.4
African-American	297.0	0.74	1.2	7.5	2.9
Native American	49.8	0.39	0.7	0.7	0.3
White	94.9	0.56	0.9	1.9	0.7
Total Population	*120.3*	*0.59*	--	*2.6*	--
Chronic Lung Disease (Study: Moolgavkar): Elderly (65+)					
Asian-American	992.2	0.67	1.2	11.7	0.7
African-American	1,956.7	0.74	1.3	26.3	1.6
Native American	1,148.9	0.39	0.7	8.4	0.5
White	1,540.5	0.54	1.0	15.8	1.0
Total Population	*1542.4*	*0.57*	--	*16.5*	--
Chronic Lung Disease (Study: Ito): Elderly (65+)					
Asian-American	992.2	0.67	1.2	7.4	0.7
African-American	1,956.7	0.74	1.3	16.7	1.6
Native American	1,148.9	0.39	0.7	5.3	0.5
White	1,540.5	0.54	1.0	10.1	1.0
Total Population	*1542.4*	*0.57*	--	*10.5*	--
Chronic Lung Disease (less Asthma): Adults (18-64)					
Asian-American	30.7	0.69	1.2	0.3	0.2
African-American	162.5	0.74	1.2	2.6	1.4
Native American	31.1	0.39	0.7	0.2	0.1
White	156.2	0.56	0.9	2.0	1.0
Total Population	*146.2*	*0.59*	--	*1.9*	--
Pneumonia: Elderly (65+)					
Asian-American	1,857.9	0.67	1.2	43.2	0.8
African-American	2,026.5	0.74	1.3	59.6	1.1
Native American	4,974.3	0.39	0.7	58.1	1.1
White	2,375.1	0.54	1.0	51.7	1.0
Total Population	*2327.1*	*0.57*	--	*51.9*	--

Table 6. Absolute and Relative Reduction in Mean $PM_{2.5}$ Concentrations and Incidence of *All-Cause Mortality* (per Million Population)

Age / Race	Baseline Incidence per 100,000 Pop.	Absolute Reduction in $PM_{2.5}$ Level (ug/m^3)	Relative Reduction in $PM_{2.5}$ Level (ug/m^3)	Absolute Reduction in Incidence per Million Pop.	Relative Reduction in Incidence per Million Pop.
Infants (Age 0)					
Asian-American	290.7	0.71	1.2	13.6	0.7
African-American	954.3	0.74	1.2	46.9	2.3
Native American	416.6	0.38	0.6	9.0	0.4
White	400.5	0.57	0.9	15.3	0.8
Total Population	*481.6*	*0.61*	--	*20.2*	--
Adults (30-64)					
Asian-American	177.1	0.69	1.2	6.6	0.6
African-American	518.3	0.73	1.2	21.6	2.0
Native American	258.7	0.40	0.7	4.6	0.4
White	302.7	0.55	0.9	9.7	0.9
Total Population	*322.5*	*0.58*	--	*11.1*	--
Elderly (65+)					
Asian-American	2,041.1	0.67	1.2	77.6	0.6
African-American	3,978.3	0.74	1.3	170.1	1.4
Native American	2,534.4	0.39	0.7	52.8	0.4
White	3,794.5	0.54	1.0	119.6	1.0
Total Population	*3686.3*	*0.57*	--	*121.3*	--

3.2.1 Subgroup-specific reductions in $PM_{2.5}$ concentrations

On average, Asian-Americans and African-Americans are predicted to experience relatively larger reductions in $PM_{2.5}$ concentrations as a result of the HDD rule. Asian-Americans are predicted to experience about 20 percent greater reductions, on average, than the total population (i.e., a *relative* reduction of 1.2), while African-Americans are predicted to experience from 20 percent to 30 percent greater reductions (i.e., relative reductions of 1.2 or 1.3), on average, depending on the age group considered. Native Americans, on the other hand, are predicted to experience reductions that are relatively smaller than the general population, on average – about 70 percent of the reduction for the total population. Finally, whites are predicted to experience reductions in air quality that are basically the same as those of the total population (relative reduction of 0.9 for ages 18 – 64, and 1.0 for the young and the elderly).

All of these relative reductions largely reflect the confluence of population distributions and baseline air quality. Native Americans are predicted to see relatively small reductions in the $PM_{2.5}$ concentrations to which they will be exposed, on average, because they live largely in areas in which the baseline $PM_{2.5}$ concentrations were low to begin with. In contrast, Asian-Americans and African-Americans are predicted to experience relatively larger reductions in $PM_{2.5}$ concentrations as a result of the HDD rule, on average, because they tend to live in areas with higher baseline concentrations (e.g., a relatively high proportion of African-Americans live in the Eastern United States, with its higher concentrations of $PM_{2.5}$).

3.2.2 Subgroup-specific health effects

The relative reductions in air quality predicted to be experienced by the different racial and ethnic subgroups as a consequence of where they live, however, do not necessarily translate into the same relative reductions in health effects. This is because the reductions in health effects depend, in addition, on the baseline incidence rates of the health effects, and these differ substantially across the subgroups. For example, the baseline rate of hospital admissions for asthma among children (ages 0 – 17) is 190 per 100,000 Asian-American children, whereas it's 719 per 100,000 African-American children. Both groups are predicted to experience about 20 percent greater reductions in $PM_{2.5}$ concentrations, on average, relative to the total population. However, Asian-American children are predicted to experience a reduction in asthma-related hospital admissions that is only 70 percent of the reduction that will be experienced by children in the total population, whereas African-American children are predicted to experience a reduction that is over 300 percent of the reduction for children in the general population. This reflects the underlying greater vulnerability of African-American children to hospitalization for asthma, reflected in their much higher baseline incidence rate, relative to the general population (or, for that matter, to any other subgroup).

The relative reductions in PM-related health effects incidence for the different racial and ethnic subgroups predicted to result from the HDD rule in 2030 thus reflect both the relative reductions in $PM_{2.5}$ concentrations experienced by the different subgroups and their underlying baseline incidence rates. As shown in the example above, the latter can differ substantially among the subgroups, often reflecting underlying socioeconomic and/or genetic differences. Thus even if two subgroups are predicted to experience the *same* reduction in $PM_{2.5}$ concentrations as a result of the HDD rule, the reduction in health effects incidence rates that will result can be very different in the two groups, reflecting differences in their baseline incidence rates which, in turn, reflect differences in their underlying susceptibilities to these environmental insults.

Even within the same broad category of health effect – hospital admissions – there can be substantial differences in incidence reduction across specific types of hospital admissions, as can be seen in Table 5. For example, African-Americans are predicted to experience reductions in $PM_{2.5}$ concentrations that are 20 or 30 percent greater than those that will be experienced by the total population. For some types of hospital admissions (e.g., for ischemic heart disease or dysrhythmia among the elderly) this is predicted to result in decreases in incidence among African-Americans that are not very different from those for the total population; for other types of hospital admissions (e.g., for asthma among children or adults), the predicted reductions in incidence among African-Americans are much greater (about 3 times as much as for the total population).

3.3 Other Characterizations of Differences Among Subgroups

In sections 3.1 and 3.2, we compared mean baseline $PM_{2.5}$ concentrations, and mean reductions in $PM_{2.5}$ concentrations and health effect incidence, across racial and ethnic subgroups, where these subgroup-specific population means were calculated as the population-weighted average of grid cell-specific values, as shown in equation (3) above. The population-weighted average is a

good metric for highlighting broad patterns – e.g., that one racial group tends to experience higher baseline levels of a pollutant and/or greater reduction in pollutant concentrations than others, or that the corresponding reductions in health effects incidence do not necessarily follow the same patterns as the reductions in pollutant concentration exposures.

3.3.1 Comparing other characteristics of subgroup-specific distributions

The mean is just one characteristic of an entire distribution of baseline $PM_{2.5}$ concentrations a subgroup is predicted to experience – i.e., not everyone in a subgroup experiences the same baseline concentration. Similarly, each subgroup can be characterized by an entire distribution of reductions in $PM_{2.5}$ concentrations as a result of the HDD rule. Another way to compare subgroups, then, is to compare these distributions. This can be done either graphically or in a tabular presentation of key percentiles of the distributions. We illustrate a graphical approach using a comparison of the subgroup-specific cumulative distributions of baseline $PM_{2.5}$ concentrations, shown in Figure 6, and subgroup-specific cumulative distributions of reductions in $PM_{2.5}$ concentrations as a result of the HDD rule, shown in Figure 7.[16] We illustrate the tabular approach to comparing distributions using the same comparisons of the subgroup-specific distributions of baseline $PM_{2.5}$ concentrations and reductions in $PM_{2.5}$ concentrations as a result of the HDD rule, in Tables 6 and 7, respectively.

In Figure 6 and Figure 7, any point (x, y) along the cumulative distribution shows that $100y$ percent of each subgroup is exposed to more than x $PM_{2.5}$ concentration. Figure 6 shows that Native Americans are exposed to relatively low baseline $PM_{2.5}$ levels, whereas greater percentages of African-Americans, Asian-Americans and Whites are exposed to relatively high baseline $PM_{2.5}$ levels.

Figure 7 shows the distribution of reductions in $PM_{2.5}$ levels for each subgroup. The pattern of the curves is similar to that in Figure 6. Native-Americans are predicted to experience smaller reductions in $PM_{2.5}$ levels than other groups. Figure 6 and Figure 7 jointly indicate that subgroups exposed to higher levels of pollution in the baseline will receive larger benefits as a result of the HDD rule.

[16] Note that Figure 6 and Figure 7 both show the proportions of each subgroup experiencing *greater than or equal to* some level – i.e., $1 - F_q$ in equation (2).

