National Guidance for Conducting Environmental Justice Analyses

Peer Review Version

Office of Environmental Justice
US Environmental Protection Agency
October 1, 1998

Executive Summary

This document was produced in response to Executive Order 12898: Federal Actions to Address Environmental Justice in Minority and Low-income Populations. It covers: the concept of Environmental Justice (EJ) and associated terminology; available data for conducting EJ analyses; and the implementation of such an EJ analysis. It provides recommendations, practical examples, and a discussion of the available resources and tools with which to conduct an EJ analysis.

Environmental Justice (EJ), as described in this document, focuses on reducing disproportionate human health and/or environmental effects endured by communities characterized by minority and/or low-income populations. The concept of an EJ analysis is distinct from analyses done under Title VI of the 1964 Civil Rights Act. By definition EJ attempts to deal with disproportionate risk to *minority and/or low-income* communities, whereas a Title VI prohibits discrimination based on *race, color, or national origin,* and the measure of discrimination is whether or not an adverse disparate impact exists.

This document provides a working set of EJ definitions and analytical concepts as a means of standardizing EJ terminology across various platforms and facilitating communication and peer review among EJ analysts. The terms in this document are not intended to have any legal significance, but simply attempt to consistently describe the various steps and issues in EJ analyses.

The information resources discussed in this document include demographic, environmental and human health effect, and location data. There are a variety of options for obtaining these various data: through federal databases; commercial vendors; or a combination of both. Depending on the particulars of an assessment, an analyst may choose to collect data to supplement the most readily available resources. For example, data describing other sources of possible effects such as: motor vehicles; overflowing sewers; ozone or particulate levels resulting from many sources; and actions of other agencies subsequent to EPA intervention. Most importantly, an analyst should be familiar with the nature and limitations of each data type and source as it relates to the assessment at hand. This document makes recommendations concerning the use of available data for varying assessments and presents the advantages and drawbacks of alternative options.

Developing a conceptual framework, or strategic plan, is critical to structuring a practical and analytically sound EJ assessment. Issues that should be considered in structuring a conceptual framework include: the purpose and significance of the decision-making activity the EJ assessment is expected to support; the 'level of uncertainty' appropriate for

that decision-making activity; and the available resources. Based on the purpose and significance of a decision-making activity, an appropriate 'level of uncertainty' should be established, i.e., how accurate does the result have to be in order to support the nature of the decision being made. For example, if the outcome of an assessment were likely to generate a legal challenge, the level of uncertainty would need to be fairly low and thus require a fairly rigorous and defensible methodology. Finally, an analyst should understand the resource limitations for conducting the EJ assessment (e.g., time, money, data availability, required hardware and software, etc.). The process for developing a conceptual framework and practical examples for illustrating the implementation of those concepts are provided in this guidance.

The two most common types of EJ assessments are targeting/screening assessments and site specific assessments. Targeting/screening assessments are proactive analyses aimed at characterizing areas for further action/assessment. Site specific assessments are analyses performed in reaction to a known source(s) or issue(s) of potential disproportionate risk; usually involving environmental and/or human health effects, or any number of other possible effects such as economic, social, nuisance, etc..

Depending on the type of assessment and the conceptual framework, the methods for EJ assessment may vary. This document provides several qualitative and quantitative methods and recommendations for using them. Qualitative methodologies utilize the opinions of experts, including the community residents. These methods assist analysts in understanding the characteristics and specific needs of a community, including special vulnerabilities and possible cumulative and synergistic effects. Methods used to characterize geographic areas quantitatively include: absolute and relative thresholds; ranking and comparison approaches; and statistical significance testing. These varying approaches may be used to characterize a target or study area, a comparison or reference area, or the difference between the two areas. While specific methods are recommended for different types of assessments, alternatives are also presented with associated advantages and disadvantages.

In addition to these methods, this document provides a discussion of other tools employed in EJ assessments such as geographic information systems (GIS) and environmental risk exposure models. GIS allows an analyst to manage, analyze and display data spatially. EPA users have performed analyses using a variety of such tools, or "platforms," predominantly ARC/INFO, ArcView, and LandView. However, it should be noted that it is more important to use accurate and documented data and apply accepted methodologies than to use a particular GIS platform. Exposure models can be used for assessments in the absence of empirical data, as a supplement to limited data, or to mathematically describe complex exposure situations. These tools range from individual air and water models available to integrated analysis tools such as OPPT's PC-GEMS (Personal Computer - Graphical Exposure Modeling System) system. Models may be very useful,

as long as their correct application and their limitations are understood and considered with respect to the objectives of the study.

This document provides recommendations for using available resources, methods, and tools for performing EJ analyses. However, it is up to the analyst to study and understand the particulars of each assessment and proceed in a practical and methodologically appropriate manner. This guidance is considered a work in progress, and as such will change with the continual exploration of EJ concepts and methods.

Table of Contents

Acknowledgments

The United States Environmental Protection Agency's Office of Enforcement and Compliance Assurance (OECA) produced this report with editorial support from Indus Corporation, under EPA Contract Number 68-W7-0034, and Abt Associates Inc., under EPA Contract Numbers 68-W6-0022 and 68-C4-0060. Special thanks go to Dr. Gaylord; former Director of the Office of Environmental Justice who provided the inspiration and support needed to make this project a success. Also recognized are the following individuals from EPA Headquarters Program Offices and Regional Offices who participated in this project by assisting in planning efforts, writing significant portions of the text, and reviewing draft documents.

Albright, David
Region IX
San Francisco, CA

Carney, Gerald
Region VI
Dallas, TX

Forman, Debra
Region III
Philadelphia, PA

Fox, Catherine*
Region IV
Atlanta, GA

Hall, Loren*
Office of Pollution
Prevention and Toxics
Washington, DC

Halper, Marty*
Office of Environmental
Justice
Washington, DC

Hoffman, Bill
Region III
Philadelphia, PA

Levine, Ronnie
Region I
Boston, MA

Lubin, Arthur
Region V
Chicago, IL

Moralis, Susan
Region X
Seattle, WA

O'Lone, Mary
Office of General Counsel
Washington, DC

Phillips, Ginny
Office of Solid Waste
Washington, DC

Potash, Cory
Region VIII
Denver, CO

Reyes, Deldi
Region VIII
Denver, CO

Zimmerman, Rae
Region II
New York, NY

*Denotes project leads.

1. Introduction to Environmental Justice Analysis

1.1 Background

Environmental Justice (EJ) places emphasis on reducing disproportionate adverse human health and/or environmental effects endured by communities typically characterized by minority and/or low-income populations. As a social and political issue, EJ is concerned with establishing greater representation and equity in the administrative, political, and economic decisions that affect public health, welfare, and the development of communities (Goldman, 1993).

Pursuant to Executive Order (EO) 12898 (given in Appendix A) signed February 11, 1994, the Environmental Protection Agency (EPA) and other Federal Agencies are required to:

> "make achieving environmental justice part of its mission by identifying and addressing, as appropriate, disproportionately high and adverse human health or environmental effects of its programs, policies, and activities on minority populations and low-income populations."

In response to EO 12898, EPA's Office of Environmental Justice (OEJ) developed and released an agency-wide document describing its "Environmental Justice Strategy" (US EPA, 1995a). A year later an "Environmental Justice Implementation Plan" (US EPA, 1996a) was released. These two documents define EPA's goals in the pursuit of EJ. Further, the documents discuss the Agency's planning activities, policy formation, programs, and public participation processes, and possible changes in rule-makings necessary in the pursuit of EJ. Several EPA offices such as the Office of Enforcement and Compliance Assurance, the Office of Solid Waste and Emergency Response, and the Office of Federal Activities have or are developing supplementary plans and guidance to enable the successful execution of Executive Order 12898 and EPA's Environmental Justice Implementation Plan.

1.2 A Need for Guidance

One barrier that exists with respect to implementing EO 12898 is the lack of consistent agreement on how to conduct an EJ analysis. Regardless, many analysts at EPA, other Federal agencies, state agencies, universities, and other organizations have or are currently in the process of identifying and assessing EJ areas of concern (defined as an area having potential or actual disproportionate adverse effects on minority and/or low-income populations). The comparability of such studies is difficult due to the great

variety of data and methodological approaches used in their analysis process. In addition, because the purpose and nature of these applications vary so greatly, there is no single approach that can meet the individual objectives of all studies. Nevertheless, some standardization is necessary to allow for meaningful comparisons of results across the Agency.

Dr. Clarice Gaylord, former Director of EPA's Office of Environmental Justice, recognized the need for a consistent, reproducible approach for identifying and assessing instances of disproportionate adverse effects on minority and/or low-income populations. Dr. Gaylord also sought awareness of instances in which the Agency's actions may result in disproportionate adverse effects (EJ Workshop, Denver, CO, May 1997). The Enforcement Subcommittee of the National Environmental Justice Advisory Council (NEJAC) recognized similar issues and requested the development of an EJ technical guidance that would be used to prioritize activities that will ultimately reduce disproportionately adverse effects on minority and/or low-income populations (Enforcement Resolution No. 3, NEJAC Meeting, Baltimore, MD, 1996).

Established in 1995, the U.S. Environmental Protection Agency (EPA) Office of Enforcement and Compliance Assurance (OECA) Risk-Based Targeting (RBT) Work Group met monthly to exchange ideas and data related to targeting of sensitive populations and ecosystems for a variety of purposes. Dr. Gaylord tasked the RBT Work Group in 1997 with reviewing EPA's approaches to targeting facilities which may disproportionately affect minority and/or low-income communities and to develop written guidance that includes recommendations concerning appropriate applications, technical methodologies, and procedures to be used in the conduct of EJ analysis. Membership of the RBT Work Group included Agency experts in the following fields: ecological and human health risk assessment; Geographic Information Systems (GIS); analytical techniques; and experts in the field of EJ itself.

In order to respond to Dr. Gaylord's request and find consensus on EJ issues, the RBT Work Group held a series of meetings within EPA that included presentations on risk-based targeting and EJ identification initiatives given and critiqued by both Headquarters and Regional representatives. The group also discussed and prepared background material on various data types and methodologies related to EJ analysis. In addition, the group put forth default recommendations on procedures to be used in the conduct of EJ analysis. This document represents the Work Group's response to Dr. Gaylord's request.

1.3 Purpose and Use of this Document

The purpose of this document is to serve as guidance to analysts so that their EJ assessments may be: consistent and comparable across EJ study areas; conducted in a

non-arbitrary manner; and conceptually and methodologically defensible when presented to others.

This guidance is written for a wide audience interested in the assessment of environmental justice. This includes analysts within EPA and other Federal agencies; state and local environmental authorities; academics; consultants; and non-government/non-profit organizations such as community or business groups. This guidance should be used where possible to address many of the current problems and issues raised by Federal and state agencies in their attempts to conduct EJ analyses, although there is no formal requirement to use it. The concepts and recommendations presented focus on how EPA currently identifies EJ areas of concern for a variety of purposes, including: broad-scale area characterization; identification of geographic areas qualifying for assistance (e.g. grants); targeting/priority setting for geographic areas or facilities (e.g. inspections); regulatory effect analyses; review of permitting/regulatory practices; and new permit analyses/siting decisions.

Some of the concepts and recommendations in this document may also be applicable to conducting Title VI analyses to determine disparate impact. However, it is important to note that the terms "environmental justice" and "Title VI" as well as corresponding analyses cannot be used interchangeably. The reference to "Title VI" comes from the section of the Civil Rights Act of 1964, that prohibits discrimination based on *race, color, or national origin* (emphasis added) by recipients of Federal financial assistance. The "Environmental Justice" (EJ) concept is defined by Executive Order 12898 which states that: "each Federal agency shall make achieving environmental justice part of its mission by identifying and addressing, as appropriate, disproportionately high and adverse human health and environmental effects of its programs, policies, and activities on *minority populations and low-income populations...*" The distinction between Title VI and EJ is important to those doing EJ related assessments, because it may change the nature of the data used in the assessment as well as the type of assessment done. If a Title VI assessment is being done, then the demographic data used must focus on race, color, or national origin. If a more general EJ assessment is being done, the demographic data should assess both race and/or income information. This document includes guidance for the latter; it is important to note that the agency has promulgated separate guidance for assessing Title VI cases.

As presented thus far, section one of this document provides an introduction to EJ analysis as well as the purpose and use of this document. Section two, *Key Concepts and Terminology*, defines important terminology used in EJ analysis. Section three, *A Conceptual Framework for Conducting EJ Analyses*, provides a context for conducting EJ analyses in environmental decision-making. Section four, *Information Resources for Conducting EJ Analyses*, provides information on demographics, environmental, and human health data sources, as well as existing data gaps and data quality issues. The next section, *Methodologies for Conducting EJ Analyses*, recommends suitable methodologies

and sets research and system development priorities which will in due time improve the overall quality of EJ assessments. Finally section six, *Recommendations and Conclusions*, presents a discussion of the RBT Workgroup's recommendations on the EJ assessment process. Appendices A through E provide supporting information to this study including: EO 12898, a list of EPA Regional and Program Office EJ coordinators, a glossary of terms, an EJ bibliography, and EPA database information.

Many of the concepts, example applications, and methodologies presented here are derived from the ongoing efforts of the RBT Work Group, from the draft document "NEPA Environmental Justice Guidance" (US EPA, 1996b), and from the draft document "Interim Region II Policy on Identification of Environmental Justice Areas" (US EPA, 1996c).

2. Key Concepts and Terminology

A consistent set of terms are provided below to clarify understanding of concepts presented in this report and to encourage use in Agency-wide EJ assessments conducted in the future. The definitions are not intended to carry legal significance, but simply to be a useful way to consistently describe the issues and ideas pertinent to EJ analyses. Additional terminology is presented in a Glossary (Appendix B). Additional information on these concepts, e.g., reference area and disproportionate effect, may be found in later sections of this document.

2.1 Target Area

A *target area* (study area) is a geographical area that is potentially affected by an action falling under EO 12898. A target area is usually proximate to and may surround a source(s) of potential adverse environmental and/or human health effects, often including, but not restricted to, a polluting facility. Other sources of possible effects include motor vehicles, overflowing sewers, ozone or particulate levels resulting from many sources, and actions of other agencies subsequent to EPA intervention.

2.2 Target Population

The *target population* includes the potentially affected residents of the target area. Depending upon the objective and context of the analysis, the target population may also include transient residents such as migrant workers, commuters, and seasonal visitors. A target population may constitute an entire population or a subset within the population (children, or low-income fishermen, for example). Exposure of the target population to an environmental hazard, may be the result of a source(s) *within* the target area or a source *external to* the target area (consumers of a contaminated drinking water or fish, for example).

2.3 Reference Area/Reference Population

A *reference area* is the area that is used as a benchmark of comparison when determining whether a *target area* suffers from a disproportionate effects to its minority and/or low-income populations. A reference population includes the residents of the reference area. Therefore, both the reference area and population provide a context for the interpretation of data from the target area and population.

2.4 *Disproportionate Effect*

A *disproportionate effect* is an incidence (or prevalence) of an effect, a risk of an effect, or likely exposure to environmental hazards potentially causing such adverse health effects on a minority and/or low-income population, or sub-population such as children, that significantly exceeds that experienced by a comparable reference population. In estimating effects, the possible cumulative nature of these effects should be considered. For example, if the effect of concern is the level of truck traffic from a proposed facility, the fact that the local streets are already carrying the traffic from several similar facilities could be significant. The concept of disproportionate effect on sub-populations (such as high numbers of minority children) is discussed later in this document.

2.5 *EJ Area of Concern*

An *EJ area of concern* is a target area that has been demonstrated to experience disproportionate effects and has a significant minority and/or low-income population relative to an appropriate reference area. A *potential EJ area of concern* is a target area that contains a significant minority and/or low-income population but the existence of disproportionate effects has not yet been shown. Depending on the objective of the study, further analysis to determine which, if any, potential EJ areas of concern are actual EJ areas of concern, may not be necessary (e.g. broad scale area characterization).

2.6 *Population Areal Unit*

A *population areal unit* is a geographic unit containing populations which in aggregate are used to define a target area. The sum of the populations of these units comprise the target population. Examples of areal units include Census blocks, block groups, tracts and sometimes zip codes or counties. In some analyses a single population areal unit is assumed to define the target area.

3. A Conceptual Framework for Conducting EJ Analyses

3.1 Objective

The purpose of this section is to present a clear and concise overview of the procedural issues associated with conducting an Environmental Justice (EJ) assessment. It is assumed that the reader is familiar with topics frequently encountered in the consideration of environmental issues and has a working knowledge of the statutory and legal conditions that prevail. In addition, experience with the tools employed in the assembly, display, and analysis of demographic and environmental data, ecosystem processes, and related effects on human health will also be an asset to understanding and applying the information presented in this document.