Figure 6. Racial and Ethnic Group-Specific Distributions of *Baseline* $PM_{2.5}$ Concentrations

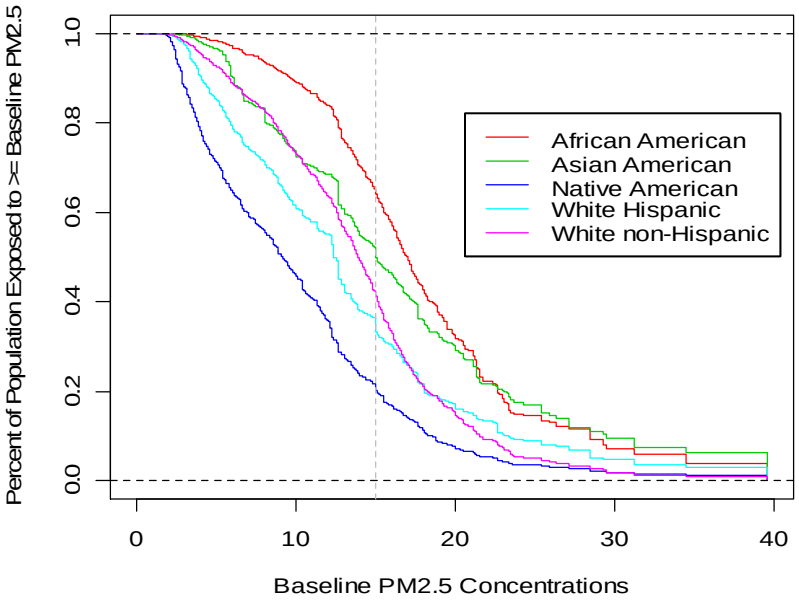


Figure 7. Racial and Ethnic Group-Specific Distributions of *Reduction* in $PM_{2.5}$ Concentrations

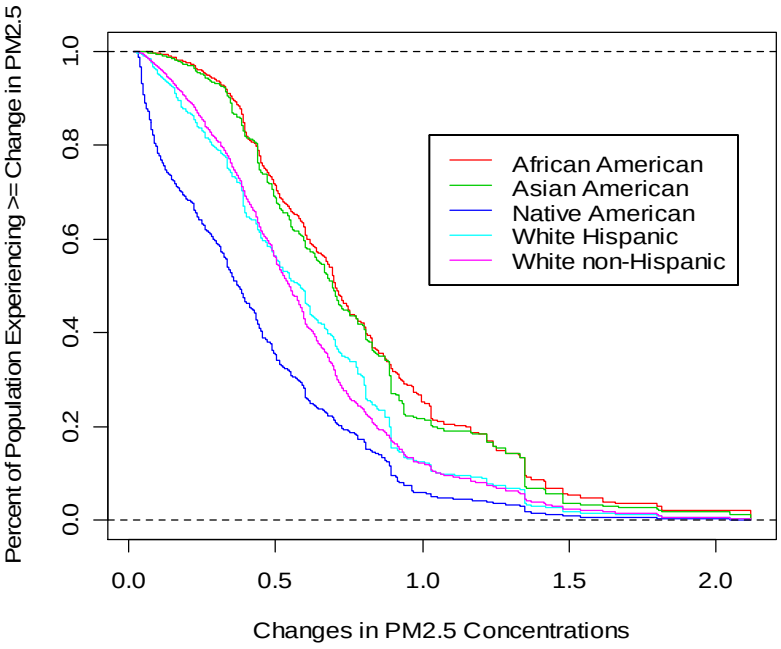


Table 7. 2030 Projected *Baseline* Annual Average $PM_{2.5}$ Concentrations (ug/m^3) by Race and Ethnicity

Racial/Ethnic Subgroup	Mean	Std. Deviation	5th Percentile	25th Percentile	50th Percentile	75th Percentile	95th Percentile
Asian-American	16.705	9.133	5.637	9.525	15.029	21.308	39.587
African-American	18.130	7.5023	7.424	13.219	16.991	21.469	34.467
Native American	10.219	6.972	2.466	4.426	9.171	13.739	22.612
White Hispanic	13.385	8.210	3.382	6.779	12.401	17.278	29.319
White non-Hispanic	14.068	6.447	4.162	9.610	13.892	17.224	25.351
Total Population	14.654	7.388	4.050	9.517	14.045	18.019	28.441

Table 8. 2030 Projected *Reductions* in Annual Average $PM_{2.5}$ Concentrations (ug/m^3) by Race and Ethnicity

Racial/Ethnic Subgroup	Mean	Std. Deviation	5th Percentile	25th Percentile	50th Percentile	75th Percentile	95th Percentile
Asian-American	0.766	0.406	0.252	0.445	0.693	0.932	1.480
African-American	0.792	0.419	0.265	0.465	0.706	1.008	1.539
Native American	0.438	0.354	0.042	0.125	0.373	0.622	1.048
White Hispanic	0.616	0.374	0.105	0.341	0.585	0.828	1.347
White non-Hispanic	0.606	0.365	0.128	0.358	0.549	0.777	1.347
Total Population	0.644	0.385	0.134	0.383	0.585	0.828	1.355

We can see from Tables 6 and 7 that there is substantial variability in both baseline $PM_{2.5}$ concentrations and reductions in $PM_{2.5}$ concentrations as a result of the HDD rule, even within subgroups. Among African-Americans, for example, while the mean baseline annual average $PM_{2.5}$ concentration is 18.13 ug/m^3, five percent of this subgroup is predicted to experience baseline concentrations less than half that concentration (7.42 ug/m^3), while another five percent is predicted to experience baseline concentrations almost 5 times that concentration (34.47 ug/m^3). The general patterns seen in the subgroup-specific means, however, is also seen in the distributions as a whole. While African-Americans and Asian-Americans have the highest mean baseline $PM_{2.5}$ concentrations, they also have the highest 75th and 95th percentile concentrations. So, for example, 25 percent of African-Americans will experience $PM_{2.5}$ concentrations in excess of 21.47 ug/m^3. This is higher than the 25th percentile concentration for any other subgroup. The 25th percentile concentration for Native Americans is, like their mean, the lowest among the subgroups.

In addition, the subgroup-specific reductions in $PM_{2.5}$ concentrations as a result of the HDD rule tend to follow a pattern that is the reverse of the pattern seen in the baseline concentrations -- i.e., the subgroups that experience the worst baseline conditions tend to enjoy the greatest reductions in $PM_{2.5}$ concentrations – as was seen in the population-weighted means. For example, while African-Americans are predicted to experience the highest baseline $PM_{2.5}$ concentrations, they are also predicted to experience the greatest reductions in $PM_{2.5}$ concentrations as a result of the HDD rule – both at the mean and in the upper tail of the distribution. Five percent of African-Americans are predicted to experience reductions in $PM_{2.5}$ concentrations of at least 1.54 ug/m^3,

whereas for the total population that 95th percentile point is only 1.36 ug/m^3, and for Native Americans it's only 1.05 ug/m^3.

3.3.2 Comparing subgroups using Lorenz curves and inequality measures

We introduced the Lorenz curve and inequality measures in Section 2.2.3. Lorenz curves, presented in Figure 8, show graphically how subgroup-specific inequality in baseline $PM_{2.5}$ concentrations varies across the subgroups. The subgroups whose curves are closest to the 45° line have the least inequality; those whose curves are furthest from the 45° line have the most. One can see at a glance that there is less inequality among African-Americans and white non-Hispanics in contrast to more inequality among Asian-Americans, white Hispanics, and Native Americans. Note that the Lorenz curve for the general population lies within the locus of group-specific Lorenz curves, which implies that the between-group variability is not the major source of overall variability.

Recall that inequality measures that are decomposable are particularly useful in distributional analyses comparing subgroups, because they allow us to compare the inequality among members within subgroups to the inequality between subgroups (analogous to the comparison of within-group variability to between-group variability in an analysis of variance to determine if the group-defining variable significantly affects the dependent variable).

As we noted in Section 2.2.3, the Generalized Entropy (GE) indicator depends on a parameter, θ. For $\theta > 0$, the measure is more sensitive to differences in the higher end of the distribution, while for $\theta < 0$ it is more sensitive to differences in the lower end of the distribution. In order to see the results under different choices of θ, we calculated the GE indicator for baseline $PM_{2.5}$ concentrations, for $\theta = -1$, 0, 1, and 2, first for the total population, and then for each racial/ethnic subgroup. The results are given in Table 9.

Figure 8. Subgroup-Specific Lorenz Curves for *Baseline* $PM_{2.5}$ Concentrations in 2030

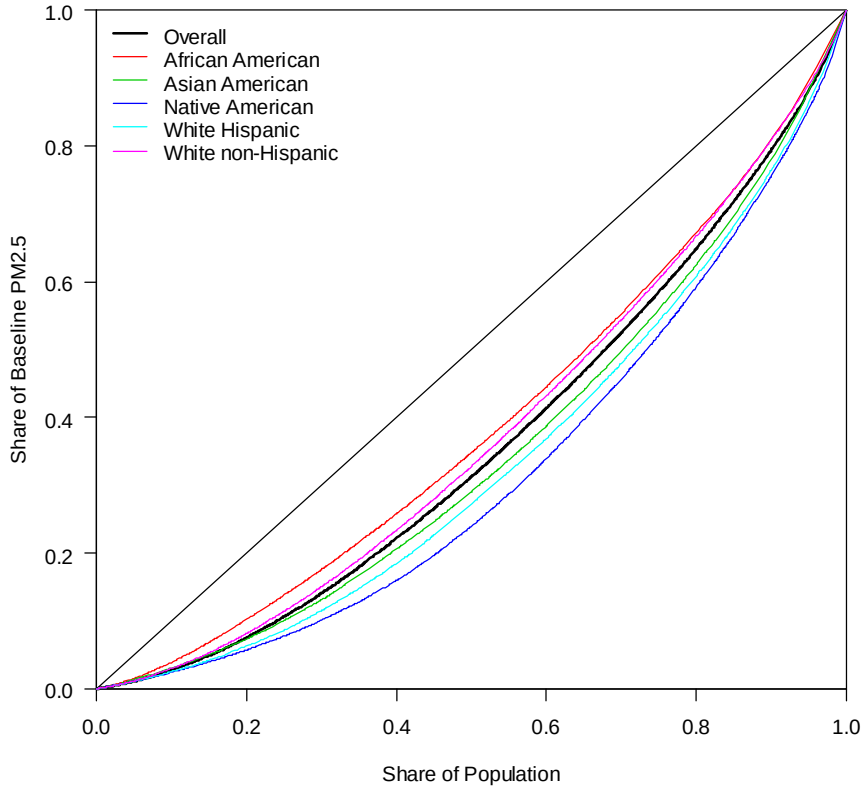


Table 9. Generalized Entropy (GE) Indicator of Inequality in *Baseline* $PM_{2.5}$ Concentrations in the Total Population and in Racial/Ethnic Subgroups

Racial/Ethnic Subgroup	Generalized Entropy Indicator (GE)			
	GE(θ =-1)	GE(θ = 0)	GE(θ = 1)	GE(θ = 2)
Total Population	0.189	0.140	0.124	0.127
Asian-American	0.196	0.155	0.143	0.149
African-American	0.109	0.089	0.083	0.086
Native American	0.309	0.233	0.212	0.233
White Hispanic	0.257	0.194	0.176	0.188
White non-Hispanic	0.163	0.120	0.105	0.105
Within-Group Inequality	0.183	0.134	0.118	0.121
Between-Group Inequality	0.006	0.006	0.006	0.006
Share of Total Inequality Attributable to Between-Group Inequality	3.1%	4.3%	4.9%	5.0%

We similarly calculated the Atkinson index for baseline $PM_{2.5}$ concentrations for the total population and for each of the racial and ethnic subgroups. The Atkinson index depends on a parameter, $\varepsilon>0$. When $\varepsilon<1$, more weight is placed on the differences between individuals in the higher end of the distribution; when $\varepsilon>1$, the index is more sensitive to differences in the lower end of the distribution. We calculated the Atkinson index using values of $\varepsilon = 0.5$, 1, and 2. The results are given in Table 10.

Table 10. Atkinson Index of Inequality in *Baseline* $PM_{2.5}$ Concentrations in Total Population and in Racial/Ethnic Subgroups

Racial/Ethnic Subgroup	Atkinson Index		
	ε = 0.5	ε = 1	ε = 2
Total Population	0.064	0.131	0.274
Asian-American	0.072	0.144	0.281
African-American	0.042	0.085	0.179
Native American	0.106	0.207	0.382
White Hispanic	0.088	0.176	0.340
White non-Hispanic	0.054	0.113	0.246
Within-Group Inequality	0.060	0.123	0.255
Between-Group Inequality	0.004	0.009	0.026
Share of Total Inequality Attributable to Between-Group Inequality	5.67%	6.69%	9.30%

Holding the parameter (θ or ε) fixed, larger indicators indicate more inequality. Table 9 and Table 10 show that the results are consistent no matter which index we use. For all choices of θ or ε, we see that Native Americans have more inequality than any other subgroup and African-Americans have the least inequality. That is, Native Americans have a wider spread of exposures to $PM_{2.5}$ and African-Americans are more equally exposed to $PM_{2.5}$ compared with other subgroups.

These inequality measures are particularly useful for showing at a glance the extent to which the total inequality in the total population is due to inequality among subgroups. In the case of baseline $PM_{2.5}$ concentrations projected to 2030, both inequality measures strongly indicate that only a small share of the total inequality among individuals in the total population is due to inequality among subgroups. Using the Generalized Entropy Indicator with $\theta = 2$, for example, slightly less than 5 percent of the total inequality in baseline $PM_{2.5}$ concentrations to which individuals in the general population are predicted to be exposed in 2030 is due to inequality among the racial/ethnic subgroups. There is far more inequality within subgroups than between them. Although the percentages are slightly different, the Atkinson Index tells the same basic story. Using $\varepsilon = 1$, for example, less than 7 percent of the total inequality in baseline $PM_{2.5}$ concentrations is due to inequality among the racial/ethnic subgroups. The between-group inequality is at least an order of magnitude smaller than the within-group inequality for any of the racial/ethnic subgroups. Thus, the differences between subgroup-specific means that we see in Tables 1 and 2 seem much less substantial when seen in this broader context – i.e., there are differences between the subgroups, on average, but these differences are very small compared to the within-subgroup differences.

The inequality in reduction in $PM_{2.5}$ concentrations in 2030 as a result of the HDD Rule within and between racial and ethnic subgroups, measured using the Generalized Entropy indicator and the Atkinson Index is shown in Tables 11 and 12, respectively. The results show a similar pattern to the baseline case. Within the subgroup of Native Amercians, there is greater variability in reduction of levels of exposure as a result of the HDD rule. In contrast, African-Americans experience more equal reductions compared with other subgroups.

As with baseline $PM_{2.5}$ concentrations, the inequality in reductions in $PM_{2.5}$ concentrations is substantially greater within subgroups than between subgroups. Although the percentages vary with the inequality measure and the parameter value chosen, the basic picture is clear.

Table 11. Generalized Entropy (GE) Indicator of Inequality in *Reduction* in $PM_{2.5}$ Concentrations Due to the HDD Rule in the Total Population and in Racial/Ethnic Subgroups

Racial/Ethnic Subgroup	Generalized Entropy Indicator			
	GE(θ =-1)	GE(θ = 0)	GE(θ = 1)	GE(θ = 2)
Total Population	0.335	0.203	0.172	0.179
African-American	0.192	0.145	0.133	0.140
Asian-American	0.200	0.147	0.134	0.140
Native American	0.846	0.421	0.318	0.328
White Hispanic	0.396	0.226	0.183	0.184
White non-Hispanic	0.325	0.201	0.172	0.181
Within-Group Inequality	0.329	0.197	0.165	0.172
Between-Group Inequality	0.006	0.006	0.007	0.007
Share of Total Inequality Attributable to Between-Group Inequality	1.83%	3.10%	3.77%	3.75%

Table 12. Atkinson Index of Inequality in *Reduction* in $PM_{2.5}$ Concentrations Due to the HDD Rule in the Total Population and in Racial/Ethnic Subgroups

Racial/Ethnic Subgroup	Atkinson Index		
	ε = 0.5	ε = 1	ε = 2
Total Population	0.088	0.184	0.401
African-American	0.0670	0.135	0.278
Asian-American	0.068	0.137	0.286
Native American	0.167	0.344	0.629
White Hispanic	0.096	0.202	0.442
White non-Hispanic	0.088	0.182	0.394
Within-Group Inequality	0.085	0.175	0.375
Between-Group Inequality	0.004	0.010	0.042
Share of Total Inequality Attributable to Between-Group Inequality	4.45%	5.59%	10.00%

Similarly, indices for the incidence of different health effects could also be calculated using the same method. Due to limited time and resources, however, we do not present those results here.

3.3.3 Comparisons by Region

For some rules, impacts are geographically clustered – i.e., reductions are predicted to occur as a result of the rule only, or primarily, in some regions of the country. For such rules, some (possibly substantial) proportion of the population will experience no reduction in the pollutant, and correspondingly no reduction in incidence of any health effects, as a result of the rule. In

such cases, a large proportion of zeros will substantially affect summary statistics such as population-weighted means or population-weighted percentiles of distributions of reductions. Regional analyses, which allow us to compare subgroups within specified "hot spot" regions, may be more informative in such cases.

Regional clustering of impacts is not as pronounced for the HDD Rule as for some other rules such as the PM NAAQS (U.S. EPA, 2006). There are, however, regional differences that may be worth highlighting. We illustrate this in Table 13, which shows regional results for hospital admissions for all cardiovascular illnesses (except myocardial infarctions) among adults, age 18 – 64.

Table 13. Absolute and Relative *Reduction* in $PM_{2.5}$ Concentrations and Hospitalizations (per Million Population) for All Cardiovascular Illnesses (Except Myocardial Infarctions) Among Adults, Ages 18 – 64: Regional Results

Region / Race	Baseline Incidence per 100,000 Pop.	Absolute Reduction in $PM_{2.5}$ Level (ug/m^3)	Relative Reduction in $PM_{2.5}$ Level (ug/m^3)	Absolute Reduction in Incidence per Million Pop.	Relative Reduction in Incidence per Million Pop.	Population (millions)	Percent of Population in Region
Region 1: The Northeast							
Asian-American	351.2	0.93	1.2	4.5	0.4	3.9	24.1%
African-American	1429.2	0.94	1.2	18.7	1.6	5.1	16.4%
Native American	1369.9	0.77	1.0	14.2	1.2	0.2	7.6%
White	1130.3	0.76	0.9	11.8	1.0	26.7	16.1%
Total Population	1088.5	0.80					
Region 2: The Midwest							
Asian-American	636.6	0.64	1.1	5.8	0.8	2.2	13.8%
African-American	1501.5	0.70	1.2	14.7	2.0	5.5	17.7%
Native American	1148.8	0.39	0.7	6.5	0.9	0.4	15.6%
White	789.8	0.58	1.0	6.3	0.9	36.2	21.9%
Total Population	872.9	0.59					
Region 3: The South							
Asian-American	419.5	0.88	1.3	5.0	0.5	4.1	25.3%
African-American	1510.1	0.75	1.1	15.6	1.7	17.3	56.1%
Native American	885.6	0.57	0.9	7.0	0.7	0.7	28.2%
White	925.2	0.63	0.9	7.9	0.8	59.8	36.2%
Total Population	1023.2	0.67					
Region 4: The West							
Asian-American	896.1	0.43	1.3	5.5	1.7	6.0	36.8%
African-American	1426.2	0.39	1.2	7.7	2.4	3.0	9.9%
Native American	921.3	0.23	0.7	3.0	0.9	1.2	48.6%
White	588.0	0.32	1.0	2.6	0.8	42.5	25.7%
Total Population	678.7	0.34					

We noted above that for some rules pollutant changes are clustered regionally. In such cases, there may be differential impacts on different subgroups if some are located disproportionately in pollutant "hot spots" or in "hot spots" of pollutant change. Table 13 suggests that, for the HDD rule, this is the case to only a limited extent. The population-weighted average absolute reduction in $PM_{2.5}$ concentration ranges from 0.80 ug/m^3 in the Northeast to 0.34 ug/m^3 in the

West. The relative reductions in $PM_{2.5}$ concentration predicted to be experienced by the different subgroups, however, do not change substantially from one region to another. For some subgroups, however, baseline incidence rates vary across the regions, and that in turn affects regional health impacts. Asian-Americans, for example, are predicted to experience a 20 percent greater reduction in $PM_{2.5}$ concentration, relative to the total population in the Northeast, and a 10 percent greater relative reduction in the Midwest. However, the baseline incidence rate for cardiovascular hospital admissions in this subgroup is almost double in the Midwest what it is in the Northeast. Because of this, a very minor difference in the relative reduction in $PM_{2.5}$ concentration experienced by Asian-Americans in the Northeast versus the Midwest becomes a doubling in relative reduction in health effect (from 0.4 in the Northeast to 0.8 in the Midwest).

4 Discussion

We saw that, for a national air pollution rule, those subgroups that are disproportionately exposed to higher baseline pollutant concentrations tend to enjoy greater reductions in pollutant concentrations as a result of the rule. This is not surprising, since rules tend to target the areas of worst pollution. Therefore national air pollution rules would be expected to have an equalizing effect (although the differences within subgroups were much greater than those between subgroups for both baseline $PM_{2.5}$ concentrations and reductions in concentrations).

We also saw that the reduction in air pollutant concentrations did not necessarily translate into an equivalent reduction in health effect incidence in the different subgroups – e.g., the subgroup that experiences the largest reduction in pollutant concentration as a result of a rule does not necessarily also experience the largest reduction in incidence of adverse health effects associated with the pollutant. This is because another factor – the baseline incidence rate of the adverse health effect – affects each subgroup's population health response to a reduction in pollutant concentration, and these baseline incidence rates vary substantially across racial and ethnic subgroups.