Recognizing the variety of possible applications of EJ analysis, this guidance does not recommend any single approach but explores the types and nature of approaches one may adopt in different applications, while working within a common conceptual framework. This conceptual framework provides an EJ analyst with some procedural guidance through the identification and evaluation of the following steps:

1. Establish the purpose and significance of the decision-making activity the EJ assessment is expected to support;

2. Determine the 'level of uncertainty' appropriate for that decision-making activity, i.e., how accurate does the result have to be in order to support the nature of the decision being made; and

3. Identify resource limitations for conducting the EJ assessment, e.g., time, money, data availability, required hardware and software, etc..

Establishing the purpose and significance of a decision-making activity involves identifying contextual issues associated with the consequences of the decision. Possible contextual issues to be considered include: the magnitude of the potential adverse effects and the number of people potentially affected; the potential for legal challenge or in the context of a court case; the level of public awareness/media coverage; the direct and indirect economic factors of investment and development issues (e.g., land or property value, cost of the project, new jobs generated, etc.); and the scope of the project (e.g., national, state, county, etc.). It may be appropriate to loosely quantify the significance of the decision-making activity for comparative purposes, e.g., ranking the significance of different projects in relation to each other.

Based on the significance of the decision-making activity, a 'level of uncertainty' may be determined for the related EJ assessment. The level of uncertainty concept refers to the confidence an analyst has in the accuracy of the assessment leading to the eventual decision. A level of uncertainty is dependent on factors influencing the validity of that assessment, such as the accuracy of data, the rigorousness of the methodology, or the level of effort. For example, if the decision-making activity were likely to generate considerable public interest, the assessment methods that lead to the decision would likely be subject to rigorous scrutiny. Under those circumstances, a relatively low level of uncertainty (i.e., high level of accuracy) would be appropriate. As such, the data *and* methodologies used in the assessment would need to be both accurate and defensible. On the other hand, if the decision-making activity was an internal initial review of a broad geographic area, the level of uncertainty in the result could be higher since the analysis would be supporting the development of general working knowledge of possible EJ issues. In this case, the precision and accuracy of the data and methodologies are not likely to be critical to the outcome.

Once the level of uncertainty is determined, it is necessary to assess the feasibility and options for conducting the EJ analysis, given the available resources. It is at this point that an analyst may decide to plan and implement the assessment requested, modify the scope of the assessment, or not move forward at all. This decision is entirely dependent on the resource limitations involved and the allowable level of uncertainty. For example, if the assessment requires a low level of uncertainty, but the time allowed to complete the assessment is not sufficient for a rigorous in-depth analysis, the analyst may not be able to comply with that request within the specified constraints.

3.2 Scenarios of Decision-Making Supported by EJ Assessment

The following two possible scenarios are briefly presented below to illustrate how an EJ assessment could be used in environmental decision-making. These scenarios are discussed in later text to help illustrate important EJ analytical concepts and recommendations. They are presented in order of greatest to least allowable levels of uncertainty.

3.2.1 Development of Descriptive Characterizations of a Geographic Area

In this scenario, the objective is to identify potential EJ areas of concern within a major geographic area (a state, a region, or the entire U.S. for example). The intent is to form a summary-level working knowledge of a given geographic area (i.e., what's the general nature of the area; population characteristics; environmental status and condition in the area; major roads; other possible sources of effects). This information may be used to characterize the communities surrounding facilities or a specific industrial sector, e.g., a

pulp and paper mill. Sometimes this may entail summarizing environmental information together with population data on a broad geographic scale, e.g., TRI air releases by county along with income data. This approach may also be used to identify potential EJ areas of concern within a given sector of facilities, e.g., federal facilities. This type of analysis is essentially descriptive and therefore does not require a low level of uncertainty. As such, it involves the use of few, if any, demanding quantitative methodologies and could be performed with minimal lead time.

3.2.2 Permitting and Facility Siting Process (non-delegated)

EPA is in charge of reviewing permit applications for environmental programs that have not been delegated to their respective state, such as the RCRA permits in the State of Indiana. However, this situation is relatively rare since the majority of environmental permits are issued by states through delegated authorities. At the conclusion of such a review, the EPA may render a decision to approve, approve with a required modification, or disapprove a permit application. Alternatively, sometimes EPA will oversee the permit review process for a particular state if the community residing proximate to the proposed site has expressed concerns about EJ issues.

If the objective of the assessment is to support a permit review, it is important to first figure out who may be affected by the permitted activity, i.e., are there any minority or low-income communities affected. This information is critical to developing a communication strategy, since the nature of the community may impact how to best reach them. Once the communities are identified, issues to be considered would include language barriers and a mechanism for getting information out (e.g., radio, TV, publications, public meetings, etc.). Once the community in question has been adequately informed about the proposed site and does not voice any concerns, then an EJ analysis should not be needed. However, if the community *is* concerned about EJ issues, then an EJ assessment should be conducted. Accordingly, the nature of that assessment should be responsive to the concerns raised by the communities during the comment period and public meeting process. The necessary degree of certainty/uncertainty in this case supports the decision to approve or disapprove the permit. Depending on the type of assessment called for, a fairly low level of uncertainty would generally be required, based on community concern and the possible denial or revocation of a permit and the probability that the decision will be further challenged.

4. Information Resources for Conducting EJ Analyses

4.1 Introduction

Implementing Executive Order 12898 requires the EJ analyst to identify disproportionately high and adverse human health and environmental effects on minority populations and/or low-income populations. The availability and characteristics of the information resources necessary to fulfill this requirement is discussed in this section.

Information on receptor population characteristics and environmental risks within the target and reference areas are key to the conduct of a successful EJ study. Access to current, accurate, and comprehensive data enables the analyst to study and fully describe the characteristics of, and potential adverse effects faced by populations within the target and reference areas. Collectively, the gathering and preparation of data will consume significant time and financial resources. Analysts may at times feel compelled to use data that are less than optimal for the application at hand, but should always use techniques and data that are appropriate for the ultimate use of the results.

4.2 Demographic Data

The term "demographic" describes information pertaining to characteristics of a population such as its size, poverty status, density, geographic distribution, and vital statistics (birth, deaths and marriages), and usually also refers to family, income and housing characteristics. Collection of demographic data is undertaken by both public agencies and commercial firms. The Bureau of the Census conducts decennial census of the US population in order to determine its demographic characteristics and publishes its findings for congressional redistricting. While the main census is conducted every ten years, there are many intermediate data collection efforts, as well as the use of forecasting techniques to provide demographic data on time periods intermediate to the decennial census. Commercial firms usually provide more geographically detailed estimates (e.g., to the "block group" level) between the decennial census than the Census Bureau itself.

4.2.1 Data Available from the Bureau of the Census

Where acceptable, EPA researchers should choose to use data provided by the Bureau of the Census. The Census provides low cost, consistent, nationwide data coverage. Their data collection and aggregation methodology are well documented. The Census conducts surveys and other auxiliary data collection activities on a continuous basis. They provide

periodic updates to subsets of the total data structures that they make available after the decennial survey. However, these updates are rarely more specific than the county geographic level.

The Census provides information on a wide range of demographic, economic, and ethnic variables. The data collected by the Bureau are aggregated to several geographic boundary-defined areas. From the smallest area to the largest area covered these enumeration units are:

- Block
- Block Group
- Tract/Block Numbering Area
- Minor Civil Division
- County
- State
- Region
- Country (National)

The number of persons recorded within each enumeration unit type is designed to be similar. Block records contain information on approximately 50-200 people; block groups: 800-1,000 people; and tracts: 3,500-4,500 people. There are about six million block areas, about 300,000 block group areas, and about 60,000 tract areas.

Other summary levels available from the Bureau of the Census include:

- ZIP code areas;
- Urbanized Areas;
- Places (incorporated, non-incorporated, or Census-designated place types);
- County groups which form Metropolitan Areas;
- Congressional Districts;
- Tribal Lands.

The US Postal Service defines about 30,000 five digit ZIP code delivery zones. The five digit ZIP code data aggregates released for the 1990 Census were not from household level data like other Census geographic areas, but were derived by assigning 1990 block group data to 1992 vintage ZIP boundary information with some mismatch of data vintage for the two files. The land area covered by five digit ZIP code areas varies widely from a few square miles in urban areas to thousands of square miles in rural areas. It is important to note that ZIP codes change in their designated number and in their location over time depending on post office jurisdictional requirements. Since it is generally accepted that the use of the smallest available enumeration unit yields the most accurate

population characterization results, zip code data should be avoided since the zip code areas are subject to change, relatively large and more heterogeneous than tracts.

Census data are released in the form of printed reports and as digital files that may be obtained in the form of downloads via the Internet; or on CD-ROM, diskette, and tape. The 1990 Census data are primarily released as Summary Tape Files (STF). Data products are usually designated by number and summary levels, with all STF1 products (A - D) referring to the 100% count data (block level), while STF3 products include original sample data (block group level). 100 percent count means that "all" individuals responded to these questions while sample data indicate 1 in 6 individuals responded. Often the data files will contain more detailed data breakdowns than the printed reports. A series of additional products focus on particular sub-populations, such as members of Indian tribes, the elderly, and the black population. These products are designated as Special Summary Tape Files (SSTF).

Census data are available from sources within EPA. For example, the Office of Information Resources Management (OIRM) has created complete national ARC/INFO and ArcView GIS compatible coverages for the Public Law (PL) 94-171 (block level) data and the STF 3A (block group level) data and included them as part of EPA's Spatial Data Library System (ESDLS). The ESDLS is currently available from EPA servers and data are/will be made available via the Internet. As part of its Environmental Justice Spatial Analysis Tools project, the Office of Pollution Prevention and Toxics (OPPT) has selected the PL 94-171, a subset of the existing OIRM STF 3A files, as well as subsets of the STF 3B and STF 1A files and prepared them as ARC/INFO GIS format coverages at the block, block group, tract and county levels. The coverages include a number of derived variables, such as percent below poverty and percent children under five years old below poverty. There are plans to press these data onto CD-ROM for distribution in the future, but this has not been implemented in a central location or within ESDLS (Hall, 1997).

4.2.2 Commercial Data Vendors

In addition to the Census, a number of corporations generate and provide data on a commercial basis due to the data's utility to firms that undertake marketing activities. Firms such as Claritas and Urban Decision Systems (UDS) provide data that are updated on an annual basis for summary levels such as block groups, census tracts, ZIP codes, counties, and states. The estimation methods such firms use involve calibration to Bureau of the Census survey results and employ the use of projection techniques. Such data may be licensed either to individual users or to larger groups within the Agency.

The Office of Water (OW) and the OPPT acquired UDS data for the distribution of population by race and by income level. They are available in block group or county

enumeration units for the census year 1980 reformatted to 1990 census geography, 1990, 1996, and as estimates for 2001. The license permits unlimited use *within* EPA, but does not allow for distribution *outside* the Agency. Data products *derived* from these data, such as OW's population counts and densities for watershed areas, are not limited by the license agreement. The UDS data are available on CD-ROM in a common database format and portions of the data have been converted to ARC/INFO coverages. For example, total population data derived for watersheds are available in ARC/INFO format files and is available via the Internet. The UDS data are available from the EPA FTP site at: public/data/R02earth/cendata.

4.2.3 Minority and Low-income Data for an EJ Analysis

4.2.3.1 Ethnicity Data - A Definition

The term "ethnicity" generally describes information pertaining to a population's *religious*, *national*, *racial*, or *cultural* composition. While we all understand the concept of a population adhering to a *religion* or forming a *national* entity, the meaning of the other terms merits clarification. A *race* is a population group distinct from other groups distinguishable by one or more genetically transmitted physical characteristics such as skin color. A *culture* is distinguishable by the totality of its socially transmitted behavior patterns, arts, beliefs, institutions (including political), and other like products of a community's activity. It is important to note that this social definition of ethnicity is different from the EJ specific definition (presented in the next section) which distinguishes race from ethnicity.

4.2.3.2 Minority Data - A Definition and Discussion of Data Aggregation

Executive Order 12898 instructs the EJ analyst to examine environmental effects on "minority" populations. The definition of *minority* is a population group that is smaller in number when compared to the *majority* population group that resides within an population areal unit. A *minority group* is defined by a composite of, or an element of, its ethnic (racial, religious, national, and cultural) characteristics.

The racial classification used by the Bureau of the Census adheres to the guidelines in Federal Statistics Directive No. 15, issued by Office of Management and Budget (OMB) in 1977 and recently revised. The directive provides classification definitions to be used by all Federal agencies for ethnic and racial statistic reporting. The directive identifies *race* categories as conceptually separate from *ethnic* categories, such as Hispanic. It defines two possible methods for collecting and reporting Federal statistical data, with the preferred approach being to collect race and ethnicity data separately. In the application of the first method, the four racial categories defined in Directive 15 include:

- American Indian or Alaskan Native;
- Asian or Pacific Islander;
- Black; and
- White.

Ethnicity is defined as either:

- Of Hispanic Origin;
- Not of Hispanic Origin.

Directive 15 states that if race and ethnicity data are collected in combination, the categories should be:

- American Indian or Alaskan Native;
- Asian or Pacific Islander;
- Black Persons Not of Hispanic Origin; and
- Persons of Hispanic Origin and White Persons Not of Hispanic Origin.

The 1990 Census either directly reports such information or gives information by which to derive these categories.

The 1980 and 1990 Census reports include an additional racial category for *Other*. The five race categories used in the Census are also reflected in the Hispanic ethnicity question, so that users can distinguish between Hispanics and non-Hispanics within each race category. The Census *Other Race* category is not explicitly included in Statistical Directive 15, or in the draft definitions of CEQ and the IWG. Although the vast majority of reporters to the *Other Race* category also report Hispanic ethnicity, the differences can be significant in some areas.

Since the Hispanic ethnicity category includes members of each race class, simply adding those reporting Hispanic ethnicity to the race categories may double-count substantial numbers of persons. For example, in the 1990 Census data for the Los Angeles metropolitan area, a significant number of *Hispanics* were reported in the *Asian or Pacific Islander* (44,000) and *American Indian* (25,000) race categories which would be double-counted. In addition, about 30,000 *Other Race* respondents who did not claim Hispanic ethnicity would not be counted as minorities in the draft definitions.

Since publication in 1977, Directive 15 has been the target of criticism, mostly with respect to the following:

(1) The distinction between race and ethnicity;

(2) The lack of an Other and/or Multi-Racial category; and

(3) The ambiguity of classifications based on combining factors (such as ancestry and geographic origin) which may omit consideration of groups such indigenous peoples of Central and South America, for example.

The directive was recently revised, and final changes will be adopted for use in the collection of Census 2000 data.[1]

4.2.3.3 Current Options for Using Minority Data

Several guidance documents that define data collection methodologies regarding racial classification have been developed for the purpose of EJ identification.

- EPA's Office of Environmental Justice defines a minority population to include: Hispanic, Asian-Americans and Pacific Islanders, African-Americans, American Indians, and Alaskan Natives.

- The Council on Environmental Quality (CEQ) Environmental Justice Guidance Under NEPA (Dec. 10, 1997) uses the following classification (which parallels Directive 15): American Indian or Alaskan Native; Asian or Pacific Islander; Black; not of Hispanic origin; or Hispanic.

- Most EPA and other geographic analyses (e.g., Goldman and Fitton, 1994) have been based on aggregations of the Census classifications. These refer to the total of Black, American Indian/Alaska Native, Asian or Pacific Islander, Other race, and White Hispanics as minority populations.

These three alternatives are by far the most frequently used race classification methodologies.

4.2.3.4 Recommendations for Defining Minority Status

An inclusive definition of "minority" is most appropriate and common practice in geographically based EJ studies. This method incorporates the Census Other race category as a portion but does not double-count Hispanics. This can be conceptualized as the sum of the populations identified as members of *American Indian or Alaskan Native, Asian or Pacific Islander, Black*, and *Other races* together

---

1 OMB adopted two modifications. The Asian or Pacific Islander category will be separated into two categories. These are: Asian and Native Hawaiian or Other Pacific Islander. The term Hispanic will be changed to Hispanic or Latino.

with *Hispanics* with a listed race of *White*. This is equivalent to an alternate calculation of total population minus *White non-Hispanics*.

4.2.3.5 Income - A Definition and Discussion of Data Aggregation

The term *income* describes the amount of money received in exchange for labor, services, sales of goods or property, or as profit from investment, whereas the term *wealth* describes the sum of all goods and financial resources and possessions having economic value. A common way of reporting income is to sum the incomes of all persons living in the same family or household. However, another way to report income is as a per capita (per person) value averaged within an administrative unit. Personal and household wealth statistics are generally available for a wide variety of aggregation levels. Income and wealth data collection is undertaken by both public entities and private firms.

Information concerning residential income reported in the Census of 1990 reflects the economic conditions of respondents in 1989, reported in 1989 dollars. Under privacy protection laws, specific persons' or households' responses may not be disclosed. Although this information is not released at the household level (except in micro data samples), it is tabulated and summarized for several geographic enumeration units, such as household income levels within block groups.