Environmental justice or distributional analyses were originally developed to address a common hypothesis that environmental disamenities locate disproportionately in poor or predominantly minority communities in part *because of* the socio-demographic makeup of those communities. While $PM_{2.5}$ is generated to some extent by stationary sources (e.g., power plants), where someone had to decide where to locate the source of pollution, $PM_{2.5}$ can travel great distances and it can form in so-called "secondary" reactions in the atmosphere, many miles from the original sources of the precursor emissions. This is an important consideration, particularly in interpreting the results of a national distributional analysis of air quality. If we see differences in pollutant concentrations to which the members of one subgroup are exposed versus those in other subgroups, it does not necessarily follow that these differences are the result of unfair intent. On the other hand, a finding that minorities and/or poorer households are disproportionately exposed to higher levels of air pollution might reflect underlying injustices to the extent that housing patterns reflect unequal access to higher quality areas (Morello-Frosch and Jesdale, 2006). In general, it is more difficult to discern the *why* of any observed differences among subgroups for regional air pollutants than for local pollutants.

Therefore, in our distributional analysis of a national air quality regulation, we are not asking *why* there are differences in the levels to which different groups are exposed, but only whether there are differences. We believe that the individual-based method discussed and illustrated above is best suited to answering this question, because it effectively considers all members of each subgroup and tallies results within subgroups (as opposed to, e.g., comparing "minority communities" which also contain whites, with "white communities" which also contain minorities).

Similarly, we are not asking why different subgroups may benefit differentially from a rule or regulation, but simply whether or not they do benefit differentially – in terms of the reductions in air pollution they experience as a result of the rule and in terms of the health risk reductions they

enjoy as a result of the reductions in air pollution. We believe that the individual-based method discussed above is similarly best suited to answering these questions.

As we noted above, however, in the assignment of pollutant concentrations, or reductions in pollutant concentrations as a result of the HDD rule, our analysis could only approximate an individual-level analysis, because estimating truly individual-specific pollutant concentrations was not feasible. Instead we assigned the same baseline (and control scenario) pollutant concentration to all individuals within a grid cell, and as a result, any intra-grid cell differences between subgroups were obscured. This is likely to be less of a problem, however, for regional pollutants such as particulate matter and ozone than for more local pollutants, such as carbon monoxide, whose concentrations tend to vary more within any given grid cell.

In the HDD analysis, the grid was relatively coarse – each grid cell is roughly 36 kilometers by 36 kilometers, and even in more recent national analyses, such as the recent Marine Engine Rule (U.S. EPA, 2008), the resolution is still relatively coarse at roughly 12 kilometers by 12 kilometers. The more the grid cell size can be reduced in such analyses, the better will be the approximation to a truly individual-level distributional analysis. Still, because of an individual's normal mobility during the course of the day, it isn't clear how much reduction is actually necessary to achieve reasonable estimates of individual-specific pollutant concentrations. It would be instructive to progressively reduce the grid cell size in a distributional analysis and observe how grid cell size affects the results of the analysis.

To assess whether pollution affects some subgroups disproportionately, some studies (e.g., Apelberg, et al., 2005; Morello-Frosch, et al., 2002) have applied regression techniques and statistical tests to what appear to be complete censuses rather than random samples (e.g., all the census tracts in a given state), and have reported "statistically significant" results. "Statistical significance," however, is a meaningful concept only when an analysis is based on a random sample (rather than the entire population of interest). "Statistical significance" suggests that what we observe in the sample indicates something real about the population, rather than being due to random chance (i.e., to the particular sample we randomly drew from the population). If we are observing the entire population (e.g., all the census tracts in a state), then we should not use statistical tests, as "statistical significance" is meaningless.

Rather than "statistical significance," the relevant question is whether observed differences between populations (e.g., between minorities and non-minorities) are worthy of concern. At what point should any observed differences be considered disproportionate? There is no easy answer.

We *can* say that the differences we observe in our distributional analysis are likely to be real differences – in particular, we know that they are not just due to sampling error, since our distributional analysis uses a *complete census* of each subgroup rather than samples. However, because we cannot measure actual individual-level pollutant concentrations and must instead assign the same concentration to all individuals in a grid cell, our method may misstate the degree of difference between subgroups, since it obscures any intra-grid cell heterogeneity. In particular, our method will understate differences between subgroups if the differences we see when we look across grid cells also exist within grid cells. For example, suppose that, looking

across grid cells (but ignoring heterogeneity within grid cells), we observe that minorities tend to be exposed to higher pollution levels than non-minorities. If this also holds true within grid cells, ignoring these intra-grid cell differences will cause us to understate the overall difference between minorities and non-minorities.

However, even if there were no bias in our results, there is a legitimate question as to what magnitude of differences between subgroups constitutes "environmental injustice." Since it is highly improbable that all subgroups would have *exactly* the same baseline pollutant concentrations or reductions in pollutant concentrations, there will necessarily be differences. There is no objective degree of difference beyond which we definitively conclude that there is "environmental injustice."

One useful set of tools for considering this problem, as noted above, are inequality indices that allow a comparison of within-group and between-group variability. In our HDD Rule case study, for example, we found that there is far more inequality in pollutant concentrations among individuals within subgroups than between them. It would be instructive to examine the extent to which this result holds as we decrease the grid cell size in our distributional analysis. If (1) there is little variability in ambient PM concentrations within grid cells and/or (2) the intra-grid cell distributions for the different subgroups are similar, then the result should basically hold.

Appendix A. Methodological Details of Distributional Analysis of Benefits of the Heavy Duty Diesel Rule

This Appendix presents the methodology and a description of the inputs used for estimating air quality and health risks for the Heavy Duty Diesel rule.

A.1. Overview of a Typical Benefit Analysis

The typical benefit analysis for a rule or regulation targeting a criteria air pollutant proceeds through the following seven steps:

Step 1: Estimate baseline and control scenario emissions from industrial and other emission sources.

Step 2: Input the baseline emissions to an air quality model (AQM) that incorporates atmospheric chemistry and relevant weather and climate variables. The AQM provides as output the ambient levels of the relevant air pollutant(s) in each grid cell in a specified grid covering the United States. Different AQMs use different grid sizes which depend to a large extent on the computational abilities of the computers doing the modeling.

Step 3: Input the control scenario emissions to an AQM which provides as output the ambient levels of the relevant air pollutant(s) in each grid cell in a specified grid covering the United States.

Step 4: Input the grid cell-specific baseline and control scenario ambient air pollutant concentrations output from the AQM to a benefits model such as BenMAP. The benefits model may have its own grid (in which case, air pollutant concentrations in the grid cells must be interpolated from the air pollutant concentrations in the grid cells of the AQM, or from monitors, if monitor data are used instead). In the case of BenMAP, the grid can be specified to conform to the grid of any AQM.

Step 5: Calculate the number of cases of each specified health effect avoided. Given the *change* in air pollutant concentration in a grid cell, from baseline to control scenario, the benefits model calculates the number of cases of each specified health effect avoided (e.g., the number of cases of mortality avoided) as a result of that change for each grid cell in its grid. The benefits model requires the following inputs to calculate cases avoided within a grid cell:

a) The population within the grid cell;

b) The baseline incidence rate for the specified health effect;

c) The change in ambient air pollutant concentration (from baseline to control scenario); and

d) A health impact function giving the change in the number of cases of the specified health effect corresponding to a given change in ambient air pollutant concentration.

Step 6: Aggregate across grid cells to derive total number of cases avoided. After grid-cell-specific numbers of cases avoided are calculated, the benefits model aggregates across grid cells to derive a total number of cases avoided.

Step 7: Cases avoided may be monetized. This is typically done in benefits assessments; however, it seems unnecessary for a distributional analysis, since health impact valuation for key health effects, such as mortality, does not vary by demographic subgroup.

A.2. Methodological Details of the Heavy Duty Diesel Distributional Analysis

The benefit analysis for the Heavy Duty Diesel (HDD) Rule was completed by Abt Associates in 2000. Therefore, the baseline and control scenario emissions, and the corresponding baseline and control scenario ambient pollutant concentrations (in this case, $PM_{2.5}$ concentrations) have already been estimated. Abt Associates thus completed Steps 1 through 3 of a typical benefit analysis, as described above. We had also completed Steps 4 through 6 using Criteria Air Pollutant Modeling System (CAPMS), the precursor to BenMAP. However, Steps 4 through 6 may also be completed using the most recent version of BenMAP (version 3), which is the approach we chose. In particular, we took the air quality modeling data that came out of the AQM used in the benefit analysis for the HDD rule (the output from Steps 2 and 3 above) and input them to the most recent version of BenMAP, which already has incorporated in it 304 demographic subgroups that allows for easy analysis of distributional issues.[17] The inputs to the distributional analysis of the HDD rule are described below.

A.2.1 Air Quality

To estimate air quality, we followed the same general approach used in the regulatory impact analyses (RIAs) for the HDD rule. A key difference is that we used BenMAP, version 3.0, as opposed to using the CAPMS benefit models used in the original HDD analysis. BenMAP 3.0 incorporates the necessary demographic variables for an environmental justice analysis that are absent from the earlier benefits models.

EPA used the Regulatory Model System for Aerosols and Deposition (REMSAD) to model $PM_{2.5}$ levels for the HDD rule. Unlike in more recent RIAs, which combined both modeling and monitoring data to forecast $PM_{2.5}$ levels, EPA forecast $PM_{2.5}$ for HDD using only modeling data.

[17] The calculation of health effects at the 8 x 8 kilometer grid cell level is internal to CAPMS which reports results only at the county level. Furthermore, CAPMS applies county-level data evenly across all grid cells in a county which is not representative of the demographic composition of the grid cells. BenMAP's grid cell-level data negates the necessity to make assumptions about grid-cell level demographic composition.

The technical support document associated with the RIA for the HDD rule describes this in detail (Abt Associates Inc., 2000; 2003).

Table 14 presents the population-weighted $PM_{2.5}$ levels for the baseline and for the difference between baseline and the control, or "delta," where the delta equals the baseline minus the control. The baseline and delta $PM_{2.5}$ values reported by BenMAP 3.0 compare reasonably well with those from the RIA for the HDD rules. To the extent that there are differences, this is most likely due to the different approaches used in BenMAP and CAPMS to forecast 2030 population. For example, CAPMS assumes the demographic structure of the population remains constant over time, while BenMAP takes into account a changing demographic structure, such as greater life expectancies and a greater proportion of elderly in the future.

Table 14. Air Quality Metrics for HDD Rule

					Baseline		Delta	
Analsyis	**Year**	**Pollutant**	**Location**	**Metric**	**Ver 3.0**	**Original**	**Ver 3.0**	**Original**
HDD	2030	$PM_{2.5}$	U.S.	Annual mean	14.65	14.85	0.64	0.65

NOTES: Units = $PM_{2.5}$: micrograms per meter cubed (ug/m^3).
Delta = Baseline minus control.
Source for statistics from "original" analysis: *see*: U.S. EPA (2000, Tables VII-2 and VII-4).

A.2.2 Population Forecast for 2030

The air quality model grid cells typically cross Census and jurisdictional boundaries, so population data – a critical component for a distributional analysis – are not readily available for each grid cell. A separate application developed by Abt Associates, called "PopGrid," assigns year 2000 Census block data to the REMSAD grid cells used in the calculation of air quality and health impacts. As described below, BenMAP then combines the year 2000 population data at each REMSAD grid cell with county-level population forecasts to estimate 2030 population levels for each REMSAD grid cell.

To calculate the population in each REMSAD grid cell, PopGrid aggregates year 2000 block data, which is the most detailed data available from the Census Bureau. Each block generally has a few hundred individuals.[18] If the center of a block falls within a grid cell, PopGrid assigns the block's population to that grid cell. Figure 9 in Appendix B graphically shows this relationship.

After the aggregation, BenMAP has 304 unique race-ethnicity-gender-age groups in each REMSAD grid cell: 19 age groups by 2 ethnic groups by gender by 4 racial groups (19*2*2*4=304). Table 15 presents the 304 population variables available in BenMAP. As discussed below, BenMAP uses these variables to develop the necessary population estimates for each race-ethnicity-age subgroup.

[18] Blocks and blockgroups are defined at: http://www.census.gov/geo/www/geo_defn.html. Blockgroups generally have 600 to 3,000 individuals. Since blocks comprise blockgroups, we estimate blocks generally have a few hundred individuals.

Table 15. Demographic Groups and Variables Available in BenMAP

Racial/Ethnic Group	Ethnicity	Age	Gender
White, African American, Asian, American Indian, Other, Hispanic	Hispanic, Non-Hispanic	<1, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+	Male, Female

In calculating the population in age groups that may include a portion of one of the pre-specified demographic groups in Table 15, BenMAP assumes the population is uniformly distributed in the age group. For example, to calculate the number of children ages 3 through 12, BenMAP calculates:

$$age_{3-12} = \frac{1}{2} \cdot age_{1-4} + age_{5-9} + \frac{3}{5} \cdot age_{10-14} \quad .$$

To estimate population levels for the years after the last Census in 2000, BenMAP scales the 2000 Census-based estimate with the ratio of the county-level forecast for the future year of interest over the 2000 county-level population level. Woods & Poole (2007) provides the county-level population forecasts used to calculate the scaling ratios; these data are discussed in detail below.

In the simplest case, where one is forecasting a single population variable, say, children ages 4 to 9, CAMPS calculates:

$$age_{4-9,\, g,\, 2010} = age_{4-9,\, g,\, 2000} \cdot \frac{age_{4-9,\, county,\, 2010}}{age_{4-9,\, county,\, 2000}}$$

where the gth population grid-cell is wholly located within a given county. (Note that while this example is for 2010, the same process holds for the 2030 population estimates used in our analysis.)

In the case where the gth grid-cell includes "n" counties in its boundary, the situation is somewhat more complicated. BenMAP first estimates the fraction of individuals in a given age group (e.g., ages 4 to 9) that reside in the part of each county within the gth grid-cell. BenMAP calculates this fraction by simply dividing the population all ages of a given county within the gth grid-cell by the total population in the gth grid-cell:

$$fraction\ of\ age_{4-9,\, g\ in\ county_c} = \frac{age_{all,\, g\ in\ county_c}}{age_{all,\, g}}$$

Multiplying this fraction by the number of individuals ages 4 to 9 in the year 2000 gives an estimate of the number of individuals ages 4 to 9 that reside in the fraction of the county within the gth grid-cell in the year 2000:

$$age_{4-9,\, g\ in\ county_c,\, 2000} = age_{4-9,\, g,\, 2000} \cdot fraction\ age_{4-9,\, g\ in\ county_c}$$

To then forecast the population in 2010, we scale the 2000 estimate with the ratio of the county projection for 2010 to the county projection for 2000:

$$age_{4-9,\, g\ in\ county_c,\, 2010} = age_{4-9,\, g\ in\ county_c,\, 2000} \cdot \frac{age_{4-9,\, county_c,\, 2010}}{age_{4-9,\, county_c,\, 2000}} \quad .$$

Combining all these steps for "n" counties within the gth grid-cell, we forecast the population of persons ages 4 to 9 in the year 2010 as follows:

$$age_{4-9,\, g,\, 2010} = \sum_{c=1}^{n} age_{4-9,\, g,\, 2000} \cdot \frac{total\ pop_{g\ in\ county_c}}{total\ pop_g} \cdot \frac{age_{4-9,\, county_c,\, 2010}}{age_{4-9,\, county_c,\, 2000}}$$

In the case where there are multiple age groups and multiple counties, BenMAP first calculates the forecasted population level for individual age groups, and then combines the forecasted age groups. In calculating the number of children ages 4 to 12, BenMAP calculates:

$$age_{4-9,\, g,\, 2010} = \sum_{c=1}^{n} age_{4-9,\, g,\, 2000} \cdot \frac{total\ pop_{g\ in\ county_c}}{total\ pop_g} \cdot \frac{age_{4-9,\, county_c,\, 2010}}{age_{4-9,\, county_c,\, 2000}}$$

$$age_{10-14,\, g,\, 2010} = \sum_{c=1}^{n} age_{10-14,\, g,\, 2000} \cdot \frac{total\ pop_{g\ in\ county_c}}{total\ pop_g} \cdot \frac{age_{10-14,\, county_c,\, 2010}}{age_{10-14,\, county_c,\, 2000}}$$

$$age_{4-12,\, g,\, 2010} = age_{4-9,\, g,\, 2010} + \frac{3}{5} \cdot age_{10-14,\, g,\, 2010} \quad .$$

A.2.3 Concentration-response functions

Table 16 presents the health impact functions and estimated cases of adverse health effects from the original HDD analysis and the present analysis. The estimated numbers are not directly comparable because different health impact functions, different incidence rates and different population estimates are used. For the present analysis, we have used more recent functions.

Table 16. Epidemiological Studies Used and Estimated Cases of Adverse Health Effects

Analysis	Health Effect	Age	Current Analysis	Estimated Cases
Original HDD Analysis	Mortality	Infants	Woodruff et al (1997)	34
		30+	Krewski et al (2000)	8,300
	Hospital Admissions			
	COPD	65+	Samet et al (2000)	900
	Pneumonia	65+	Samet et al (2000)	1,100
	Asthma	0-64	Sheppard et al (1999)	880
	Cardiovascular	65+	Samet et al (2000)	2,700
	ER Visits, Asthma	< 65	Schwartz et al (1993)	2,100
Present Analysis	Mortality	Infants	Woodruff et al (2006)	100
		30+	Pope et al (2002)	10,100
	Hospital Admissions			
	COPD (less asthma)	18-64	Moolgavkar (2000)	410
	COPD	65+	Moolgavkar (2003)	1,100
			Ito (2003)	730
	Pneumonia	65+	Ito (2003)	3,600
	Asthma	0-64	Sheppard et al (2003)	1,100
	Heart Attacks	18-64	Peters et al (2001)	5,700
		65+		12,100
	Ischemic Heart Disease	65+	Ito (2003)	1,300
	Cong. Heart Failure	65+	Ito (2003)	2,600
	Dysrhythmia	65+	Ito (2003)	660
	All cardiovasular	16-64	Moolgavkar (2003)	1,700
		65+	Moolgavkar (2003)	4,400
	ER Visits, Asthma	0-17	Norris et al (1999)	8,400

A.2.4 Baseline incidence data

Concentration-Response (C-R) functions developed from log-linear or logistic models estimate the percent change in an adverse health effect associated with a given pollutant change. In order to estimate the absolute change in incidence using these functions, we need the baseline incidence of the adverse health effect. This is typically calculated as the product of the incidence rate (per person) and the population. Below, we describe the approach we used to calculate incidence rates for mortality, hospital admissions, and ER visits. For mortality and hospital admissions, we calculated incidence rates varying by race and age. And in the case of ER visits, we calculated incidence rates varying by race, ethnicity, and age.

A.2.4.1 All-cause mortality

Age, race, and county-specific mortality data were obtained from the U.S. Centers for Disease Control (CDC) for the years 1996 through 1998. CDC maintains an online data repository of health statistics, CDC Wonder, accessible at http://wonder.cdc.gov/. The mortality rates provided are derived from U.S. death records and U.S. Census Bureau post-central population estimates. Mortality rates were averaged across three years (1996 through 1998) to provide more stable estimates.

When estimating rates for age groups that differed from the CDC Wonder groupings, we assumed that rates were uniform across all ages in the reported age group. For example, to estimate mortality rates for individuals aging between 1 and 17, we scaled the 15-19 year old death count and population by 3/5 and then generated a population-weighted mortality rate combining data for the younger age groups.

CDC data record three race groups: White, black and other. The mortality rates were first calculated for these three races. Then we split the "other" race to Native American and Asian race groups and assigned them the same mortality rates as those in the "other" race group in order to match the input format of BenMAP.

The county-level mortality rates from the CDC Wonder website are not considered reliable if the number of deaths is less than 20. In these cases we summarized the death counts and population to state level and calculated the state-level mortality rates.

To obtain the predicted mortality rates in 2030, we divided the projected mortality rates from census life tables[19] by the estimated rates of 1997 to calculate the calibrated ratios. Then we applied the calibrated ratios to adjust our estimated age, race and county-specific mortality rates in order to get the predicted age, race and county-specific mortality rates in 2030. Table 17 presents the national mortality rates (all-cause) by age group and race.

Table 17. Mortality Rates for All-Cause Mortality, by Age Group and Race

Race	Mortality Rate by Age Group (deaths per 1000 people per year)									
	Infant	1-17	18-24	25-34	35-44	45-54	55-64	65-74	75-84	85+
White	6.0	0.3	0.8	1.0	1.8	3.9	9.9	24.3	56.7	153.7
Black	14.4	0.5	1.5	2.2	4.1	8.2	17.2	34.6	66.3	141.9
Other	4.9	0.3	0.6	0.7	1.2	2.5	6.2	15.3	38.3	105.8

Source: Original data were from 1996-1998 from the CDC Wonder (http://wonder.cdc.gov/) and were summarized by age group and race. Predicted county-specific rates of 2030 are used in the C-R functions.