The Bureau reports several categories and sub-categories of income, either by source, or by social group. The STF3 data product includes tabulations noting wage or salary income, income from self-employment, farm income, Social Security and public assistance income, as well as limited information on cash benefits. Several key sources are *not* included in income calculations, such as "in kind" income from food stamps, public housing subsidies, medical care, or lump sum figures such as tax refunds. Income data may be tabulated by family (a group of related persons living together, with possibly multiple families per household) or by household (all persons living in the same housing unit). Consider Figure 1 which illustrates how the Bureau of the Census tabulates (aggregates) data in pre-defined income ranges (categories). Thus household counts are available for income ranges such as $5,000-$9,999 or $50,000-$74,999.

Figure 1: Income Ranges Reported in STF3A	
Less than $5,000 **$5,000 to $9,999** **$10,000 to $12,499 ($10,000 to $14,999)** **$12,500 to $14,999** **$15,000 to $17,499 ($15,000 to $24,999)** **$17,500 to $19,999** **$20,000 to $22,499** **$22,500 to $24,999** **$25,000 to $27,499 ($25,000 to $34,999)** **$27,500 to $29,999** **$30,000 to $32,499** **$32,500 to $34,999** **$35,000 to $37,499 ($35,000 to $49,999)**	**$37,500 to $39,999** **$40,000 to $42,499** **$42,500 to $44,999** **$45,000 to $47,499** **$47,500 to $49,999** **$50,000 to $54,999 ($50,000 to $74,999)** **$55,000 to $59,999** **$60,000 to $74,999** **$75,000 to $99,999 ($75,000 to $99,999)** **$100,000 to $124,999 ($100,000 or more)** **$125,000 to $149,999** **$150,000 or more**
Bold values denote income ranges that are also available for tables tabulated by householder age, by race and Hispanic ethnicity.	

Data are tabulated as counts on each of the variables on the census questionnaire form (total population, for example). The value of some variables (per capita income, for example) are derived by calculation. Many cross-tabulations of variables (such as per capita income by the five major racial categories) are also produced. The distribution of income data is presented as numbers of respondents (counts) in a particular income range. For example, a Block Group might be estimated to include 20 households reporting total income between $12,500 and $15,000. Tabulations of "Household Income by Race of Householder", by "Hispanic Origin of Householder", and by "Age of Householder" are published by nine income categories (see Figure 1 above). In addition to these categories, family income data are also available by family type; and by the presence and age of children (e.g., "Female Householder, No Husband Present: With Own Children under 18 years").

Mean household income is obtained by dividing the summed (the total) household income by the number of households. Median income is obtained by dividing the distribution of the variable into two equal parts, one having incomes above the median (or middle) value and the other having incomes below the median. For households and families, the median income is based on the distribution of the total number of units (including those with no income). The Bureau of the Census STF 3A technical documentation states that of these two measures, median values are generally preferred for evaluation of incomes in small geographic areas because a few very high or low values may substantially affect the mean. Per capita income is computed by dividing the aggregated/summed income for an area by the number of residents for which income data were available.

4.2.3.6 Poverty - A Definition and Discussion of Data Availability

Poverty statistics presented in Bureau of Census publications are based on a definition originated in the 1960's and identified by the OMB as a statistical standard for Federal

agencies.[2] Income level cutoffs used by the Census to determine the poverty status of families and unrelated individuals living together are given as a set of thresholds arranged in a cross tabulation consisting of family size (from one person to nine or more persons) by presence and number of family members under 18 years old (from no children present to eight or more children present). This cross tabulation is presented in Figure 2 below.

The number of persons below the poverty level is defined as the sum of the number of persons in families with incomes below the poverty level and the number of unrelated individuals with incomes below the poverty level. The poverty thresholds are revised annually to allow for changes in the cost of living as reflected in the Consumer Price Index. For example, the average poverty threshold for a family of four persons was $12,674 in 1989 (See Figure 2 below). Poverty thresholds are applied on a national basis and are not adjusted for regional, state or local variations in the cost of living.

Figure 2: Poverty Thresholds in 1989 by Family Size and Number of Related Children Under 18 Years

Size of Family Unit/Age of Head of Household	Weight Average Threshold In '000's Dollars	Number of Related Children Under 18 Years								
		0	1	2	3	4	5	6	7	8
1 Person	6.3									
< 65 years	6.5	6.5								
> 65 years	5.9	5.9								
2 Persons	8.1									
< 65 years	8.3	8.3	8.5							
> 65 years	7.5	7.5	8.5							
3 Persons	9.9	9.7	10.0	10.0						
4 Persons	12.7	12.8	13.0	12.6	12.6					
5 Persons	15.0	15.4	15.6	15.2	14.8	14.6				
6 Persons	16.9	17.8	17.8	17.4	17.1	16.6	16.3			
7 Persons	19.2	20.4	20.5	20.1	19.8	19.2	18.6	17.8		
8 Persons	21.3	22.8	23.0	22.7	22.2	21.7	21.1	20.4	20.2	
=,> 9 persons	25.5	27.4	27.6	27.2	26.9	26.4	25.7	25.1	24.9	24.0

2 At the core of this definition was the 1961 Economy Food Plan. It was determined from an Agriculture Department's 1955 survey of food consumption that families of three or more persons spend approximately one-third of their income on food; hence, the poverty level for these families was set at three times the cost of the Economy Food Plan. For smaller families and persons living alone, the cost of the economy food plan was multiplied by factors that were slightly higher to compensate for the relatively larger fixed expenses of smaller households.

4.2.3.7 Current Options for Using Income Data

Figure 3 summarizes the income and poverty data elements and availability by EPA data source or system. Other data summaries may also be available by Region.

Figure 3: Income Data Elements and Availability by EPA Data Source or System

EPA Source / System	ESDLS / STF3A	LandView III	OPPT EJ Tools	Census Updates Yearly Small Area		Commercial (UDS)
Years	1990	1990	1990	1991, 1993 - 1995		1991 - 1996
Geographic Summary Levels	Block Group	Block Group	Block Group, Tract, County	State, County		Block Group[1], Tract, County[1], State
Income Range[2]						
– Household	Yes	No	Yes	Yes	No	Yes[1]
– Family	Yes	No	No	No	No	Yes
Statistics						
-- Median	Yes	Yes[3]	Yes[3]	Yes	Yes	Yes[1]
-- Mean	Yes	No	Yes[3]	No	No	Yes[1]
– Per Capita	Yes	No	Yes	No	Yes	
Poverty Status	Yes	No	Yes	Yes	Yes[4]	No
Income/ Poverty Ratio	Yes	No	Yes	No	No	No

Notes for Figure 3:

1. Data for these summary levels have been acquired by OPPT from Urban Decision Systems but not processed into a proprietary GIS format.

2. Different sources or portions of sources may use different income ranges. The STF3A/ESDLS data includes some tables with 25 income ranges, and others with nine. The OPPT tools product includes five classes for all tables, while the Urban Decision Systems products include 15 classes.

3. Household income only.

4. With yearly adjustments in the poverty cutoff values.

There is little income-specific information available at or below the county level between the decennial Censuses. This is due to the small sample sizes of the intermediate surveys. While a limited amount of housing cost and value information (a surrogate for income and wealth) is available from the short Census form, most income-related data are based on the 1 in 6 sample long Census form. This limits the geographic level of specificity (to block group and higher summary levels) and complicates comparison with population

count information for which the short form data are used to derive the more accurate 100 percent counts.

4.2.3.8 Census Year to Year Adjustments and Updates

As part of the Current Population Survey (CPS) the Bureau of the Census releases annual estimates of the number of persons below poverty by state, using a poverty threshold updated by the value of the Consumer Price Index. The Census released county-level income and poverty statistics for 1993 in 1997. These estimates provide Median Household Incomes, Per Capita Income, and Number of Persons below Poverty in four age classes. Inter-census estimates of persons in poverty cannot be calculated except by the Bureau, since the process requires knowing the family size and composition for individual households.

The lack of official estimates for inter-census population and income statistics for the smallest geographic areas such as block groups and tracts is a serious impediment to assessing current income status in EJ studies. While income data are available from commercial sources, their use in studies with potential legal implications requires that the assessor be confident of the quality of the estimates and the sampling and/or forecasting methods used in their creation. In addition, income data reported for 1990 (or any other year other than that in which the analysis is undertaken) needs to be converted to account for the current value of the dollar. Adjustment values are published in many general reference almanacs.

4.2.3.9 Geographic Adjustments for Income and Cost of Living

Comparing income levels adjusted for buying power across the U.S. is not straightforward, since the total cost of living data are not aggregated and available at the smaller summary levels such as block group or tract. The Bureau compiles data for housing value (median and distribution), mortgage and rental costs (including median gross rent, distribution of gross rent) in a range of choices, depending on whether one wishes to include consideration of non-cash benefits such as included meals, mortgage, etc. These data usually do not include non-housing costs such as food, transportation, etc, making choice of a single cost-of-living comparison variable difficult.

HUD computes the HUD-Adjusted Median Family Income (HAMFI) values annually for metropolitan areas and counties within states. Several housing assistance programs define "low-income" households as being below 80% of these values, and "very low-income" as being below 50%, and makes such families eligible for housing assistance. However, these values are predominantly calculated specifically for this purpose (determining applicant eligibility, based on submission of proof of a qualifying income) rather than using them to assess the geographic distribution of low-income residents. The HUD "low-income" thresholds for 1996 generally range between about $27,000 and

$42,000, while the "very low-income" values range between about $16,000 and $27,000. Income information available between the decennial Censuses takes the form of either distribution tables with predefined ranges (e.g. 10,000 - 15,000) or as a mean or median statistic. Unless the income decision criterion involves determining whether 50% of the population lies below a threshold (e.g. the HUD "low-income", the poverty threshold, or some multiple), it would not be possible to determine whether an area met the criterion.

HUD has published a set of "qualified Census tracts" (QCTs) in the U.S. for the Low-income Housing Tax Credit program (FRxxx) which is based on low-income thresholds for 1994 and deflated to 1989 dollars for use with the 1990 Census data. HUD defines QCT's (including Census Block Numbering Areas) as those in which at least 50% of the households have an income less than 60% of the Area Median Gross Income (AMGI), which is further defined to be equal to 120% of HUD's Very Low-income Limits. The calculation procedure involves computing the tract's average household size, adjusting the income standard to this size, and computing the number of households with incomes below this standard. Tracts with 50% or more of the households below the limit may qualify, subject to a restriction on total population.

There is a restriction on the number of tracts which are determined to be QCT's within a single metropolitan statistical area. The total population of the QCT's must not exceed 20% of the total MSA population. Thus, some metropolitan areas may include tracts which meet the income criteria but which were excluded from HUD's published list because tracts with a higher percentage of eligible households sum to 20% of the population.

4.2.3.10 Recommendations for Defining Low-Income Status

In general, the term "low- income" is described in comparison to some benchmark. For example, both the NEPA Guidance document (draft, July 1997) and the April 1997 draft CEQ Guidance recommend that: "Low-income populations in an affected area should be identified with the annual statistical poverty thresholds from the Bureau of the Census' Current Population Survey Reports, Series P-60 on Income and Poverty." In other words, these documents recommend using poverty thresholds as a benchmark for low-income. While this benchmark may be appropriate for some EJ assessments, depending on the associated levels of uncertainty and data attribute requirements, it may not be feasible for others.

This document recommends a two-tiered approach which includes several practical options for defining the concepts of both low-income and very low-income populations. Figure 4 below presents these recommended benchmarks along with their associated data sources.

Figure 4: Benchmarks for Defining Low-Income Populations			
Data Source	**Poverty Thresholds**	**Income Ranges in STF3A**	**Qualified Census Tracts**
Low-Income	Below two times the poverty threshold	Below $25,000	Below 60% of AMGI values (imbedded)
Very Low-Income	Below the poverty threshold	Below $15,000	Not Available

These definitions should be used at the analyst's discretion, given the particulars of the EJ assessment being conducted and the attributes of the data as described in this document. Some of the advantages and drawbacks to using the different benchmarks are described below.

An advantage of using the poverty thresholds as benchmarks for low-income status is that the associated data adhere to a Federal statistical standard. The data are available in a variety of geographic levels; block group, tract, MCD, county, MSA, place, state, Census region, U.S., zip code, and tribal land. In addition, the data are available in a wide range of cross tabulations, such as race and age, and will facilitate some types of assessments, e.g., young children below poverty as an indicator of potentially high lead paint exposure. Poverty data are also adjusted for family size and number of dependents. A drawback to using poverty thresholds is that the associated data are adjusted for cost of living on a national basis but not for regional, state or local variations. If poverty status are thought to underestimate the number of low-income individuals, two times the poverty benchmark should be used. For regional assessments, an analyst might base the decision to use two times the poverty threshold, on whether an area is likely to have higher average incomes and living expenses.

The main advantage to using the set income ranges as benchmarks for low-income status is that the associated data are updated for population counts more frequently than poverty data and thus are more current. In addition, the data are available for most of the same geographic summary levels and cross tabulations of poverty thresholds. A drawback to using income ranges is that associated data are not adjusted for family size or cost of living by geographic area. If the $15,000 benchmark is thought to underestimate the number of low-income households, the $25,000 benchmark should be used. For regional assessments, an analyst might base the decision to use the $25,000 benchmark, on whether an area is likely to have higher average incomes and living expenses.

An advantage to using the HUD Qualified Census Tracts (QCTs) in defining low-income status is that they are available nationally and at a refined geographic level which is useful

for assessing income levels relative to local geographic medians. In addition, they are adjusted for household size, defined by HUD, which is approximately equal to the Census family definition. A drawback to using QCTs is the restriction on the number of tracts which are determined to be QCTs within a single metropolitan statistical area. In addtition, QCTs, as the name implies, are only available at the tract geographic summary level.

4.3 Environmental and Human Health Effect Data

Along with the use of demographic data, it is essential to gain access to information resources that will aid the analyst in the assessment of the environmental risk in an EJ study. The nature of this risk, as well as the ability to measure it accurately, is variable. For example, the risk may involve the release of a hazardous, toxic substance(s). Information of what these substances are, how they are released into the environment, what mechanisms of transfer exist between the source and human populations, and what the effects are on a human population are important. Other risks may involve adverse effects not associated with direct releases of toxic chemicals such as radon or noise from heavy vehicle traffic. Of equal importance is gaining accurate information on the *location* of these substances in space and over time. This issue will be discussed in detail in a later section of the document.

This section provides general information on data sources of environmental and human health effect data. Appendix E describes in considerable detail how to obtain access to the EPA-held data.

4.3.1 Hazardous Substance Release Data in EJ Assessments

There are many data sources for hazardous substance release information that can be used to identify adverse human health and/or environmental effects. This section contains descriptions of these information sources, including sources tracked as mandated by the Clean Air Act; the Clean Water Act; the Emergency Planning and Community Right-to-Know Act; the Comprehensive Environmental Response, Compensation, and Liability Act; and the Resource Conservation and Recovery Act. The types of releases provided in these data sources include both direct and indirect emissions from facilities and potentially hazardous activities and events.

4.3.2 Limitations

It is important to note that limitations exist with respect to adequate characterization of hazardous releases to the environment. Point source release data may be based on actual measurements (e.g., PCS) or estimates (e.g., TRI) that can be off by an order of magnitude or more. The availability and quality of permitted data also varies geographically. Data on a significant percentage of permitted discharges may be unavailable or of poor quality due to insufficient monitoring of releases. Accurate non-point source release information of hazardous chemicals is especially problematic to obtain. The combination of these problems with that of non-permitted discharges often makes the quantitative assessment of risk a difficult task. In addition, many of the greatest risks to the environment are not tracked in databases (SAB report, "Relative Risk", 1990). These include: motor vehicle emissions; non-TRI point sources; area sources, such as gas stations, dry cleaners, etc.; consumer product use; pesticide use; etc.. Thus, it is important to keep these limitations in mind when undertaking an EJ analysis for they will impact on the trustworthiness of the decision-making which the activity supports.

4.3.3 Databases that Track Sites and Facilities

4.3.3.1 The Resource Conservation and Recovery Act Information System (RCRIS) and The Biennial Reporting System (BRS)

The Resource Conservation and Recovery Act (RCRA) established a statutory definition of hazardous waste. The Act directs the EPA and the states to compile information on the operations of hazardous waste handlers. General information on waste handlers (such as name, address, and activities); permit information; corrective action program status; and enforcement and compliance data are stored in the RCRIS database.

The Hazardous and Solid Waste Amendment (HSWA) requires all hazardous waste generators; and Treatment, Storage, and Disposal (TSD) facilities that handle hazardous waste to file reports with EPA administrators at least once every two years. Information gathered and entered in the Biennial Reporting System (BRS) database. Data are reported to BRS administrators by the various categories of waste stream that are either generated, received, managed, or shipped. These waste streams can be either 100 percent RCRA-listed wastes, or a mixture of one or more toxic materials contained at various concentrations in a non-hazardous matrix (for example, railroad gravel). In addition, a waste stream can be described by multiple RCRA waste codes (e.g., a waste stream can simultaneously be ignitable, contain spent halogenated solvents, and contain benzene). At present, there is no mechanism to apportion the waste stream volume to particular waste codes where multiple codes are reported. Waste codes that qualify as hazardous under BRS are waste characteristics (such as ignitable, corrosive, reactive); individual chemicals (such as arsenic, chromium, lead), and process by-products (such as distillation still bottoms).

Data contained in the BRS database that may be of use in EJ analyses include facility status (such as active, inactive, closed); enforcement and compliance data; type of activity (such as disposal methods, incineration); the quantity and type of hazardous waste at the facility; and facility capacity information. The following is a list of the sites that are tracked in the BRS database:

- *Treatment, Storage, Disposal (TSD) facilities* - Wastes classified as hazardous under RCRA are managed in TSD facilities. These facilities are potential sources of exposure via air, soil, groundwater, or surface water through normal operations or through accidental releases.

- *Corrective Action Sites (CAS)* - These are specific sites, facilities, or areas that are currently regulated due to past (pre-RCRA) contamination from toxic waste generation and/or management activities. CAS are ranked for remediation according to the National Corrective Action Prioritization System (NCAPS) as of High, Medium, or Low priority. These sites present potential exposure - principally through groundwater contamination.

- *Hazardous Waste Generators* - Facilities that actively generate hazardous wastes are tracked by the BRS. However, only Large Quantity Generators (LQG) of hazardous waste report activities through the system. LQG facilities are potential sources of exposure (principally via the air) through normal operations or accidental releases. LQGs are a surrogate indicator of industrial activity. It should be noted that a LQG facility may be both a waste generator and a TSD site, and that individual TSDs and LQGs may have a Corrective Action Site on their property.

Facilities regulated under RCRA (especially TSDs) are frequently examined in EJ analyses (for examples see: United Church of Christ, 1987; Sadd/Occidental, 1996; Anderton et al., 1994; Mohai and Bryant, 1992) that have addressed the proximity of hazardous waste facilities to low-income and minority communities. There are current efforts underway to compile and rank toxicity data of RCRA chemicals (Waste Minimum Report...). This information can be used to evaluate and compare potential risk posed by RCRA facilities and in EJ analyses.

4.3.3.2 Comprehensive Environmental Response, Compensation, and Liability Act Information System (CERCLIS)

CERCLIS is the EPA repository for site and non-site specific Superfund data in support of the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA). CERCLIS contains information on hazardous waste site assessment and remediation (from 1983 to the present) for all potential Superfund sites, as well as "Proposed" and "Final" sites that have been listed on the National Priority List (NPL).

Sites that are listed on the NPL can also be found in the NPL database that is a subset of CERCLIS. The NPL database contains approximately 1,200 sites and provides more detail on the sites than the CERCLIS database. Additional information concerning Superfund sites (such as nearby stream or groundwater flows) may be available from Superfund site managers.

Data available through the CERCLIS and the NPL that can be used to support the assessments of risk include the stage of activity/priority (such as discovery phase, preliminary assessment, site investigation, or removal action); major constituents and contaminants; and a Hazard Ranking System (HRS) score. The HRS is a measure of the relative threat posed by the site to human health and to the environment. The HRS score is used to decide whether the site should be on the NPL. Scores vary from 0 (lowest threat) to 100 (highest threat). A site scoring 28.5 or greater is eligible for the NPL.

One caveat to using CERCLIS data for assessments of risk is that Superfund sites typically have no single point of toxic material release (such as a factory smokestack). Larger Superfund sites can span areas more than 20 square miles. Another note worthy point is that human exposure to Superfund sites is limited and so risk may be less than what might be expected based on concentrations of chemicals present.

4.3.3.3 Toxic Release Inventory Information System (TRIS)

The Toxic Release Inventory (TRI) is a facility-level database created under Section 313 of the Emergency Planning and Community Right-to-Know Act (EPCRA). TRI contains information about releases and transfers of more than 650 toxic chemicals to the environment. Under §313 of EPCRA manufacturing facilities with 10 or more full-time employees that meet the established "manufacture," "process," or "otherwise use" thresholds for the listed chemicals must report their releases and transfers. The "manufacture" or "process" threshold is currently 25,000 pounds per year for each listed chemical. The "otherwise use" threshold is 10,000 pounds per year. Manufacturing facilities include (but are not limited to): chemical manufacturers; petroleum refiners; primary metal processors; fabricated metal processors; paper, rubber and plastic manufactures; and transportation equipment producers. The EPCRA does not require such facilities to monitor their releases. Instead, EPA provides guidelines for estimating releases along with the TRI reporting package.

When using TRI data for EJ assessments it is important to recognize that the system reports on only larger facilities (both in terms of number of employees and amount of materials involved), on a limited number of chemicals, and on specific manufacturing sectors. Additional sources of release data may therefore be required for a more complete assessment of risk. The chemical toxicity of the chemicals released are also available from OPPTS Risk Screening Environmental Indicators Model. This information can be used to more accurately estimate potential risk from TRI chemical discharges.

4.3.3.4 AIRS Facility Subsystem (AFS)

The AFS database is a subset of EPA's Aerometric Information Retrieval System (AIRS) which addresses the regulatory requirements of the Clean Air Act. The AFS contains criteria pollutant emissions data; regulatory compliance data; and permit data for point sources tracked by EPA and by state and local air pollution agencies. Regulated sources cover the spectrum from large industrial sites to smaller operations such as dry cleaning facilities. Point sources included in AFS are those emitting: 100 tons per year of VOCs, NO_2, SO_2, or PM-10; 1,000 tons per year of CO; or 5 tons per year of Lead. Emissions data from these point sources are collected by the states and local agencies, and then reported to AFS administrators. Compliance data recorded in the AFS are maintained at the facility level and at the sub-facility level (for individual pieces of equipment or line processes, for example). AFS data may be useful for identifying facilities within non-attainment areas, especially where the designated non-attainment area is fairly large.

One caveat to using AFS release data for risk assessment is that most AFS facilities prepare emissions inventories only once every five years. It is therefore possible that the emissions data recorded in the AFS are somewhat out of date. Also note worthy is that release information is generally available for only five criteria air pollutants: SO_2, NO_2, CO, O_3, PM-10, and Pb. Release estimates can be made for many other toxic chemicals using a model available from OAR (Ref). Use of this model, like other models, provide additional information to the analyst but also introduce greater uncertainty to the analysis being performed.

4.3.3.5 Permit Compliance System (PCS)

The PCS tracks the permit, compliance, and enforcement status for facilities that discharge under the National Pollutant Discharge Elimination System (NPDES) permit program of the Clean Water Act. PCS tracks information on Federal, municipal, and industrial discharges into surface waters. Information in the PCS that may be useful in EJ identification efforts includes discharge characteristics; permit conditions; inspections; enforcement actions; and compliance schedules.

Discharge data submitted by facility owners is in the form of Discharge Monitoring Reports (DMRs). Discharge amounts can be reported as either daily or monthly average concentrations (recorded as releases assessed in milligrams per liter, or parts per million, for example), or as a daily or monthly average loading (recorded as pounds per day, or kilograms per day, for example). Monthly and annual loading measures have been estimated using the Effluent Data Statistics (EDS) system (Office of Enforcement and Compliance, US EPA (19??)) and are available from BASINS (Better Assessment Science Integrating Point & Nonpoint Sources) (Office of Science and Technology, Office of Water, US EPA (1998)), and LandViewIII (Chemical Emergency Preparedness and Preparation Office, US EPA (1998)).

PCS distinguishes between major discharges and minor discharges based on potential threat to health and to the environment. Only discharge information from major facilities are required to be entered into the PCS, and so minor facilities are under-represented. The utility of PCS data in EJ analyses is lower when the location of a permitted facility's out-fall pipe is not available. Some PCS records only show the location of the principal facility and not, when they exist, secondary facilities. Some PCS records only indicate the corporate address rather than the give information on the actual location of the toxic material release point.

4.3.3.6 Solid Waste Landfills

The location and characteristics of solid waste landfills are not tracked or managed at the national level. However, many EPA Regions have developed data systems that store information on landfills. Due to database design variations there is not consistency among the systems in terms of the location and attribute characteristics tracked. State and local environmental agencies generally have more information on solid waste landfills than EPA since such agencies in most instances have primary authority for regulating the sites.

Solid waste landfills handle mostly municipal waste produced by households and so are contaminated with hazardous waste that are largely exempt from regulation (such as cleaning agents and paints used by household members) or regulated materials disposed of illegally as "household" waste. The principal exposure pathway related to substances in solid waste landfills is through groundwater contamination, which can affect drinking water supplies, although there can be some ambient air releases from the sites. These sites often receive heavy truck traffic that may result in localized air pollution concerns.

4.3.4 Databases that Track Multi-Media Enforcement and Compliance Data

4.3.4.1 American Indian Lands Environmental Support Project (AILESP)

The goal of the American Indian Lands Environmental Support Project (AILESP) is to improve understanding and management of sources and impacts of chemicals on and near Tribal lands. Staff from all EPA regions with Tribal lands and eight tribes (St. Regis Mohawk Tribe, Red Lake Tribe, San Carlos Apache Tribe, Omaha Tribe of Nebraska, Cheyene-River Souix Tribe, Shone-Bannock Tribes, Tuscarora Nation, and Penobscot Nation) have participated in the development and testing of the AILESP over the past eighteen months. Phase One of the project involves compilation and distribution of the draft AILESP database and ArcView project for comment to pilot tribes and regions (Summer and Fall of 1997). In Phase Two, the project will be refined, expanded, and updated on a regular basis (based comments received from EPA regions and Tribes) to serve as an effective tool to improve understanding and management of sources and

impacts of toxic chemicals on and near Tribal lands. It is envisioned that this tool will be used by EPA regions and tribes both directly to evaluate and map facilities on and near Tribal lands, and indirectly by soliciting specific database query and mapping requests to the OECA Tribal Program in Washington D.C.

The AILESP database is a compilation of available data from EPA datasystems, specifically the Facility Indexing System (FINDS), AIRS Facility Subsystem (AFS), Permit Compliance System (PCS), Toxics Release Inventory System (TRI), and Biennial Storage System (BRS), on names and locations, recent compliance and enforcement histories, as well as estimates of the amounts and types of chemicals being released to the air, land, and water for permitted federal, industrial, commercial, and municipal facilities regulated under the Clean Air Act, the Clean Water Act, and the Resource Conservation and Recovery Act located on and near (5 km) Tribal lands. Because AILESP data comes directly from EPA datasystems, errors that exist in the core databases are also present in AILESP. In addition to facility-specific data, aquatic impact monitoring data, such as stream reaches with fish consumption advisories, contaminated fish tissues, and contaminated sediments were also included in the AILESP database. The database has been integrated into an ArcView 2.1 geographic information system (GIS) project which allows mapping and analysis of information at the national, regional, and/or reservation level.

The AILESP is available online at: http://es.epa.gov/oeca/ailesp/index.html.

4.3.4.2 Sector Facility Indexing Project (SFIP)

The goals of the Sector Facility Indexing Project (SFIP), a pilot program developed by EPA, are to make environmental information about industrial facilities and regulatory compliance more accessible to the public; to expand the comprehensiveness and improve the accuracy of data for analyzing the environmental track record of an industrial sector; to provide industrial and government stakeholders with better analytical tools for permitting, reporting, compliance, benchmarking, self-policing, and pollution prevention purposes; and, to help all stakeholders take a more holistic, multi-media approach to environmentally sound performance.

SFIP integrates and provides public access to environmental information on approximately 650 individual facilities in five industrial sectors: automobile assembly, pulp manufacturing, petroleum refining, iron and steel production, and the primary smelting and refining of aluminum, copper, lead, and zinc (nonferrous metals).

SFIP provides environmental data about each facility, such as the number of inspections, compliance with federal regulations, enforcement actions taken, chemical releases, and spills. SFIP also includes background information on the location and production capacity of each facility, as well as information on the population of the surrounding area.

EPA worked for three years to identify the facilities to be included in SFIP, and to collect and verify the data. Each facility received a copy of its compliance and enforcement data for review, to make sure that any problems were identified before the information was distributed through SFIP. Based upon these data reviews, EPA believes that the information in the underlying databases is generally of high quality. EPA will continue to work with stakeholders to ensure the greatest possible quality of data in SFIP.

The SFIP is available online for use by the general public. The SFIP Internet address is as follows: http://www.epa.gov/oeca/sfi.

4.3.5 Databases that Track Potentially Hazardous Activities and Events

4.3.5.1 Pesticide Applications

Over one billion pounds of pesticides are applied in the US every year. Applications occur mostly on agricultural farms, but also on parks and other recreational areas, on residential lawns, and in and around institutions or areas which people frequent. A number of these chemicals are potentially toxic to humans and the environment if not used according to regulation. Improper pesticide applications can result in a number of hazardous exposure scenarios. These include occupational exposures to farm workers (many of whom are lower-income and/or minority workers); spray drift to nearby residential areas, schools, and community institutions; contamination of underground aquifers and drinking water wells; and ecologically damaging runoff to nearby surface waters.

Pesticide application and poisoning incidents are tracked by state, local, and agricultural agencies. Reporting varies from state-to-state. Some states keep detailed records of pesticide applications (California, for example), while many others do not. Thus, the consideration of pesticide applications as part of an EJ analysis may be difficult or impossible in certain geographic areas. Since the great majority of all pesticides are applied in rural areas, a lack of information may not necessarily impact in a negative way on urban-centered EJ analysis. However, when good pesticide application data are available an attempt should be made to factor in not only the volume of pesticides applied, but also the pesticide's relative toxicity and pathway.

4.3.5.2 Hazardous Material Spills

The Emergency Response Notification System (ERNS) is a national database that stores information on release notifications of oils and other hazardous substances to the air land or water. Information stored in ERNS may include: the date and location of the spill or release; the amount and type of material released; any damage or injuries related to the release; the environmental medium into which the release occurred; and other

circumstances of the spill or release (US EPA OECA, 1995). Some caution must be employed if ERNS data are used in EJ identification analyses since a single incident (a spill or release) may result in more than one report logged into the system. As ERNS data identifies areas that have experienced hazardous material spills in the past, the data may be useful in predicting the location and risk of future spills and releases.

4.3.6 Human Health and Ecosystem Effects of Adverse Environmental Conditions

4.3.6.1 Ambient Air Pollution: Non-Attainment Areas for Criteria Pollutants

The AIRS air quality system also contains measurements of ambient concentrations of criteria air pollutants (SO_2, NO_2, CO, O_3, PM-10, and Pb). The data are used to assess the status of the nation's air quality and to identify areas where levels of one or more air pollutants exceed National Ambient Air Quality Standards (NAAQS). Information on which geographic areas are in exceedence of air quality criteria may be obtained from OAR's web site. Areas that exceed the NAAQS would be important components of a cumulative risk analysis of a potential EJ community.

4.3.6.2 Water Pollution (Surface)

Under section 305(b) of the Clean Water Act (CWA) every state must submit an assessment of its water quality on a five-year cycle. These assessments vary greatly in terms of the quality and quantity of information given in the 305(b) reports. States assess the quality of their surface waters by determining if they meet designated beneficial use standards (e.g., drinking water supply, aquatic life support, primary contact recreation and swimming). This data is currently being digitized through a joint effort between EPA and the states. A consistent national georeferenced database of this information may be available from the Office of the Wetlands, Oceans, and Watersheds (OWOW), EPA headquarters and regional offices in January 1999. Data from many individual states may be available from EPA before then. The Storage and Retrieval (STORET) database of ambient water quality data contains primarily chemical and physical water quality data generated from monitoring activities. STORET contains information of varying quality and quantity, much of which is older than five years and may not be reflective of current conditions. Both 305(b) and STORET data may be used to identify "impaired" surface water bodies and water quality exceedances that may have an effect on a nearby community. Water pollution may pose a potential public health concern due to an effect on drinking water supplies or contaminated fishing/recreational areas. In absence of an explicit exposure pathway, proximate impaired water bodies can still be considered as a component of a "disproportionate effect" assessment.

4.3.6.3 Water Pollution (Ground)

States are also required to assess the quality of their groundwater in their CWA section 305(b) reports. In general, groundwater pollution tends to be a more localized problem than surface water pollution as the pollution often results from a specific contaminant source (e.g., a leaking underground tank, a nearby farm, or landfill). Nationally, groundwater provides drinking water for 53 percent of the US population and nearly all of the rural population (OW RTC, 1994). Frequently identified contaminants of groundwater include pesticides, nitrates, volatile organic chemicals, and radio nuclides. Available groundwater monitoring data, collected on a state-by-state basis, is quite variable. Thus, the availability of data may limit the instances where groundwater contamination may be used as part of an EJ analysis. There currently does not exist a national database of groundwater contamination. However, when there are reliable data available, it may be possible to generate some quantitative risk estimates (assuming the chemical contaminant is identified and has some known measures of toxicity) of contaminated wells or aquifers used for drinking water.