A.2.4.2 Hospitalization

Regional hospitalization counts were obtained from the National Center for Health Statistics' (NCHS) National Hospital Discharge Survey (NHDS). NHDS is a sample-based survey of non-Federal, short-stay hospitals (<30 days),[20] and is the principal source of nationwide hospitalization data. The survey collects data on patient characteristics, diagnoses, and medical procedures.

[19] Data source: http://www.census.gov/population/www/projections/natdet-D5.html

[20] The following hospital types are excluded from the survey: hospitals with an average patient length of stay of greater than 30 days, federal, military, Department of Veterans Affairs hospitals, institutional hospitals (e.g. prisons), and hospitals with fewer than six beds.

Public use data files for the year 1999 survey were downloaded and processed to estimate hospitalization counts by region, race, age group and endpoint.[21] NCHS groups states into four regions using the following groupings defined by the U.S. Bureau of the Census:

- Northeast - Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, New York, New Jersey, Pennsylvania
- Midwest - Ohio, Indiana, Illinois, Michigan, Wisconsin, Minnesota, Iowa, Missouri, North Dakota, South Dakota, Nebraska, Kansas
- South - Delaware, Maryland, District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida, Kentucky, Tennessee, Alabama, Mississippi, Arkansas, Louisiana, Oklahoma, Texas
- West - Montana, Idaho, Wyoming, Colorado, New Mexico, Arizona, Utah, Nevada, Washington, Oregon, California, Alaska, Hawaii

The race categories recorded by NHDS are as follows:
1 = White
2 = Black
3 = American Indian
4 = Asian/Pacific Islander
5 = Other
9 = Not Stated

The "other" and "not stated" are nuisance categories and we assigned them to the other four race categories based on the existing distribution of cases among the other four races. For example, for the Northeastern region and infant age group, there are 32432 (65.9%), 14563 (29.6%), 1121 (2.3%) and 1094 (2.2%) cases of "All Respiratory Hospital Admission" for white, black, American Indian and Asian race groups respectively. The total cases for "other" and "not stated" are 15803, so we assign 69.5%, 29.6%, 2.3% and 2.2% of 15803 cases to white, black, American Indian and Asian respectively.

We calculated per capita hospitalization rates, by dividing these counts by the estimated 2000 population estimates in each subgroup defined by region, age group, race and endpoint combination that were derived from the U.S. Bureau of the Census. Note that NHDS started with hospital admission counts, based on a sample of admissions, and then they used population estimates to generate population-weighted hospital admission counts that are representative of each region. This weighting used forecasts of 1999 population data. Ideally, we would use these same forecasts to generate our admission rates. However, while NHDS presented counts of hospital admissions with a high degree of age specificity, it presented regional population data for only four age groups: 0-14, 15-44, 45-64, and 65+.[4] Using only the NHDS data, we would be limited to calculating regional admission rates for four groups. Because we are interested in a broader range of age groups, we turned to the 2000 Census[5].

[21] Data are available at ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHDS/

[4] See: 1999nhds_summary.pdf (p. 187) for published regional population estimates for 1999.

[5] We realized that using the 2000 population and 1999 hospitalization counts could underestimate the rates a little bit given the consideration of population growth.

The endpoints in hospitalization studies are defined using different combinations of ICD codes. For the purposes of this analysis, we identified a core group of endpoints and calculated their incidence rate for use in the C-R functions:

- 1= Acute Myocardial Infarction, Nonfatal (ICD-9 410)
- 2= HA, All Cardiovascular less Myocardial Infarctions (ICD-9 390-409, 411-459)
- 3= HA, All Respiratory (ICD-9 460-519)
- 4= HA, Asthma (ICD-9 493)
- 5= HA, Chronic Lung Disease (ICD-9 490-496)
- 6= HA, Chronic Lung Disease less Asthma (ICD-9 490-492, 494-496)
- 7= HA, Congestive Heart Failure (ICD-9 428)
- 8= HA, Dysrhythmia (ICD-9 427)
- 9= HA, Ischemic Heart Disease less Myocardial Infarctions (ICD-9 411-414)
- 10= HA, Pneumonia (ICD-9 480-487)

For each C-R function obtained from the epidemiologic studies, we selected the baseline rate or combination of rates that most closely matches to the study endpoint definition. For studies that define chronic lung disease as ICD 490-492, 494-496, we subtracted the incidence for asthma (ICD 493) from the chronic lung disease (ICD 490-496). In some cases, the baseline rate will not match exactly to the endpoint definition in the study. For example, Burnett et al. (2001) studied the following respiratory conditions in infants <2 years of age: ICD 464.4, 466, 480-486, 493. For this C-R function we apply an aggregate of the following rates: ICD 464, 466, 480-487, 493. Although they do not match exactly, we assume that relationship observed between the pollutant and study-defined endpoint is applicable for the additional codes. Table 18 presents the hospitalization rates estimates by endpoint, race and age group.

Table 18. Hospitalization Rates by Endpoint, Race and Age Group

Endpoint*	Race	Hospitalization Rate by Age Group (deaths per 1000 people per day)									
		0-1	*2-17*	*18-24*	*25-34*	*35-44*	*45-54*	*55-64*	*65-74*	*75-84*	*85+*
1	ASIAN	0.0000	0.0000	0.0000	0.0000	0.0022	0.0028	0.0123	0.0211	0.0529	0.0016
1	BLACK	0.0000	0.0000	0.0001	0.0002	0.0019	0.0067	0.0192	0.0236	0.0421	0.0609
1	NATA	0.0000	0.0000	0.0000	0.0000	0.0034	0.0021	0.0088	0.0451	0.0923	0.0438
1	WHITE	0.0000	0.0000	0.0000	0.0004	0.0026	0.0090	0.0182	0.0315	0.0469	0.0625
2	ASIAN	0.0000	0.0000	0.0000	0.0020	0.0050	0.0220	0.0526	0.0989	0.2113	0.4385
2	BLACK	0.0029	0.0007	0.0011	0.0100	0.0286	0.0677	0.1130	0.2089	0.2419	0.2981
2	NATA	0.0000	0.0002	0.0001	0.0049	0.0136	0.0247	0.1072	0.2486	0.4456	0.4288
2	WHITE	0.0025	0.0006	0.0015	0.0027	0.0101	0.0306	0.0720	0.1464	0.2285	0.3012
3	ASIAN	0.1027	0.0098	0.0047	0.0026	0.0053	0.0092	0.0212	0.0747	0.1536	0.2903
3	BLACK	0.2230	0.0279	0.0134	0.0158	0.0241	0.0436	0.0652	0.1030	0.1839	0.3279
3	NATA	0.1272	0.0122	0.0022	0.0106	0.0090	0.0090	0.0320	0.0924	0.2957	0.6510
3	WHITE	0.1384	0.0127	0.0065	0.0079	0.0109	0.0187	0.0440	0.0965	0.1718	0.2612
4	ASIAN	0.0102	0.0051	0.0003	0.0004	0.0018	0.0013	0.0037	0.0043	0.0099	0.0568
4	BLACK	0.0430	0.0174	0.0038	0.0066	0.0079	0.0099	0.0139	0.0119	0.0213	0.0054
4	NATA	0.0156	0.0073	0.0000	0.0024	0.0020	0.0000	0.0017	0.0014	0.0082	0.0000
4	WHITE	0.0161	0.0040	0.0020	0.0025	0.0020	0.0033	0.0034	0.0043	0.0050	0.0065
5	ASIAN	0.0102	0.0051	0.0003	0.0006	0.0022	0.0018	0.0071	0.0193	0.0320	0.0744
5	BLACK	0.0458	0.0175	0.0043	0.0067	0.0086	0.0170	0.0307	0.0449	0.0680	0.0665
5	NATA	0.0156	0.0073	0.0000	0.0023	0.0020	0.0024	0.0035	0.0179	0.0296	0.0919
5	WHITE	0.0169	0.0042	0.0022	0.0030	0.0035	0.0075	0.0190	0.0380	0.0517	0.0406
6	ASIAN	0.0000	0.0000	0.0000	0.0001	0.0004	0.0006	0.0037	0.0149	0.0215	0.0419
6	BLACK	0.0029	0.0000	0.0004	0.0000	0.0011	0.0070	0.0173	0.0333	0.0466	0.0603
6	NATA	0.0000	0.0000	0.0000	0.0000	0.0000	0.0023	0.0022	0.0166	0.0203	0.0870
6	WHITE	0.0008	0.0002	0.0002	0.0005	0.0015	0.0042	0.0155	0.0336	0.0467	0.0339
7	ASIAN	0.0000	0.0000	0.0000	0.0003	0.0001	0.0026	0.0061	0.0167	0.0630	0.2230
7	BLACK	0.0001	0.0001	0.0000	0.0015	0.0055	0.0168	0.0344	0.0653	0.0711	0.1306
7	NATA	0.0000	0.0000	0.0000	0.0026	0.0030	0.0020	0.0090	0.0596	0.0483	0.2531
7	WHITE	0.0005	0.0000	0.0002	0.0001	0.0009	0.0029	0.0108	0.0309	0.0743	0.1357
8	ASIAN	0.0000	0.0000	0.0000	0.0011	0.0002	0.0014	0.0049	0.0166	0.0284	0.0972
8	BLACK	0.0003	0.0002	0.0005	0.0016	0.0024	0.0057	0.0122	0.0226	0.0363	0.0405
8	NATA	0.0000	0.0003	0.0002	0.0001	0.0011	0.0032	0.0056	0.0316	0.1201	0.0346
8	WHITE	0.0004	0.0003	0.0005	0.0006	0.0021	0.0043	0.0110	0.0287	0.0481	0.0621
9	ASIAN	0.0000	0.0000	0.0000	0.0000	0.0001	0.0116	0.0290	0.0452	0.0852	0.0550
9	BLACK	0.0000	0.0000	0.0000	0.0003	0.0047	0.0171	0.0325	0.0537	0.0613	0.0337
9	NATA	0.0000	0.0000	0.0000	0.0004	0.0014	0.0138	0.0769	0.1362	0.1687	0.1212
9	WHITE	0.0008	0.0000	0.0002	0.0006	0.0040	0.0166	0.0386	0.0622	0.0688	0.0532
10	ASIAN	0.0303	0.0039	0.0005	0.0016	0.0019	0.0046	0.0137	0.0403	0.0684	0.1277
10	BLACK	0.0678	0.0058	0.0047	0.0055	0.0082	0.0155	0.0234	0.0318	0.0723	0.1427
10	NATA	0.0441	0.0039	0.0010	0.0092	0.0064	0.0012	0.0010	0.0441	0.1828	0.4879
10	WHITE	0.0366	0.0044	0.0015	0.0024	0.0038	0.0061	0.0149	0.0374	0.0767	0.1545

Source: As described in the text, we obtained the regional count of hospital admissions from National Hospital Discharge Survey (NHDS), and we obtained the population data from the 2000 U.S. Census. * Endpoints described on the previous page.