4.3.6.4 Drinking Water

Data that characterize the quality of drinking water sources for use by community water systems is not available nationally. However, surrogate measures may be obtained using the Public Water Systems (PWS) inventory and the Community Waters Systems (CWS) which is a subset of PWS. The PWS contains infromation on the quality of many water bodies which supply drinking water sources, but not the quality of the drinking water at the tap which can be degraded due to old pipes and lead soldering, etc.. Public water systems are required to take whatever actions are necessary to ensure the drinking water they provide meets state and EPA drinking water standards. Source waters that are impaired or threatened prompt water systems to take corrective actions, some of which can be captured by compliance data reported to EPA.

4.3.6.5 Fish and Sediment Monitoring Data

Another condition that is related to pollution is the contamination of fish tissue and the sediments under water courses and water bodies. In 1993, 93 percent of fish consumption advisories were issued after mercury, PCBs, chlordane, dioxin, and DDT contamination (OW RTC, 1994). These advisories vary considerably among states in terms of the criteria according to which advisories are issued. Minnesota, for example, uses a much higher standard than South Carolina so that even though Minnesota has many more advisories than South Carolina, levels of contaminants in South Carolina fish may overall be higher and pose a higher risk. EPA tracks issuance of such advisories in its National Fish Consumption Advisory (NFCA) database. The database is limited in that it only counts one advisory per water body, even if multiple fish species in different parts of the water body are contaminated. Fish advisories, contaminated sediments, and fish tissue

data from the recently compiled National Sediment Inventory (NSI) database have been used to link pollutant discharges to human health and aquatic life effects in minority and low-income communities (Fox, 1997).

4.3.6.6 Radon Gas

$Radon^{222}$ is a gas produced by the radioactive decay of Radium, an element that occurs naturally in most soils and rock. Due to the geographic variation in sub-surface geology Radon concentrations released into the atmosphere vary across the United States. Radon gas emitted from the soil and rock infiltrates into buildings of all types. Radon is classified as a known human carcinogen based on extensive epidemiological study and poses the greatest risks to smokers (due to a synergistic interaction between smoking and Radon) in causing lung cancers. In 1992, EPA and the United States Geological Survey (USGS) developed a map showing areas with high radon potential at the state and at the county level.

4.3.6.7 Lead

Lead is frequently present in the pipe work and in paints used on older (pre-1950) housing stock. Communities that are proximate to transportation corridors may experience high levels of lead deposited on soils after years of lead-based gasoline exhaust emissions. Young children are more susceptible to the neurological affects of lead than adults. Due to the nature of their play activities they are likely to be exposed to lead through ingestion of contaminated paint chips, household dust, and soil. The degree of lead contamination in humans is usually characterized by the amount of lead found present in blood and is expressed as the number of milligrams of lead (element Pb) per deciliter (mg/dl). The Center for Disease Control defines children with lead levels in excess of 10 mg/dl as experiencing "lead poisoning" and those with levels between 10 and 25 mg/dl are classified in an "action-level" range. Data on lead contamination are kept at state and county health departments. Non-aggregated data on human health is subject to confidentiality restrictions and thus aggregated data is often that which is made available. The Bureau of the Census provides demographic data that can be used to determine the number and percentage of children in an EJ Area of Concern. The Bureau of the Census makes age of housing stock data available too. Additionally, the location of transportation corridors can be identified using the Bureau's TIGER files. These variables may be used in combination to "predict" possible exposure to sensitive populations.

4.3.6.8 Respiratory Disease Prevalence

Tobacco smoking is the overwhelming causative factor in many respiratory diseases. Air pollution is another known causative factor and the prevalence of respiratory diseases (such as Bronchitis and Asthma) rise with the level and duration of air pollution events.

Any use of data related to the incidence of respiratory diseases must be interpreted with caution by those unfamiliar with the cause-and-effect relationships involved. African-American children have been shown to suffer higher rates of Asthma than the general population (Reference ?).

4.3.6.9 Cancer Cases/Cancer Clusters

In the aggregate, cancer is second only to heart disease as a cause of death in the US. Cancers accounts for just under one-quarter of all deaths. Because of the unknown etiology (assignment of cause) of many cancers, environmental pollutants are often blamed by the public for increases in regional cancer rates or instances of real (or perceived) cancer clusters (unusual geographic concentrations). The National Cancer Institute (NCI) supports a network of cancer registries located throughout the country. While there are environmental pollutants that are known (or highly suspected) to cause cancer in humans, the use of cancer registry data in EJ analyses should be approached with extreme caution. This caution is warranted because of the scientific uncertainty surrounding the role of environmental factors in carcinogenesis, the potential for pointed controversy among community members, health officials, and other interested stakeholders in addressing this issue; and the complex nature of the methods used in the analysis of cancer clusters.

4.4 Location Data Available for Facilities Tracked by EPA Databases

4.4.1 Location Data Collection Efforts and Limitations on Location Data Use

For many EJ analyses, the source(s) of pollution or the adverse environmental condition that is to be evaluated must be assigned a location. A number of alternatives can used to locate a source(s) and these are explained in the *Methodology* section. The sources for the data supporting those methods are presented in this section, along with EPA's location data collection efforts and associated limitations.

The locations of monitored facilities and remediation sites collected by the EPA, by state and local environmental agencies and their contractors are available via many systems. Also, the use of self-reporting by regulated operators has been a common method used to acquire location data on sites, facilities, areas, and regions of environmental concern.

The process and method of collecting location data referred to as "map interpolation" is based upon the review of property records and associated maps that are used to *estimate* location in the latitude, longitude coordinate system which are commonly used to define location on the surface of the Earth. A significant proportion of location data collected by EPA and others was generated by this method. EPA acknowledged the critical need for

the collection of higher quality location data (*better positional estimates*) in the issuance of its Locational Data Policy (LDP) (US EPA, 1991). The LDP mandates preferred location data collection methodologies, as well as defines accuracy and verification procedures, and the reporting of location data for regulated entities. The LDP directs EPA Programs to establish a standard for their location data with a default goal of ±25 meter accuracy. The LDP helps all EPA Programs and state and local agency efforts in that it promotes the use of all data with some form of embedded location information (the vast majority of all data held by EPA). Application development efforts (for use by the Programs for which the data were collected, or to foster use by others in, for example, EJ analyses) are aided by the Locational Data Policy.

EPA's OIRM developed and maintains guidance for documenting location data with its Locational Data Improvement Plan (LDIP) (US EPA 19??) and use of Method, Accuracy and Description (MAD) codes. These codes provide a standardized method of recording and storing location data in EPA systems. OIRM has performed studies evaluating existing data collection projects and programs, and estimated the costs and accuracy of collecting data using various new approaches. There are two methods in addition to map interpolation now in use. The first method is the field-based collection of location coordinates with the use of Global Positioning System (GPS) satellite receivers. The second method is referred to as "address matching" or "geocoding." The method uses software to match a reported facility address against a reference database of all known valid addresses together with their location coordinates. A successful match results in the reported facility address record being linked to the location coordinates from the reference database.

While address matching /geocoding can usually be performed relatively inexpensively (by the analyst using readily available commercial software) its accuracy depends on the "quality" of the address describing a facility location. By "quality" it is meant that the non-standardized forms by which addresses are recorded lead to numerous and complex quality and accuracy issues. There is, in large group of address records, a very wide variety of ways (and many nuances to) that an address will be recorded. Also, an address may record the location of a corporate (facility-owning) office rather than the location of the facility itself. A Post Office Box (which can only be linked to the centroid of the Post Office to which the mail is addressed) may be included in the address information. Many EPA program systems do not maintain separate (delimitated) fields for the elements of a mailing address (street number, street name, place/city, ZIP code, state). Though the Facility Indexing System (FINDS), a centralized inventory of facilities monitored by the EPA contain this information and can serve as an index database to other EPA program databases. These elements recorded in a delimited format are needed to make full use of address matching software an gain a high positive match rate between the "raw" non-standardized addresses records and the master location reference table.

One advantage of using longitude and latitude coordinate data is that they can define location with high accuracy. Location data quality is getting considerably better with the use GPS technology. Such location data are becoming more accessible from contemporary information systems with improving user interfaces that support advanced geographic (spatial) queries rather that just text-string based queries.

While Global Positioning Systems can yield exceptionally accurate location estimates, achieving these usually requires digital data analysis by trained staff subsequent to the collection in the field, substantially increasing its cost. The method is good for obtaining coordinates for centroids or reference points on the perimeter of a facility. Obtaining locations of the plant interiors usually requires additional costs for facility owners and specialized staff such as technicians who have facility access.

Over the years, an individual regulated facility will have been investigated often and some form or other of location data collected. More recent efforts to perform address matching and/or use GPS technology will often have resulted in the collection of additional location information. Collectively these efforts have frequently resulted in a *set* of location information types (including postal addresses and coordinates) being available, of varying quality in degrees of completeness and in accuracy. Thus the selection of location data for an application such as an EJ analysis is not always straightforward.

In 1996 the OIRM launched a campaign to collect latitude and longitude coordinates for EPA Program database system records. The project culminated in the development of the Locational Reference Table (LRT). Some EPA Regions, state and local agencies have also launched data collection and reconciliation efforts. New/existing facility location identification one-stop reporting initiatives have prompted discussion of the benefits and costs of various approaches to data acquisition and recording, and the types of situations in which each may be most appropriate.

Most existing program systems cannot accommodate polygonal (area) feature types nor offer a complete set of documentation on the accuracy of the data already collected. The diversity of EPA's programmatic database systems in terms of their design and implementation makes it technically difficult and expensive to integrate location (and associated attribute) data across program (multi-media) lines. Also, much of the location data are collected independently by federal, state and local agencies, and according to different criteria and methods, and are held in many physical and virtual places. Exchanging location data between EPA systems and others is sometimes problematic, but data publishing programs help to alleviate this issue.

Based on an address or other location data it is often possible to associate a facility with an administrative area such as a county, ZIP code, or census tract. The OIRM's address matching efforts have included the assignment of block group or census tract codes for a

large number of EPA-regulated facility records. There are a number of uncertainties in this process reflected, for example, in the number of facilities which cross *over* administrative boundaries.

As a result of the limitations imposed by data collection methods, data availability, and data quality, some facilities in potential EJ areas of concern cannot be included in the analysis *with confidence* because of missing or inaccurate or confusing location data. EJ analyses may be delayed, or their cost increased, by the requirement to perform primary data collection undertaken in order to resolve such problems.

4.4.2 Ongoing Location Data Activities

OIRM has invited EPA Regions and state agencies to contribute to location data held in the LRT which is accessible through ENVIROFACTS. EPA programs are collecting additional latitude and longitude coordinate data, assessing data quality, or otherwise addressing the issue. Many EPA Regions and states are collecting data on a piece-meal basis for individual projects. Some states are performing comprehensive data collection, but the results are rarely combined or reconciled with EPA program data. The OIRM is coordinating Agency activities through the IRM (?) Steering Committee process. EPA Regional GIS and OIRM staff continue to develop policies, methods and data structures towards a comprehensive collection of accurate location data..

The OIRM assessed the status of EPA location data in 1996 and found that only about 27 percent of the approximately 800,000 records in ENVIROFACTS had any latitude and longitude coordinate values recorded, and only about 10 percent were at least partially documented with MAD codes. The OIRM performed address matching on all records available, and increased the percentage of records with latitude and longitude values to about 80 percent, with about 61 percent including documentation. Of these, a little over half have locations with estimated accuracy of 500 meters or better. The OIRM is planning to provide support to Regions to document and merge existing location data collections of good quality. In cooperation with Regional and the OPPT's GIS teams, the OIRM has developed automated quality assurance checks to perform for location data. GIS teams are working together to draft methods for selecting a default preferred location coordinate when multiple values exist for the same facility, and to document any verification or quality assurance steps performed on the data.

4.4.3 Recommendations for Improving Use of Location Data

There needs to be improved coordination between staff, technical assessors, program system managers, and GIS teams regarding locational issues. Better ongoing communication and the development of a coordinated plan for addressing location data

problems across the Agency is needed. Additional research in both location data needs and methodology as well as a need to assess the magnitude of the impact of location data issues on analyses is necessary. Good information on the cost, infrastructure requirements, reporting burden, and quality of various location data acquisition and collection methods is also needed.

Despite achievements made by the LDIP, EPA needs to continue to refine policy and set appropriate goals and work to achieve them. The Agency needs to assess the status of its location data subsequent to the OIRM's enhancement efforts, and fill critical data gaps. One example of such a goal might be to obtain facility location data of 25-meter accuracy for plant, facility, or area centroids on at least 95 percent of high-priority facilities by the year 2000.

5. Methodologies for Conducting EJ Analyses

5.1 Self-Identified Communities

An analyst may become aware of a community in need through an EJ identification exercise or a community's self-identification. An EJ identification exercise is an objective study designed to identify those communities in need of further EJ investigation. Since an EJ identification study is limited to using available data, which may be out of date or incorrect, some communities in need may not be detected through this process. Self-identification is an important tool for ensuring that some of the communities that are not identified through the EJ identification process do not fall through the cracks. Self-identified communities are brought to EPA's attention by the community's first-hand experience with environmental problems or by other person's experiences with similar environmental problems. For example, representatives from self-identified EJ communities have spoken at many of the NEJAC Meetings in the past, and have brought to the Agency's attention potential EJ areas of concern.

There are several EPA sources of information for self-identified communities. These include the following:

(1) public comments to the National Environmental Justice Advisory Council (NEJAC);

(2) public comments at Executive Order-mandated public meetings (e.g., Atlanta in 1995);

(3) communities represented at the February 1994 Environmental Justice Health Symposium;

(4) Title VI cases on file with the Office of Civil Rights;

(5) Office of Environmental Justice (OEJ) hot-line database and areas of concern;

(6) American Indian Office areas of concern;

(7) Records of public hearings from permitting, EIS's, etc.

The Federal Facilities Environmental Justice Enforcement Initiative (FFEJEI) included communities which identified themselves as "at risk" to the Office of Environmental Justice, at the National Environmental Justice Advisory Council Meetings, the EJ Hot Line and the American Indian Environmental Office (Boerlage, 1997). Federal facilities in these communities were grouped with other facilities identified through the use of

more quantitative (i.e. databases) approaches. Region IV and VI similarly consider community self-identification as an important element of EJ analysis, and so directly integrate self-identified communities into their EJ identification processes.

5.1.1 Recommendations for Considering Self-Identified Communities

Keeping in mind current limitations in data quality and quantity, self-identified communities should be considered and included as much as possible in EJ analyses. EPA Regions, states, or the communities themselves should attempt to collect environmental and human health data for the identified community that may be incomplete in EPA's national databases but may be available for other sites already included in the proposed analysis. Self-identified communities should be considered potential EJ areas of concern unless they can be ruled out as a result of more rigorous analyses including a comparison to a reference area. At an early stage in the study, information regarding the purpose, method and criterion of the EJ assessment should be disseminated to Regional EJ coordinators, NEJAC, and others for comment as well as to identify potential EJ communities they feel should be included in the analysis.

5.2 Types of EJ Analyses

The two most common types of EJ assessments are targeting/screening assessments and site specific assessments. Targeting/screening assessments are proactive analyses aimed at characterizing possible areas in need of assistance. These assessments support, for the most part, EPA internal decision making practices such as allocating resources, increasing enforcement targeting, or qualifying communities for grants. These assessments are usually high level studies with the objective of broadly describing a geographic area and its associated population. Generally, they have high levels of allowable uncertainty and as such, involve the use of few, if any, demanding quantitative methodologies and could be performed with minimal lead time.

Site specific assessments are analyses performed in reaction to a known source(s) of potential adverse environmental and/or human health effects, or any number of other possible effects, such as economic, social, nuisance, etc., often including, but not restricted to, a polluting facility. Some other sources of possible effects include motor vehicles, overflowing sewers, ozone or particulate levels resulting from many sources, and actions normally associated with other Federal agencies such as new roads or other transport projects, various actions taken on Federal lands, etc.. The decisions hinging on these assessments may include determining the need to reassess agency policies, regulations, or the need for legislative changes; Title VI cases; EPA or a state approval or modification of a permit; etc.. As mentioned in the conceptual framework, these assessments may be highly visible and potentially controversial. As such, they may

require low levels of uncertainty (or high levels of accuracy) and high quality data and rigorous methodologies.

5.3 Discussion of Procedural Approaches

The Executive Order offers a fairly clear definition of an environmental justice area as a geographic area that is minority and/or low-income, and experiences a disproportionate environmental burden. However, the issue of how to begin the analysis is not clearly defined. Should an analyst begin with an evaluation of demographic data, then environmental effect data, or vice versa?

Many EPA analysts recommend looking at demographics first, and then considering environmental data when conducting an EJ analysis. The benefits to using this approach is that demographic analyses can almost always be performed quantitatively, resulting in relatively consistent and solid identifications. Subsequent evaluation of effects can be examined qualitatively, if necessary or possible. The results of the demographic analysis can still be used effectively for inspection targeting.

An alternate approach is to consider where the greatest effects are occurring first, and then out of all of the high effects areas, determine which ones have significant minority and/or low-income populations. Within a given analysis or reference area, the question becomes whether or not a disproportionate number or percentage of the high effect areas are found to be minority and/or low-income. The justification for using this approach is that analysts can avoid making implications about a geographic area simply because it has a minority and/or low-income population. A drawback to this approach is that some minority and low-income communities might be overlooked, since traditional ways of assessing effects may not sufficiently include all potential effects to a community. For example, indoor air pollution is a great concern in many communities, but EPA and others have yet to collect enough data on this problem (EJ Focus Group meeting, November 13, 1996).

A third approach involves combining minority, low-income, and environmental burden factors to determine an 'EJ score' for a particular study area, then ranking that EJ score relative to other study areas. This approach, which was developed by Region VIII, determines the study areas based on the existence of permitted facility sites. Once the study or target areas are determined, environmental burdens and demographics information are analyzed to determine the EJ score. There are several advantages to using this approach. First, it enables ranking of communities for the purpose of identifying areas in greatest need of relief. It also eliminates the need for reference areas. Finally, this approach enables consideration of communities in need that may have been overlooked because they narrowly missed the threshold levels for both minority and low-income.

A drawback to using this approach is that it only narrowly considers areas that contain permitted facilities and associated effects. Currently, the analysis of environmental burden information is limited to a raw count of proximate TRI facilities. Analysts should be encouraged to capture other environmental burdens with as much rigor as possible. For example, in order of increasing complexity, analysts might use: 1) aggregate TRI releases instead of raw count facilities, 2) aggregate TRI releases with a consideration of constituent toxicity (e.g., Deb Forman's Chronic Index), or 3) aggregate TRI releases with a consideration of constituent toxicity and other known sources of exposure (e.g., ambient air quality). When possible, other known sources of adverse effects such as motor vehicles, power generation stations, dry cleaner operations, gas stations, etc., should be considered.

A final approach was used by Region III a few years ago. This approach used existing information about Region III to establish the significance of its community's problems. After identifying a number of communities of significant concern, Region III brought together individual medium (e.g., indoor air, water, waste) experts to discuss what was known about these communities and to rank the problems for possible action. It should be noted that this approach does not result in reproducible quantitative results, but is a way of dealing with limited and/or incompatible multi-media data.

5.4 GIS Technologies Employed in an EJ Analysis

Most EJ assessments rely on Geographic Information System (GIS) tools to manage, analyze and display data. As such it is necessary to understand the fundamental concepts of a GIS and how they relate to EJ specific analyses.

5.4.1 Geographic Data and Analysis within a GIS

The traditional way to record geographic information is in the form of a map. We frequently use printed maps but the information which these maps convey is almost always stored in a digital format and manipulated with computer software specifically designed to handle geographic information. A Geographic Information System (GIS) is a computer-based environment in which digital-format map data are stored, displayed, and analyzed. In a GIS, maps are handled just like documents are handled inside a Word Processing System.

The nature of the geographic information that appears on a map exists in a number of different forms, often referred to as "geographic feature types". First consider the geographic type "point feature". A smokestack from a coal-burning power plant would be best represented as a single point on map of small scale. In other words, on a map of small scale the best way to represent the smokestack's footprint is as a single point located within the large area. The smokestack's position, other than being shown visually

on the map, is recorded numerically as a pair of coordinates (for example, the latitude and longitude of the smokestack's center when viewed from above). Once a location of a point feature is established, we can then go on to describe its characteristics. These are often referred to its "attributes". The attributes of a point feature such as a smokestack would include descriptions of what it is (e.g. "smokestack"), how high it is (e.g. "100 feet"), what it emits (e.g. "Sulfur Dioxide"), and how much (e.g. "ten tons per year").

Linear features (often referred to as "arcs") are a second geographic feature type to consider. A good example would be a river or a pipeline. These features are by nature "linear" - narrow and elongated. In the consideration of the EJ conditions prevalent along a highway corridor (which is a known source of lead emissions from transient automobiles) the highway would be best represented as a single line on map of small scale. In other words, on a map of small scale the best way to represent the highway's footprint is as a line running through the large area. The highway's position, other than being shown visually on the map, is recorded numerically as a string of points (each point as a coordinate pair). The feature is recorded as a joined string of points, but displayed as continuous line which changes direction as the actual feature does in real life. Linear features have attributes too, such as their name (e.g. "Interstate 5"), the total number of lanes (e.g. "four"), and the volume of lead emissions (e.g. "0.5 pounds per year per mile").

Area features (often referred to as "polygons") are the third geographic feature type to consider. These features are by nature "areal" - occupying a substantial extent of space. A good example would be a spatially extensive petroleum refinery. A petroleum refinery could be represented as a footprint (an outline) on map of moderate scale. The refinery's footprint, other than being shown visually on the map, is recorded numerically as a string of points where the string is closed to form an area. The feature is displayed as continuous line (or footprint) which has the same shape and dimensions as the actual feature. Area features have attributes too, such as their name (e.g. "Exxon Refinery Number 22"), its postal address (e.g. "1500 Industrial Parkway, Baltimore, MD, 21010"), and the volume of emissions (e.g. "1.2 tons general hydrocarbons"). A second example would be a Census Tract. Beyond recording its boundary, the Census Tract may have attributes such as the total population (e.g. 12,895 persons) and their average per capita income (e.g. $12,425 per person per year).

In summary, there are three geographic feature types that are represented on a map (all maps represent some combination of point, line, and area features as they occur in the real world). We have presented the concept that these features have associated attributes (perhaps better remembered as their "characteristics") and that a Geographic Information System is used to store coordinate data on the location of the features and information on their associated attributes in a digital format.

5.4.2 EPA's GIS Platforms for Conducting EJ Identification Analyses

EJ analyses frequently rely on Geographic Information System (GIS) tools to manage, analyze and display data. EPA users have performed analyses using a variety of such tools, or "platforms," predominantly ARC/INFO, ArcView, and LandView. EPA has a large installed-base of ARC/INFO GIS software and is in the process of identifying requirements for procurement of a "desktop" (e.g., PC/Client-Server technology) spatial analysis and display capability; the MapInfo GIS being an example, ArcView GIS another. Statistical analysis and visual display software tools such as SAS can also be used effectively in EJ analyses.

In the performance of an EJ analyses it is more important to use accurate and documented data and apply accepted methodologies than to use a particular GIS platform. The existing choices of GIS software offer significant tradeoffs in terms of functionality and price. Under most circumstances, any of the applications mentioned could be undertaken on the available GIS platforms. Also available are existing EPA applications such as the OPPT Graphical Exposure Modeling System (GEMS) or the OAQPS HEM model; or emerging applications such as the HUD Community Planning Software Plus (CPS+), which is based on the Mapitude commercial software package.

The use of a GIS requires significant investment in infrastructure such as software licenses, data format conversion activities and storage, training, and on-call technical support staff. The available capabilities and functionality vary significantly in cost, power, ease of use, and data availability. Most commercially available Geographic Information Systems provide a programming interface with which to develop "user friendly" customized applications which less specialized users can employ in performing analyses and creating map displays. The ARC/INFO GIS has been used to create an enormous amount of EPA-related data, including relevant Census information as well as various regulated facility records. It can directly link to systems such as ENVIROFACTS, an Oracle database that contains program system data updated monthly.

GIS mapping servers provide an extension of GIS tools to a user community. They allow the publishing of geographic databases and mapping applications on the Internet (for public, and through a private network (Intranet). These software tools allows integrated access to geographic and related attribute information stored in a wide-variety of data formats. This type of software, and especially future versions, will allow EPA to develop an Internet accessible GIS tool specifically for EJ analyses. Such tools will allow public access to query, visualization, and analytical capabilities and be organized and made operational in a manner consistent with Agency defined methodologies and policies. In addition, it will often eliminate the need to download data sets and for individual users to their own GIS software packages. EPA expects to put more emphasis on providing

access to customized applications via the Internet for the purposes of maximizing access, resource sharing, and to reduce duplication of effort.

5.5 Defining a Study or Target Area

In defining a study or target area, an analyst must determine the geographic "scope" of the overall project--specify which areas are within the scope, and which are outside the scope of the analysis. For example, consider a smokestack of a refuse incinerator that is known to release toxic substances into the surrounding atmosphere. The optimal boundary for a study or target area in this example would be the area affected by the smokestack's emissions, including persons residing within its influence (e.g., those exposed to the pollution source under consideration, as well as those affected by the source in some other way). However, determining this boundary is not usually straight forward. It might be dependent on a number of factors including the type of substance emitted and how far and in which direction it moves. For environmental or health effects, the geographic scope appropriate for any particular assessment may depend on a variety of influences including: the type of substance causing the adverse effects; the mode of movement of the substance; and the distance and direction(s) the substances travel, etc.. Other types of effects (e.g., economic, social, religious, etc) may affect a very different group of people.

It is also important to understand how the geographic scope or scale of the study or target area may influence associated demographic characteristics. For example, by changing a study or target area, a "minority group" could change dramatically in its proportion to the whole population, or appear, or disappear. Geographers refer to this as the *modifiable areal unit problem* because of the mathematical and general technical complexity of the issue and, often, lack of a satisfactory, straightforward solution. For example, consider the development of political boundary systems as they pertain to assembling a geographic voting district system that gives unfair advantage to one party over another in an election. This process, termed "gerrymandering", is an example of how *bias* is used deliberately to favor one population group over another. Such bias should be avoided in laying the boundary of the EJ study or target area.

Depending on the type of assessment, the procedural approach taken, and the available data, a study or target area may be defined. For screening or targeting assessments, this may involve using pre-defined areas, such as county or state boundaries, to define the study area.

For site specific assessments the process of selecting the geographic boundaries for a study area is more complex. Effect areas or communities have been most commonly defined as areas within a certain distance from an adverse environmental condition or facility. As such, the source of the effect(s) being evaluated has to be given a location. A

number of alternatives have been used to locate these sources, some of which make reference to Census data units.

One approach to locating a source of adverse environmental effect is to use a single point of reference on a GIS (or point feature as described in section 5.4.1), recorded numerically as a pair of coordinates, i.e., the latitude and longitude of a facility's center when viewed from above. The single point may be referenced at a variety of locations, e.g., at the entrance of a facility, at a stack, a building, or at a tank. The "best" point of reference for a source of effects is not easily determined, although a point in the approximate center of the activities which potentially result in effects is usually preferred. For locating very large sites or sources, and those with multiple, widely-distributed environmental release points, using a single point of reference may not be appropriate. In studies such as these, analysts should consider other methods of locating sources of effects, such as using the entire perimeter of a facility or some other way of determining the extent of the effect.

Advantages of using the locational methods defined above are that they give the most specific location identification available. However, while location data is becoming more accurate and accessible with the use of Geographic Positioning Systems (GPS), it is often still unavailable, and of inadequate quality, or of known quality.

Another approach used in a number of EJ studies is to use a Census enumeration unit that contains a facility or site as the defining location. A frequently used unit is the Census Tract (Anderton, et al., 1994; Cutter, 1996). Other studies have used the Block Group (Cutter, 1996; Perlin, et al., 1995); still others, ZIP Codes (Lavelle and Coyle, 1994; Hamilton, 1993 and 1995; United Church of Christ, 1987); Municipality and Township units (Greenberg, 1993; Zimmerman, 1993), or county units (Perlin, et al., 1995; Nieves, 1992; Hird, 1993).

Advantages of this locating method are that the use of enumeration units avoids the problems associated with acquiring location data, assuming you can accurately place a source in the correct enumeration unit. In addition, data for the selected enumeration unit is usually readily available.

A drawback to this approach is that, if the source of an effect is not located close to the center of the enumeration unit, but closer to the perimeter or one corner, considerable error may be introduced into the analytical process. In addition, the characteristics of the enumeration unit are usually taken as the characteristics of the site. This may be too gross a characterization, and this problem increases with the size of the enumeration unit.

Once the source of the effects are located, the outer boundary of a study or target area may be determined. In a GIS, this outer boundary would be referred to as a "buffer". A buffer surrounds the point, line or areal features that are the source(s) of the effect. The

spatial coverage and shape of the buffer depends on the a number of factors including the nature and characteristics of the source; the transfer mechanism or agent; and impact of distance on the effect. For point features, such as a smokestack, a simple buffer would be circle-shaped. A more complex buffer for that same smokestack, might include other factors of influence as mentioned above, and require an alternate shape, e.g., plume-shaped with the long axis in the prevailing wind direction. For linear source features, such as a highway, the buffer would be corridor-shaped. For areal source features, such as a refinery, the buffer would encompass the perimeter of the facility.

The most common buffer used by EJ analysts to date is a circle-shaped buffer. These buffers have ranged from a half of a mile to four miles in radius. Usually, the selection of the buffer radius is arbitrary, having little to do with the nature of the effects being studied. In some cases, smaller radii have been used to approximate the exposure to fugative emissions (e.g., emissions from tanks, leaky pumps, etc.) and larger radii, 2-4 miles, have been used when studying stack emissions. Rarely have there been attempts to determine buffer distances through the use of modeling or actual measurement.

When the buffer is arbitrarily selected, it is a good idea to conduct a sensitivity analysis by using a number of alternative distances and seeing if the demographics change significantly. This will give an indication of how the characteristics of an area change using alternative distances from a source of contamination.

Another drawback of this approach is that effects may not occur directly around a site but much further away. For example, an affected area may be one which is receiving water from contaminated sources many hundreds of miles away or contaminated fish eaten even further away. Similarly, an affected area may be a number of miles downwind from an airborne source of contamination where the immediately proximate areas are unaffected.

Another problem with this approach is that, depending on the transfer mechanism, nature of the source or effect, the effects may not be dispersed uniformly from the source. For example, air and water pollutants may move in a variety of directions due to variations in groundwater flow and air currents; truck traffic takes a single route; etc.

5.6 Defining a Reference Area

A *reference area* is used as a standard of comparison when determining whether a target area suffers from disproportionate effects as a minority and/or low-income populations. A reference area therefore provides a context for the interpretation of data from the target area. The type of assessment that is being performed will help define an appropriate reference area.

Targeting/screening assessments are usually done to differentiate amongst a number of sub-areas within a larger area. For example, screening of counties within a state on the basis of a number of factors such as race, income, and proximity to permitted facilities. Each county being screened would have its individual "score" which is then compared to the overall score for the state as a whole (e.g., state average). For these assessments, the reference area is the starting area which is being subdivided (e.g., the state).

For site specific assessments, the selection of a reference area is not necessarily straightforward. On a case-by-case basis the analyst must try to determine what reference area(s) makes sense.

If the site being assessed is self-identified, the community may have some ideas about a reference area's definition. As an example, a community claims that a waste transfer facility siting is creating an EJ issue. The facility is designed to service a limited area - the surrounding county. The community claims that the area proximate to the proposed site has a much higher African-American population than the rest of the county and that the facility could have been placed anywhere in the county. In this case the county seems a logical choice as the reference area, since it is the area serviced by this facility.

In some site specific assessments, the permit granting entity's area of influence may have some bearing on selecting an appropriate reference area. For example, if the state issuing a permit could approve it anywhere in that state, then the entire state might be an appropriate reference area.

For some assessments, certain data characterizing the reference area may not be available. As such, it may be necessary to estimate those characteristics by randomly selecting a sample of the reference area. For example, consider a target area defined as a three mile radius surrounding a proposed facility. The analyst wants to know how the cumulative effect of green particles emitted from the proposed facility and two existing facilities within the target area compares with the rest of the reference area, the entire state. Since effect data for the entire state do not exist, and developing it would be too costly, the analyst must randomly select a set of areas to represent the reference area. A set of X random points (latitude and longitude coordinates) is then selected from within the state and a three mile radius drawn around each point. The number of green particles emitted in each of the randomly selected areas is then determined. The distribution of the effects of the random areas gives a statistical representation of the reference area.

When selecting random areas for inclusion in the estimate of the reference area, it is important to keep in mind that the reference area represents a range of places that the proposed facility could have been sited. As such, an analyst may want to consider reasons for excluding a particular area, if that area can not feasibly support the proposed site. The following are a set of proposals for dealing with these issues:

- *Existing zoning* should not be the basis for excluding areas from the reference area. That is because the nexus of the EJ problem may in fact lie in the zoning process. Bias caused by the zoning process is part of what we are trying to discover when we do these analyses. We do not have to show that any sub-component of siting a facility was responsible for causing the disproportionate risk.