A.2.4.3 Emergency room visits for asthma

Regional asthma emergency room (ER) visit counts were obtained from the National Hospital Ambulatory Medical Care Survey (NHAMCS). NHAMCS is a sample-based survey, conducted by NCHS, designed to collect national data on ambulatory care utilization in hospital emergency and outpatient departments of non-Federal, short-stay hospitals (<30 days).[5]

Public use data files for the year 2000 survey were downloaded[6] and processed to estimate ER visit counts by region, age group, race and ethnicity. There are five race categories from NHAMCS as shown below. We grouped 3, 4, and 5 to be "Other" race group given the lack of data for American Indians.

1 = White
2 = Black
3 = Asian or Native Hawaiian/Other Pacific Islander
4 = American Indian/Alaska Native
5 = More than one race reported

NHAMCS also records ethnicity information that divides people into three ethnic groups: Hispanic, non-Hispanic and blank ethnicity. Blank category is the nuisance one and we assigned the Blank cases to Hispanic and non-Hispanic using the relative proportions of cases already assigned to Hispanics and non-Hispanics. For example, for Northeastern Region, age group 0-17 and black race, there are 8733 cases (32.9%) for Hispanics and 17813 cases (67.1%) for non-Hispanics, so we would take the 3285 cases in the Blank group and assign 32.9% to Hispanics and 67.1% to non-Hispanics. For cases where blank is the only ethnic group, we use the ratio of Hispanic and non-Hispanic in the corresponding age group to assign the counts in the blank category.

After obtaining the ER visit counts in each region, age group, race and ethnicity combination, we divided these counts by the corresponding population estimates from the 2000 U.S. Census to calculate the ER incidence rates. Table 19 presents the estimated asthma emergency room rates by region.

[5] The target universe of the NHAMCS is in-person visits made in the United States to emergency and outpatient departments of non-Federal, short-stay hospitals (hospitals with an average stay of less than 30 days) or those whose specialty is general (medical or surgical) or children's general.

[6] Data are available at ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHAMCS/

Table 19. Emergency Room Visit Rates for Asthma, by Region, Race, Ethnicity and Age Group

Region	Race	Ethnicity	ER Visit Rate by Age Group (deaths per 1000 people per day)								
			0-17	18-24	25-34	35-44	45-54	55-64	65-74	75-84	85+
MidW	ASIAN	HISPANIC	0	0	0	0	0	0	0	0	0
MidW	ASIAN	NON-HISPANIC	0	0	0	0	0	0	0	0	0
MidW	BLACK	HISPANIC	0	0	0	0	0	0	0	0	0
MidW	BLACK	NON-HISPANIC	0.0987	0.0665	0.0849	0.0213	0	0.0079	0.0923	0	0
MidW	NATAMER	HISPANIC	0	0	0	0	0	0	0	0	0
MidW	NATAMER	NON-HISPANIC	0	0	0	0	0	0	0	0	0
MidW	WHITE	HISPANIC	0.0448	0	0	0.0674	0.2112	0	0	0	0
MidW	WHITE	NON-HISPANIC	0.032	0.0525	0.0281	0.0197	0.0103	0.0052	0	0.017	0
NE	ASIAN	HISPANIC	0.0343	0	0	0	0	0	0	0	0
NE	ASIAN	NON-HISPANIC	0	0	0.0174	0.0246	0	0	0.3058	0	0
NE	BLACK	HISPANIC	0.1822	0	0	0	0.008	0.7873	0	0	0
NE	BLACK	NON-HISPANIC	0.0303	0.1157	0.0494	0.0592	0.003	0	0.0408	0.0795	0
NE	NATAMER	HISPANIC	0.0343	0	0	0	0	0	0	0	0
NE	NATAMER	NON-HISPANIC	0	0	0.0174	0.0246	0	0	0.3058	0	0
NE	WHITE	HISPANIC	0.0304	0.0444	0.0447	0.0228	0.0195	0.0532	0	0	0
NE	WHITE	NON-HISPANIC	0.016	0.0355	0.0188	0.0152	0.0099	0.0112	0	0.0027	0
South	ASIAN	HISPANIC	0	0	0	0	0	0	0	0	0
South	ASIAN	NON-HISPANIC	0	0	0	0	0	0	0	0	0
South	BLACK	HISPANIC	0.2121	0	0	0	0	0	0.1239	0	0
South	BLACK	NON-HISPANIC	0.0577	0.0083	0.0175	0.018	0.0176	0.0325	0.0065	0	0
South	NATAMER	HISPANIC	0	0	0	0	0	0	0	0	0
South	NATAMER	NON-HISPANIC	0	0	0	0	0	0	0	0	0
South	WHITE	HISPANIC	0.0453	0	0	0.0027	0	0	0.0221	0	0
South	WHITE	NON-HISPANIC	0.0221	0.0234	0.0195	0.0026	0.0083	0.0138	0.007	0.0034	0
West	ASIAN	HISPANIC	0.0343	0	0	0	0	0	0	0	0
West	ASIAN	NON-HISPANIC	0.0142	0	0.0076	0	0.0109	0.0048	0.0035	0	0
West	BLACK	HISPANIC	0	0	0	0.0024	0	0	0	0	0
West	BLACK	NON-HISPANIC	0.0108	0	0.1036	0.0026	0.0384	0.0436	0	0	0
West	NATAMER	HISPANIC	0.0343	0	0	0	0	0	0	0	0
West	NATAMER	NON-HISPANIC	0.0142	0	0.0076	0	0.0109	0.0048	0.0035	0	0
West	WHITE	HISPANIC	0.0136	0.0133	0.0027	0	0.0157	0	0	0	0
West	WHITE	NON-HISPANIC	0.0081	0.0326	0.013	0.0089	0.0016	0.0094	0.0026	0.0091	0

Source: We obtained ER visit counts for the year 2000 from the National Hospital Ambulatory Medical Care Survey (NHAMCS) and population data were obtained from the 2000 U.S. Census.

Appendix A.3: Definitions and Properties of Inequality Measures

Table 20. Definitions of Inequality Measures (Reprint from Cowell (2005), p. 137)

Name	Definition	Maximum	Transfer effect
Variance	$V = \frac{1}{n}\sum_{i=1}^{n}\left[y_i - \bar{y}\right]^2$	$\bar{y}^2\left[n-1\right]$	$\frac{2}{n}\left[y_j - y_i\right]$
Coefficient of variation	$c = \frac{\sqrt{V}}{\bar{y}}$	$\sqrt{n-1}$	$\frac{y_j - y_i}{n\bar{y}\sqrt{V}}$
Range	$R = y_{\max} - y_{\min}$	$n\bar{y}$	1 if $y_i = y_{\min}$ or $y_j = y_{\max}$, 2 if $y_i = y_{\min}$ and $y_j = y_{\max}$, 0 otherwise
Rel.mean deviation	$M = \frac{1}{n}\sum_{i=1}^{n}\left\lvert\frac{y_i}{\bar{y}} - 1\right\rvert$	$2 - \frac{2}{n}$	$\frac{2}{n\bar{y}}$ if $y_i \le \bar{y} \le y_j$ 0 otherwise
logarithmic variance	$v = \frac{1}{n}\sum_{i=1}^{n}\left[\log\frac{y_i}{\bar{y}}\right]^2$	∞	$\frac{2}{ny_j}\log\frac{y_j}{\bar{y}} - \frac{2}{ny_i}\log\frac{y_i}{\bar{y}}$
variance of logarithms	$v_1 = \frac{1}{n}\sum_{i=1}^{n}\left[\log\frac{y_i}{y^*}\right]^2$	∞	$\frac{2}{ny_j}\log\frac{y_j}{y^*} - \frac{2}{ny_i}\log\frac{y_i}{y^*}$
Gini	$\frac{1}{2n^2\bar{y}}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\lvert y_i - y_j\right\rvert$	$\frac{n-1}{n}$	$\frac{[j]-[i]}{n^2\bar{y}}$
Atkinson	$A_\varepsilon = 1 - \left[\frac{1}{n}\sum_{i=1}^{n}\left[\frac{y_i}{\bar{y}}\right]^{1-\varepsilon}\right]^{\frac{1}{1-\varepsilon}}$	$1 - n^{\frac{-\varepsilon}{1-\varepsilon}}$ or 1^*	$\frac{y_i^{-\varepsilon} - y_j^{-\varepsilon}}{n\bar{y}^{1-\varepsilon}[1-A_\varepsilon]^{-\varepsilon}}$
Dalton	$D_\varepsilon = 1 - \frac{\frac{1}{n}\sum_{i=1}^{n} y_i^{1-\varepsilon} - 1}{\bar{y}^{1-\varepsilon} - 1}$	$\frac{1-n^{-\varepsilon}}{1-\bar{y}^{\varepsilon-1}}$ or 1^*	$\frac{1-\varepsilon}{n}\frac{y_i^{-\varepsilon} - y_j^{-\varepsilon}}{\bar{y}^{1-\varepsilon}-1}$
Generalised entropy	$E_\theta = \frac{1}{\theta^2-\theta}\left[\frac{1}{n}\sum_{i=1}^{n}\left[\frac{y_i}{\bar{y}}\right]^\theta - 1\right]$	$\frac{n^{\theta-1}-1}{\theta^2-\theta}$ or ∞^{**}	$\frac{y_i^{\theta-1} - y_j^{-\varepsilon}}{n\bar{y}^\theta}$
Herfindahl	$H = \frac{1}{n}\left[c^2 + 1\right]$	∞	$\frac{2}{n^2\bar{y}^2}\left[y_j - y_i\right]$
Theil	$T = \sum_{i=1}^{n} s_i \log\left(ns_i\right)$	$\log n$	$\frac{1}{n\bar{y}}\log\frac{y_j}{y_i}$

Notes: * 1 if $\varepsilon \ge 0$; ** ∞ if $\theta < 1$

Table 21. Properties of Inequality Measures (Reprint from Cowell (2005), pp. 66-67)