- *Economic factors* should not be used to exclude areas from the reference area. There is no inherent right that allows any given facility to be built. One possible option is *not* to build the facility. If economic considerations or other siting criteria results in a site selection that will cause a disproportionate risk, the fact that this was the best or only site meeting some criteria does not mitigate the disproportionate risk.

- *Areas where the facility could not be physically built* (e.g., large lakes, and parkland) could be excluded from the sample.

In general, reference areas determined for an EJ study should be the same for calculating both the effects and the minority and/or low-income community presence. If an analyst were to pick and choose which communities to use as reference for the different parts of the analysis, then the validity of the analysis would be greatly diminished. Thus, an objective approach is needed, and that requires the use of a consistent set of reference areas for a given EJ determination.

5.7 Aggregating Data for a Target or Reference Area

A first step in the assessment of demographic characteristics or adverse effects of both the target and reference areas, is to aggregate available data to a geographic level appropriate for those areas.

Depending on what type of assessment is being conducted and how the associated target or reference area is defined, this step may or may not be necessary. For screening or targeting assessments, the boundaries for the target or reference areas are often derived from and equivalent to the boundaries for Census enumeration units. As such, the data are readily available at the appropriate geographic level and no data manipulation is necessary. For site specific assessments, the boundaries of the target or reference areas do not necessarily conform to Census enumeration unit boundaries. In this case, the data may have to be aggregated to the geographic level of the target or reference area (original survey data is the exception).

Census data are available for a variety of enumeration units: blocks, block groups, tracts, zip codes, municipality designations (various), counties, states, and various sub-state and

multi-state regions. On one hand, smaller Census enumeration units produce more accurate results, especially when the data need to be aggregated for a target or reference boundary that does not conform to a Census enumeration unit. On the other hand, the larger the Census enumeration unit, the more demographic parameters are available. For example, block data does not give income, necessitating the use of house value and rent.

When defining a target or reference area by a buffer around a specific site, all Census enumeration units *completely within* the buffer should be aggregated in generating a total population estimate. Those Census enumeration units lying *partly within* the buffer may be completely or partially included in the total population estimate, depending on the specific GIS method/application used. In general, two methods exist for handling Census enumeration units that lie on both sides of the buffer or target/reference boundary. The first, the *proportionate area method*, includes partial enumeration units in the total population estimate, measured in proportion to the fraction of their land area falling inside the buffer. The second, called the *centroid method*, includes the entire Census enumeration unit in the total population count, if the center of the enumeration unit (centroid) is within the buffer. The particular method employed can have a significant effect on the total population estimate. In general, if the target or reference area is small, the proportionate area method is the more accurate of the two. For larger target or reference areas, both the proportionate area and centroid methods achieve equivalent results.

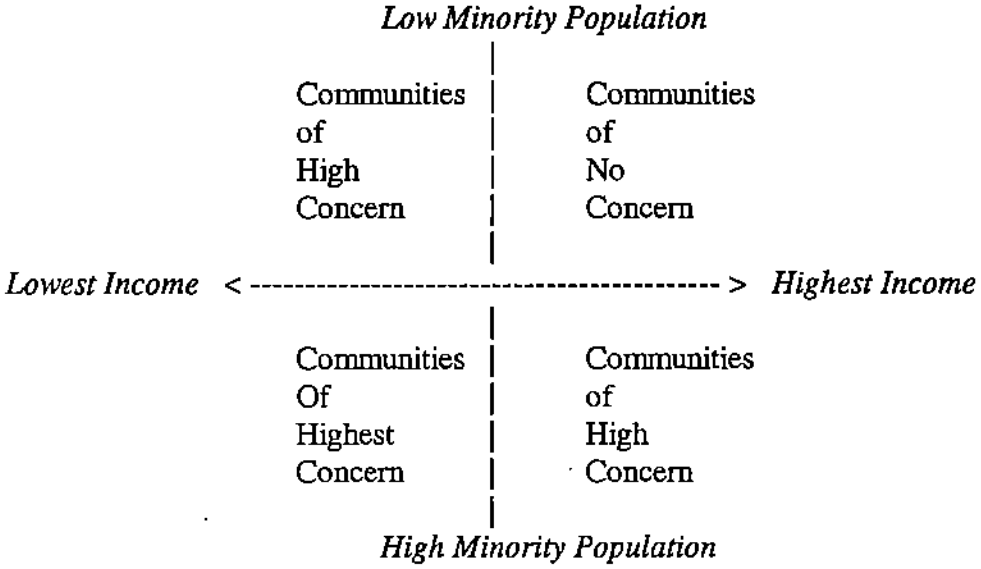
5.8 Determining Potential EJ Areas of Concern

As described in the Concepts and Terminology section, a *potential EJ area of concern* is a target area that contains a *significant* minority and/or low-income population but the existence of disproportionate effects has not been shown. In an earlier section we considered definitions of the terms "minority" and "low-income". These definitions are distinct from each other. A statistician would refer to them as "independent variables". A *variable* is a factor that may be measured, such as the proportion of a community's minority representation to a larger general population, or the fraction of a community with an income below that of the larger general population. They are *independent* variables because the value of one is not necessarily linked to the value of the other. In this context this means that some minority communities may enjoy very high levels of income, while lower income communities may not have minority representation in a given region.

In Figure 5, we represent the interactions between, and states of, the two variables as they relate to our task. On the horizontal scale we represent the income variable where we have the lowest income communities on the left, moderate in the center, and higher income communities to the right. On the vertical scale we represent the concept of a

community with low minority representation at the top of the axis and a community with high minority representation at the bottom of the axis.

Figure 5: A Bi-Variate Approach to the Identification of Target Communities



Those communities which fall into the upper right hand quadrant are of no concern with respect to the task assigned to us. They are neither lower income or minority communities. Those communities that fall into the lower right hand quadrant, or the upper left quadrant, are of concern with respect to the task assigned to us. They are, respectively, lower income (only) or minority (only) communities. Those communities that fall into the lower left hand quadrant are of great concern. They are both lower income *and* minority communities.

5.8.1 Methods for Determining Minority Population Significance

There are several methods available for assessing whether or not a target area contains a *significant* minority (as defined in section 4.2.3.3) population. The methods addressed in this guidance are: absolute thresholds, relative thresholds and a comparison approach. Another method for establishing whether or not a target area has a minority population--not necessarily significant--is ranking.

An analyst using a threshold method (absolute or relative), typically determines the percentage of minorities within the target areas, compares those percentages to the threshold selected, and then classifies the target population as a potential EJ area of concern (if the target population is equal to or greater that the threshold selected) or not (if the target population is less than the threshold selected).

An absolute threshold is a fixed percentage used as a benchmark in determining whether or not a target area has a significant minority population, e.g., 50% or the national average. The drawback to using an absolute threshold is that no *one* absolute threshold can accommodate all target areas. For example, consider a typical absolute threshold used in EJ assessments; the 50 percent minority cut-off. Its use is very common among EJ analysts, and in fact, the Bureau of the Census publishes a list of counties with 50 percent or more minority populations. However, there are difficulties inherent in using this particular threshold. Most neighborhoods in US urban areas, and many areas in the southern US are over 50 percent minority. The entire Commonwealth of Puerto Rico is over 50 percent Hispanic. Using the 50 percent absolute threshold for these areas would mean that all of those target areas would be identified as EJ areas of concern as long as their environmental effect is disproportionate. Hence, this particular threshold should not be used to determine whether or not a target area is a potential area of concern in these areas. In general, it is up to the analyst to ensure that the absolute threshold chosen makes logical sense, within the context of the particular EJ assessment.

A relative threshold is the percent minority population derived from the reference area, used as a benchmark in determining whether or not a target area has a significant minority population, e.g., the state average. It is considered *relative* because the threshold is derived from the same geographic area as the target or study area, such as a region, state, or county. As an example, consider the task of determining areas that have a significantly higher percentage of minorities than elsewhere within a particular Region. If an analyst were using the national average as a threshold to determine these areas, some areas in need might never be selected. This is because many states consistently have higher or lower percent minority than the national average. For example, California, as well as many southern states, have higher minority percentages than the national average, whereas other states, such as Idaho, have lower minority percentages. The use of a relative threshold in this example, such as state averages, would enable more meaningful determinations. The Region VI EJ methodology uses this type of approach, focusing on potentially affected areas with minority populations twice to three times the state average. Instead of using a single national figure to determine minority areas, this method takes into account regional or state differences.

A comparison approach compares the percent minority population of a target area with the percent minority population in a reference area. If the target area has a significantly higher percent minority population than the reference area it is considered a potential EJ area of concern. Two options exist for determining the significance of the difference between the two populations; a fixed percentage cut-off and/or statistical significance testing. For an explanation of statistical significance testing methods, options, and examples see section 5.8.5. Fixed percentage cut-offs are similar to absolute thresholds. That is, if the percent minority population of a target area is greater than the percent minority population of a reference area, and the difference is greater than the fixed percentage cut-off, it is considered a potential EJ area of concern. For example, Region

II, in their draft interim guide[3], use a fixed percentage cut-off of 25%. Therefore, the percent minority population of the target area must be 25% higher than the percent minority population of the reference area in order for the difference to be considered significant.

The ranking method, as the name implies, ranks the target areas by minority composition and gives priority to areas with the greatest minority populations. Two variations of this approach exist. The first one ranks a series of scored categories, with ranges of 0-25 percent, 25-50 percent, 50-75 percent and 75-100 percent minority, for the purpose of identifying those areas with high minority populations. The major disadvantage to using this approach is its tendency to over simplify the data, making meaningful interpretation very difficult, i.e., the number of possible scores resulting in the same relative rank is very large. The second variation of this is approach, continuum ranking, ranks the distribution of demographic counts (minority populations) for all target areas from lowest to highest. This approach has the advantage providing a complete description of all of the data, while at the same time selecting target areas that fall at the upper end of the distribution. However, this approach does not assign a level of significance to the upper end of the distribution. Thus, for purposes of establishing whether or not a target area has a significant minority population, an additional step to compare the upper end of the distribution to an absolute threshold level is needed.

Both the Interagency Working Group Guidance[4] and the Council on Environmental Quality NEPA Guidance document (draft, July 1997) state that a minority population of a study area should be defined as areas with *either* over 50 percent minority, *or* with a "meaningfully" greater percent minority than a reference area. Both documents take both an absolute and a relative threshold approach to defining a minority community.

5.8.2 Recommendations for Determining Minority Population Significance

In choosing a method for determining the significance of minority populations in a target area, an analyst should consider the type of assessment being conducted and its allowable level of uncertainty. For targeting and screening assessments, absolute thresholds, relative thresholds, and ranking methods are recommended. These methods require little lead time, few resources, and provide an adequate means to

---

3 Draft Implementation Guide to the Interim US EPA Region 2 Policy on Identifying Environmental Justice Areas, June 1997, IPS/Region 2 EJ Workgroup.

4 Guidance for Federal Agencies on Key Terms in Executive Order 12898, developed by the Interagency Working Group on Environmental Justice, August 1995. This IWG is a task force created by the Executive Order and is comprised of the heads (or representatives) of 17 departments and agencies.

quantify minority composition for the internal decision-making activities that they typically support. The specific thresholds recommended in this guidance are: 50% or the national average (25%) for absolute thresholds and the state average plus 20% (state average times 1.2) for relative thresholds.[5] Figure 6 below is a list of recommended minority thresholds (state average times 1.2) by state derived from the Census STF3 data.

Figure 6: Recommended Relative Minority Thresholds

State	Minority Threshold		State	Minority Threshold
Alabama	32.10%		Montana	9.76%
Alaska	31.23%		Nebraska	8.87%
Arizona	33.83%		Nevada	25.40%
Arkansas	21.34%		New Hampshire	3.24%
California	51.07%		New Jersey	30.98%
Colorado	22.92%		New Mexico	59.34%
Connecticut	19.17%		New York	36.69%
Delaware	24.70%		North Carolina	29.89%
District of Columbia	87.13%		North Dakota	6.83%
Florida	31.99%		Ohio	15.43%
Georgia	35.72%		Oklahoma	22.67%
Hawaii	82.34%		Oregon	11.05%
Idaho	9.25%		Pennsylvania	14.65%
Illinois	30.02%		Rhode	12.60%
Indiana	12.44%		South Carolina	37.68%
Iowa	4.73%		South Dakota	10.58%
Kansas	13.74%		Tennessee	20.89%
Kentucky	9.95%		Texas	47.09%
Louisiana	41.05%		Utah	10.43%
Maine	2.49%		Vermont	2.41%
Maryland	36.43%		Virginia	28.79%
Massachusetts	14.45%		Washington	15.72%
Michigan	21.10%		West Virginia	4.93%
Minnesota	7.38%		Wisconsin	10.37%
Mississippi	44.30%		Wyoming	10.63%
Missouri	15.65%			

For site-specific assessments, both the threshold and comparison methods are recommended, with the comparison method considered the more rigorous of the

5 The recommendation of 20% is a rule of thumb used in courts as a reasonable approximation in absence of a statistical analysis.

approaches. When using a comparison approach, an analyst should choose the method to determine the significance of the minority population, based on the acceptable level of uncertainty. For assessments with higher levels of allowable uncertainty, a fixed percentage cut-off would be appropriate. For assessments with low levels of uncertainty, which require more rigorous and defensible methods, an analyst should use a fixed percentage cut-off as a means of screening for meaningful differences, then employ statistical significance testing. Fixed percentage cut-offs are often used as a screening measure first, since statistical methods employed alone can overestimate the significance of the difference between two populations, especially when the study area is small. The specific percentage cut-off recommended in this guidance is 20%.[5]

5.8.3 Methods for Determining Low-Income Population Significance

In assessing whether or not a target area contains a *significant* low-income (as defined in section 4.2.3.10) population, the methods detailed in section 5.8.1 for minority populations are applicable; absolute thresholds, relative thresholds, and the comparison approach. Similarly, the ranking methods, as outlined in section 5.8.1, are used in targeting/screening assessments.

In summary, the threshold method compares the percent low-income population of a target area with either an absolute or relative threshold selected by the analyst. If the target area percent low-income population is greater than or equal to the threshold selected, the target population is classified as a potential EJ area of concern. The comparison approach compares the percent low-income population of a target area with the percent low-income population in a reference area, and determines the significance of the difference between the two populations through a fixed percentage cut-off and/or statistical significance testing (see section 5.8.5). If the percent low-income population of a target area is greater than that of the reference area, and the difference is greater than the fixed percentage cut-off, it is considered a potential EJ area of concern. Finally, the ranking method, as the name implies, ranks the target areas by low-income composition and gives priority to areas with the greatest low-income populations.

5.8.4 Recommendations for Determining Low-Income Population Significance

In choosing a method for determining the significance of low-income populations in a target area, an analyst should consider the type of assessment being conducted, the relevant definition of low-income as defined in section 4.2.3.10, and the allowable level of uncertainty. Absolute thresholds, relative thresholds, and ranking methods are recommended for targeting and screening assessments, while threshold and comparison methods are recommended for site-specific assessments. The specific

absolute and relative thresholds recommended in this guidance are presented in Figure 7 below.

Figure 7: Recommended Thresholds for Determining a Low-Income Population			
Data Source	**Poverty Thresholds**	**Income Ranges in STF3A**	**Qualified Census Tracts**
Low and Very Low-Income Definition Benchmarks	Below poverty threshold or two times the poverty threshold	Below $15,000 or $25,000	Below 60% of AMGI values (imbedded)-- very low-income N/A
Absolute Threshold (Moderate)	20% or more of the population below benchmark	20% or more of the population below benchmark	50% of households below benchmark (imbedded)
Absolute Threshold (Extreme)	40% or more of the population below benchmark	40% or more of the population below benchmark	50% of households below benchmark (imbedded)
Relative Threshold	Percent of persons below benchmark is greater than or equal to the related state percent	Percent of households below benchmark is greater than or equal to the related state percent	N/A

The absolute thresholds presented in Figure 7 pertain to both low-income and very low-income definitions. (Note that the threshold for the QCT benchmark is imbedded, or already applied, and needs no further manipulation.) The absolute thresholds of 20% and 40% were derived and are used by the Census Bureau for defining poverty areas and extreme poverty areas.