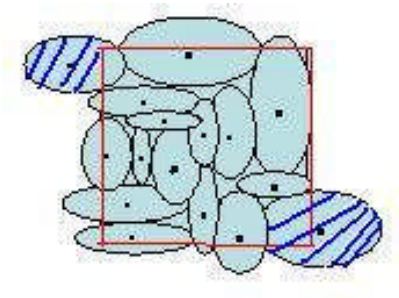
	Principle of Transfers	Distance Concept	Decomposable?	Independent of income scale & population size?	Range in interval [0,1] ?
Variance, V	strong	Absolute differences	Yes	No: increases with income	No
Coeff. of variation, c	weak	As for variance	Yes	Yes	No
Relative mean deviation, M	just fails	0, if incomes on same side of $\bar{y}$, or 1 otherwise	No	Yes	No: in [0,2]
Logarithmic variance, v	fails	Differences in (log-income)	No	Yes	No
Variance of logarithms, v_1	fails	As for logarithmic variance	No	Yes	No
Equal shares coefficient	just fails	As for relative mean deviation	No	Yes	Yes
Minimal majority	just fails	Similar to M (critical income is y_0, not $\bar{y}$)	No	Yes	Yes
Gini, G	weak	Depends on rank ordering	No	Yes	Yes
Atkinson's index, A_ε	weak	Difference in marginal social utilities	Yes	Yes	Yes
Dalton's index, D_ε	weak	As for Atkinson's index	Yes	No	No
Theil's entropy index, T	strong	Proportional differences	Yes	Yes	No
Herfindahl's index, H	strong	As for variance	Yes	No: decreases with population	Yes: but min > 0
Generalised entropy, E_θ	strong	Power function	Yes	Yes	No

Note: "just fails" means a rich-to-poor transfer may leave inequality unchanged rather than reducing it.

Appendix B. Relationship between Grid Cell and Census Blocks

The red rectangle and those ellipses represent the grid cell and Census blocks respectively. In this case, the two blocks segmented by blue lines will be excluded when BenMAP calculates the population in this grid cell.

Figure 9. Grid Cell and Census Blocks



References

Abt Associates Inc. 2000. Final Heavy Duty Engine/Diesel Fuel Rule: Air Quality Estimation, Selected Health and Welfare Benefits Methods, and Benefit Analysis Results. Prepared for U.S. EPA, Office of Air Quality Planning and Standards, Research Triangle Park, NC. Bethesda, MD. December.

Abt Associates Inc. 2003. Preliminary Nonroad Landbased Diesel Engine Rule: Air Quality Estimation, Selected Health and Welfare Benefits Methods, and Benefit Analysis Results. Prepared for U.S. EPA, Office of Air Quality Planning and Standards, Research Triangle Park, NC. Bethesda, MD. April.

Apelberg, B. J., T. J. Buckley and R. H. White. 2005. Socioeconomic and racial disparities in cancer risk from air toxics in Maryland. Environ Health Perspect. Vol. 113 (6): 693-9.

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. Vol. 85: 641-663.

Bullard, R. D. 1994. The Legacy of American Apartheid and Environmental Racism. St. John's Journal of Legal Commentary. Vol. 9: 445-474.

Cowell, F. A. 2000. Measuring Inequality, manuscript. www.lse.ac.uk.

Davidson, P. and D. L. Anderton. 2000. DEMOGRAPHICS OF DUMPING II: A NATIONAL ENVIRONMENTAL EQUITY SURVEY AND THE DISTRIBUTION OF HAZARDOUS MATERIALS HANDLERS. Demography. Vol. 37 (4): 461-466.

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.:

Gwynn, R. C., R. T. Burnett and G. D. Thurston. 2000. A time-series analysis of acidic particulate matter and daily mortality and morbidity in the Buffalo, New York, region. Environ Health Perspect. Vol. 108 (2): 125-33.

Hite, D. 2000. A random utility model of environmental quality. Growth and Change. Vol. 31: 40-58.

Ito, K. 2003. Associations of Particulate Matter Components with Daily Mortality and Morbidity in Detroit, Michigan. In: Revised Analyses of Time-Series Studies of Air Pollution and Health. Health Effects Institute. Boston, MA. May.

Krewski, D., R. Burnett, M. Goldberg, K. Hoover, J. Siemiatycki, M. Jerrett, M. Abrahamowicz and M. White. 2000. Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality. Health Effects Institute. Cambridge. July.

Lejano, R. P. and H. Iseki. 2001. Environmental Justice: Spatial Distribution of Hazardous Waste Treatment, Storage and Disposal Facilities in Los Angeles. Journal of Urban Planning and Development. Vol. 127 (2): 51-62.

Levy, J. I., S. M. Chemerynski and J. L. Tuchmann. 2006. Incorporating concepts of inequality and inequity into health benefits analysis. International Journal for Equity in Health. Vol. 5 (2):

Levy, J. I., A. M. Wilson and L. M. Zwack. 2007. Quantifying the efficiency and equity implications of power plant air pollution control strategies in the United States. Environ Health Perspect. Vol. 115 (5): 743-50.

Linder, S. H., D. Marko and K. Sexton. 2008. Cumulative cancer risk from air pollution in Houston: disparities in risk burden and social disadvantage. Environ Sci Technol. Vol. 42 (12): 4312-22.

Mantaay, J. 2002. Mapping Environmental Injustices: Pitfalls and Potential of Geographic Information Systems in Assessing Environmental Health and Equity. Environmental Health Perspectives. Vol. 110 (2): 161-171.

Mennis, J. 2002. Using geographic information systems to create and analyze statistical surfaces of population and risk for environmental justice analysis. Social Science Quarterly. Vol. 83 (1):

Moolgavkar, S. H. 2000. Air Pollution and Hospital Admissions for Chronic Obstructive Pulmonary Disease in Three Metropolitan Areas in the United States. Inhalation Toxicology. Vol. 12 (Supplement 4): 75-90.

Moolgavkar, S. H. 2003. Air Pollution and Daily Deaths and Hospital Admissions in Los Angeles and Cook Counties. In: Revised Analyses of Time-Series Studies of Air Pollution and Health. Health Effects Institute. Boston, MA. May.

Morello-Frosch, R. and B. M. Jesdale. 2006. Separate and unequal: residential segregation and estimated cancer risks associated with ambient air toxics in U.S. metropolitan areas. Environ Health Perspect. Vol. 114 (3): 386-93.

Norris, G., S. N. YoungPong, J. Q. Koenig, T. V. Larson, L. Sheppard and J. W. Stout. 1999. An association between fine particles and asthma emergency department visits for children in Seattle. Environ Health Perspect. Vol. 107 (6): 489-93.

Perlin, S. A., R. W. Setzer, J. Creason and K. Sexton. 1995. Distribution of industrial air emissions by income and race in the United States: an approach using the toxic release inventory. Environ Sci Technol. Vol. 29 (1): 69-80.

Peters, A., D. W. Dockery, J. E. Muller and M. A. Mittleman. 2001. Increased particulate air pollution and the triggering of myocardial infarction. Circulation. Vol. 103 (23): 2810-5. http://www.circulationaha.org/cgi/content/full/103/23/2810

Pope, C. A., 3rd, R. T. Burnett, M. J. Thun, E. E. Calle, D. Krewski, K. Ito and G. D. Thurston. 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. Jama. Vol. 287 (9): 1132-41.

Samet, J., S. Zeger, F. Dominici, F. Curriero, I. Coursac, D. Dockery, J. Schwartz and A. Zanobetti. 2000. The National Morbidity, Mortality, and Air Pollution Study. Health Effects Institute. Cambridge, MA. Report No. 94. May.

Schwartz, J., D. Slater, T. V. Larson, W. E. Pierson and J. Q. Koenig. 1993. Particulate air pollution and hospital emergency room visits for asthma in Seattle. Am Rev Respir Dis. Vol. 147 (4): 826-31.

Sheppard, L. 2003. Ambient Air Pollution and Nonelderly Asthma Hospital Admissions in Seattle, Washington, 1987-1994. In: Revised Analyses of Time-Series Studies of Air Pollution and Health. Health Effects Institute. Boston, MA. May.

Sheppard, L., D. Levy, G. Norris, T. V. Larson and J. Q. Koenig. 1999. Effects of ambient air pollution on nonelderly asthma hospital admissions in Seattle, Washington, 1987-1994. Epidemiology. Vol. 10 (1): 23-30.

Stretesky, P. and M. Lynch. 1999. Environmental justice and the prediction of distance to accidental chemical releases in Hillsborough county, Florida. Social Science Quarterly Vol. 80 (4):

Taquino, M., D. Parisi and D. Gill. 2002. Units of analysis and the environmental justice hypothesis: the case of industrial hog farms. Social Science Quarterly. Vol. 83 (1):

U.S. Bureau of the Census. 2002. Modified Race Data Summary File 2000 Technical Documentation. Census of Population and Housing: Washington DC. http://www.census.gov/popest/archives/files/MRSF-01-US1.pdf.

U.S. EPA. 2000. Regulatory Impact Analysis: Heavy-Duty Engine and Vehicle Standards and Highway Diesel Fuel Sulfur Control Requirements. U.S. EPA, Office of Air and Radiation. Washington, DC. EPA 420-R-00-026. December.

U.S. EPA. 2004. Final Regulatory Analysis: Control of Emissions from Nonroad Diesel Engine. U.S. EPA, Office of Transportation and Air Quality. Washington, DC. EPA 420-R-04-007. April.

U.S. EPA. 2006. Final Regulatory Impact Analysis: PM2.5 NAAQS. Office of Air and Radiation, Office of Air Quality Planning and Standards. Research Triangle Park, NC. http://www.epa.gov/ttn/ecas/ria.html.

U.S. EPA. 2008. Regulatory Impact Analysis: Control of Emissions of Air Pollution from Locomotive Engines and Marine Compression Ignition Engines Less than 30 Liters Per Cylinder. Office of Transportation and Air Quality, Assesssment and Standards Division. Washington, DC. EPA420-R-08-001. March. http://www.epa.gov/otaq/locomotv.htm.

Williams, R. W. 1999. The Contested Terrain of Environmental Justice Research: Community as Unit of Analysis. The Social Science Journal. Vol. 36 (2): 313-328.

Woodruff, T. J., J. Grillo and K. C. Schoendorf. 1997. The relationship between selected causes of postneonatal infant mortality and particulate air pollution in the United States. Environmental Health Perspectives. Vol. 105 (6): 608-612.

Woodruff, T. J., J. D. Parker and K. C. Schoendorf. 2006. Fine particulate matter (PM2.5) air pollution and selected causes of postneonatal infant mortality in California. Environmental Health Perspectives. Vol. 114: 786–790.