Relative thresholds presented in Figure 7 should be used for low-income and very low-income definitions. (Again, QCTs have imbedded or already applied thresholds and need no further manipulation.) The specific relative threshold recommended in this document is the state percent of persons or households falling below a selected benchmark. That is, if the percent of persons or households with incomes below the selected benchmark for a target area is greater than or equal to the state percent of persons or households with income below the selected benchmark, the target area is considered a potential EJ area of

concern. **Figure 8 is a listing of the recommended relative low-income thresholds by state derived from the Census STF3 data.**

Figure 8: Recommended Relative Low-Income Thresholds

State	15K Threshold	25K Threshold	Poverty Threshold	2*Poverty Threshold
Alabama	33.13%	52.27%	18.34%	40.20%
Alaska	14.71%	27.97%	9.00%	23.78%
Arizona	25.45%	45.10%	15.74%	35.78%
Arkansas	36.00%	57.78%	19.07%	44.40%
California	18.94%	34.10%	12.51%	30.12%
Colorado	22.78%	41.27%	11.68%	29.25%
Connecticut	15.23%	27.45%	6.82%	16.28%
Delaware	18.12%	33.94%	8.71%	23.00%
District of Columbia	23.68%	41.02%	16.87%	32.42%
Florida	25.01%	45.11%	12.69%	31.99%
Georgia	25.45%	43.13%	14.65%	33.29%
Hawaii	14.87%	29.82%	8.25%	22.66%
Idaho	27.23%	49.46%	13.25%	38.82%
Illinois	22.06%	38.26%	11.91%	27.05%
Indiana	23.83%	43.09%	10.68%	29.00%
Iowa	26.72%	47.49%	11.48%	31.73%
Kansas	25.71%	45.54%	11.48%	31.24%
Kentucky	34.61%	54.17%	19.03%	41.09%
Louisiana	36.26%	55.12%	23.58%	45.69%
Maine	24.87%	44.58%	10.80%	30.48%
Maryland	15.49%	28.99%	8.27%	20.05%
Massachusetts	19.92%	33.28%	8.93%	21.02%
Michigan	24.16%	40.57%	13.12%	28.94%
Minnesota	22.40%	39.93%	10.22%	26.53%
Mississippi	39.24%	58.94%	25.21%	49.76%
Missouri	28.06%	47.44%	13.34%	33.09%
Montana	32.11%	53.86%	16.07%	39.86%
Nebraska	26.58%	47.89%	11.14%	32.67%
Nevada	20.46%	39.14%	10.15%	27.99%
New Hampshire	16.73%	31.76%	6.42%	19.27%
New Jersey	16.20%	28.76%	7.58%	18.55%
New Mexico	31.15%	51.61%	20.61%	44.15%
New York	23.01%	38.07%	13.03%	27.86%
North Carolina	27.00%	46.76%	12.97%	33.22%

State	15K Threshold	25K Threshold	Poverty Threshold	2*Poverty Threshold
North Dakota	31.10%	53.42%	14.38%	37.82%
Ohio	25.10%	43.47%	12.54%	29.52%
Oklahoma	32.00%	52.51%	16.71%	39.38%
Oregon	25.44%	45.59%	12.42%	32.21%
Pennsylvania	24.77%	42.99%	11.13%	28.19%
Rhode	22.70%	38.54%	9.61%	24.29%
South Carolina	28.26%	47.56%	15.37%	36.86%
South Dakota	32.05%	54.96%	15.86%	40.91%
Tennessee	30.49%	50.33%	15.70%	36.96%
Texas	27.56%	46.32%	18.10%	38.96%
Utah	21.88%	41.36%	11.36%	34.63%
Vermont	22.41%	41.16%	9.86%	28.19%
Virginia	20.14%	36.45%	10.25%	26.22%
Washington	21.39%	39.26%	10.92%	27.53%
West Virginia	37.29%	57.81%	19.66%	43.27%
Wisconsin	23.37%	42.08%	10.70%	28.14%
Wyoming	26.33%	45.96%	11.86%	33.35%

Clearly, the more analytically defensible method is the comparison approach. When using a comparison approach, an analyst should choose either a fixed percentage cut-off or a combination of the fixed percentage cut-off and statistical testing to determine the significance of the low-income population, depending on the acceptable level of uncertainty. As mentioned previously, for assessments with higher levels of uncertainty a fixed percentage cut-off is appropriate. For low levels of allowable uncertainty, which require more rigorous and defensible methods, fixed percentage cut-offs applied to screen for practical differences followed by statistical significance testing is recommended. The specific percentage cut-off recommended in this guidance is 20%.[5]

5.8.5 Statistical Methods for Determining Minority or Low-Income Population Significance

The objective of this section is to provide credible analytical procedures to evaluate the minority and/or low-income category status in a study. This is particularly important in studies requiring a low degree of uncertainty, e.g., site-specific permitting decisions. The immediate intention is to develop an introductory discussion of statistical procedures available to EPA staff.

The initial step in a statistical analysis is to provide descriptive statistical measures to describe the demographic (minority category levels and low-income levels) in the reference and target areas. The average or mean, median and first and third quartiles should be provided for the low-income and minority category demographic variables for the reference and target areas. In addition to tabular presentation of the results, the results may be demonstrated graphically via histograms, statistical maps, polygons, Travis charts, "Starry Night" charts, image charts, qq charts, among others.

If the descriptive statistics demonstrate the possibility of excessive numbers of minority status and or low-income persons in the target areas, the next step is to use statistical tests of significance to verify the possibility. Tests of significance are statistical procedures that will be used to demonstrate whether or not the apparent differences have a high probability of being authentic (i.e., there is a low likelihood that the target area has a relatively high minority and/or low-income population by chance alone).

Tests of significance require the development of null and alternative hypotheses. In the case of Environmental Justice situations, the null hypothesis is that the observed percent low-income and/or minority category status are at least approximately equal to those of the reference area. The alternative hypothesis is that the observed percent minority category and/or low-income is significantly higher in the target area.

The probability of being incorrect often is called alpha (a Type I error) or the probability of incorrectly claiming that the null hypothesis is incorrect (a false positive conclusion). The alpha probability is the probability that the observed p probability has to be less than or equal to prior to not supporting the null hypothesis. This probability is designated prior to the computation of the statistical significance test. Common alpha probabilities are .01 through .20. The choice should depend upon the environmental, political, and programmatic consequences of a Type I error. The use of the procedure is to assess whether the minority status and/or low-income status of the target area differ greatly from the reference area.

There are two basic types of statistical significance tests: parametric and non-parametric procedures. Parametric tests require known (often normal) continuous assumed data probability distributions. Non-parametric tests do not require data distribution assumptions as stringent as those required for parametric procedures. Most importantly for environmental justice analyses an approximately normal or log-normal data probability distribution cannot be assumed. Thus, the distribution of the data must be assessed prior to applying a test of statistical significance. Several graphical and computational procedures are available. Commonly applied graphical procedures to assess for normality include probability plots, histograms, and stem-and-leaf plots. Commonly applied computational procedures include Kolmogorov-Smirnov test, the Shapiro-Wilk test, the Filliben statistic, range test, and coefficient of kurtosis and skewness tests. Specifics on how to compute the statistics are available in the "Guidance

for Data Quality Assessment: Practical Methods for Data Analysis." EPA QA/G-9, QA96 Version, EPA/600/R-96/084, Office of Research and Development, Washington, DC, 1996.

There are several tests of statistical significance available depending on whether the data are at least approximately normally distributed. All tests of significance should be done separately for proportion minority status and proportion low-income. If the data are approximately normally distributed, the recommended procedure is the T test with proportions. The essence of the test is to assess whether or not the differences between the target area and the selected reference area (in terms of proportion minority category status and proportion low-income) are large enough relative to the measure of dispersion or standard deviation. If they are large enough, the analyst is able to claim with one minus the pre-selected alpha probability that the differences are authentic.

If the data are not approximately normally distributed, there are several procedures that are applicable. The chi-squared test is an alternative procedure that uses contingency data to compare the differences between the distribution of the two demographic characteristics (within blocks or block groups of the target area) and the distribution in the reference area. The differences must be large enough to conclude, with at least a one minus alpha probability, that the inconsistencies are not due to chance. It should be noted that the chi-square test is applicable regardless of whether the data are disaggregated or not for the target area. In other words, the data can be essentially a two by two contingency table.

If the data from the target area are aggregated rather than disaggregated, the Fisher's Exact Test may be utilized to assess the differences between the reference and target area for the demographic characteristics. The test procedure provides an actual estimate of the probability of a Type I error. The procedure is approximately equivalent to a chi-square test for a two by two contingency table assuming the chi-square test had a correction for continuity.

5.8.6 Examples of Statistical Methods Employed

The purpose of this section is to illustrate the utility of several of the statistical procedures previously discussed. The exemplary analysis is designed to demonstrate the determination of whether or not a target area is a potential EJ area of concern. The exemplary case is derived from 1990 US Census data provided for block groups. The analysis compares the block groups per cent low-income and per cent minority group membership within four contiguous Illinois counties with the analogous descriptive statistics for the remainder of the state. Consequently, the initial step was to calculate the descriptive statistics for each of the areas being compared for per cent low-income and per cent minority group. The statistical results in this discussion were computed using the SAS software package.

The most significant descriptive statistics for each of the areas were the mean or average, median, first and third quartiles and the standard deviation. The average and median is included to provide an indication of central tendency among the block groups within the comparison geographical areas. The geometric means are provided because the raw data may not adhere to an approximately normal distribution, but rather the log base 10 of the data log-normally distributed. Parametric tests of statistical significance may be used with the log transformed data. In addition, if the T test is used as the test of statistical significance to compare the areas, the means for the comparison areas are compared. Furthermore, the arithmetic and geometric standard deviations and the first and third quartiles are provided to show dispersion within the areas. Also, information about the dispersion is used in significance tests to determine whether or not the differences between areas are authentic rather than chance data fluctuations. The descriptive statistics are shown below:

Area	Per cent low-income	Per cent minority Remainder of Illinois
# Block Groups	10,550	10,550
Mean	27.87	24.08
Median	23.70	7.13
Standard Deviation	21.17	33.29
First Quartile	10.80	1.41
Third Quartile	40.40	31.28
Geometric Mean	1.87	1.81
G. Standard Deviation	0.93	1.41

Four Contiguous Counties

# Block Groups	246	246
Mean	34.03	32.25
Median	30.45	8.38
Standard Deviation	22.21	29.58
First Quartile	18.20	20.56
Third Quartile	44.80	35.67
Geometric Mean	2.11	2.04
G. Standard Deviation	1.02	0.98

Because the descriptive statistics are indicative of a possible difference between the two areas to be compared (for both per cent low-income and per cent minority), the four contiguous counties seem to have higher proportions. However, statistical significance

tests are needed to demonstrate whether or not the differences are authentic. As previously mentioned, there are two basic types of tests of statistical significance: parametric and non-parametric. Therefore, in order to select the appropriate test it is necessary to assess whether or not the block group data have at least an approximately log-normal data distribution. In order to assess the adherence to the normal and log-normal distributions, the Shapiro-Wilk's test and normal and log-normal probability plots were individually utilized for each of the areas for each of the demographic variables. Low values of the test demonstrate deviations from normality. Nonlinear plots demonstrate deviations from normality. The plots are not provided but are available upon request. The Shapiro-Wilk's as well as the graphs consistently reflected substantial deviations from data normality. However, the deviations appear to be less substantial for the log-transformed data. The Shapiro-Wilk's results are shown below:

Remainder of Illinois

Area	Test Statistic Value	P Probabilities
Per cent low-income	0.09	<.01
Per cent minority	0.25	<.01

Four Contiguous Counties

Per cent low-income	0.92	<.01
Per cent minority	0.64	<.01

Log Transformed Data

Remainder of Illinois

Per cent income	0.26	<.01
Per cent minority	0.40	<.01

Four Contiguous Counties

Per cent low-income	0.80	<.01
Per cent minority	0.46	<.01

The results demonstrate substantial deviations from normality. Thus, non-parametric tests are most appropriate. Nevertheless, the results of the mean comparison T tests are shown along with the non-parametric chi square procedure to illustrate the test. The parametric T test procedure was done individually for each variable for the raw and log transformed data. The non-parametric procedure was only utilized with the raw data for each of the variables. The test results uniformly show that the four contiguous counties have higher per cents minority and low-income than the remainder of Illinois. Thus, the area is a potential environmental justice area. This observation is true because the value

of the test statistics consistently indicate a p probability value less than 0.01. The general rules of thumb for the p probability value usage is as follows:

p probability	Decision
0.00- 0.05	Strong evidences to reject the null hypothesis that the areas being compared have equal proportions of the demographic characteristics in question.
0.06 - 0.10	Substantial evidence to reject the null hypothesis.
0.11 - 0.15	Suggestive but probably sufficient evidence to reject the null hypothesis.
0.16 - 0.20	Borderline evidence suggestive of the need for further assessment.

Of course, these procedures to address the demographic potentiality of being an environmental justice area should be followed by environmental calculations to address whether the area is at a disproportionate level of risk and should be an area of EPA focus. Although statistical significance testing methods are not currently used to determine risk and associated effects, there are ongoing efforts to develop and standardize the necessary data and methods for future efforts. At this point in time, the statistical tools imbedded in various risk exposure models may be used to quantify the exposure risk and associated effects, provided that they are applied correctly and their limitations are understood. These models are addressed in the next section.

5.9 Determining Disproportionate Effects

To date, there is no official definition for the terms disproportionate effect or disproportionate risk. For the purposes of this document, disproportionately high effect is defined as: an incidence; prevalence of an effect; risk of an effect; or exposure or environmental hazards causing such adverse effects on a low income or minority population that significantly exceeds that of a comparable reference area or population. In estimating such effects, cumulative and synergistic effects or exposures should be considered, when possible.

The term "disproportionate"denotes a relative or comparative measure and is therefore not any finite value. In the analysis and in the interpretation phase one can consider the disproportionality of a range of types of effects, including human health. Depending on the nature of the analysis, surrogate estimates of risk ranging from proximity to measured exposure may be used. Disproportionate effect or risk will be a somewhat subjective determination based on more than the analytical approach, such as the purpose of the study, the type of effect, and the quantity and quality of availability data.

Surrogate variables (proxy data) may be available and used when the primary measurements on a variable are unavailable. Some EPA regions currently use such values as the overall density of permitted facilities or the presence of fish advisories in an area as a surrogate variable for individual facility or site effects. However, application of surrogates should be coupled with an understanding of the impact of using surrogates on the level of certainty of the result.

The use of various toxicity models is an approach that has recently been used, e.g., OPPTS's TRI Indicator Model and Region III's Chronic Index Model. These models attempt to rank emission data relative to their potential human health effects, based on some relative measure of toxicity or nature of the possible effect. They are not necessarily very good predictors of risk since they don't deal with all of the factors actually affecting levels of exposure to the population of concern. The relative pounds of individual chemical emissions may not be the same ratios at which the population will be exposed, because what happens to the chemicals after emission (e.g., environmental fate and transport) may not be considered.

To try to deal with what happens to the emissions and when and how exposure will occur requires the use of exposure models. Exposure models can be used for assessments in the absence of empirical data, as a supplement to limited data, or to mathematically describe complex exposure situations.

Generally, models transform known inputs or conditions into an exposure/risk analysis using mathematical algorithms created to predict an output for the given scenario.

Although both exposure and toxicity models can be powerful tools for the assessment process, assessors should be careful to abide by a "good modeling practice" approach. This must include consideration of the level of validation as well as other risk characterization issues associated with any model that has been utilized to address an exposure or toxicity scenario. Assessors should include discussions of risk characterization issues during the calculation process and more importantly while interpreting the results of an assessment. The "good modeling practice," essentially that all assessments completed using a model should be mathematically transparent and reproducible by reviewers.

Additionally, the rationale pertaining to how decisions regarding data inputs and the use of empirical data were made should be clear from any assessment. The level of validation of any selected model is a primary risk characterization issue for the Agency. Available information pertaining to the development and validation of the selected model should be included (at least in summary form) in any assessment as this will provide insight into the regulatory importance that should be associated to the assessment. Additionally, a discussion of the appropriateness of the model should also include several other risk characterization issues including: a discussion of the anticipated levels of accuracy and

precision; a discussion of the reliability of the inputs; and a discussion of any other critical issues that should be considered in the interpretation of the assessment.

Assessors should also abide by other general requirements of the exposure assessment process, including compliance with the U.S. EPA Exposure Assessment Guidelines.

Exposure modeling parameters can be divided into two categories: those that pertain to the estimation of the chemical concentration in a localized environment, and those that pertain to the estimation of individual exposure in an environment. Further discussion of the specifics of exposure modeling will not be addressed here, but there are many choices of tools available. These tools range from individual air and water models available to integrated analysis tools such as OPPT's PC-GEMS (Personal Computer - Graphical Exposure Modeling System, June 1995).

GEMS supports exposure and risk assessments by providing access to single medium and multimedia fate and exposure models, physical and chemical properties estimation techniques, statistical analysis, graphics and mapping programs with related data on environments, sources, receptors, and populations. Under development since 1981, GEMS provides an interactive, interface to various models, programs, and data needed for exposure and risk assessments.

The environmental models in GEMS are atmospheric, surface water, land unsaturated (soil) and saturated (groundwater) zones, and multimedia in nature. Methods for estimating octanol-water partition and adsorption coefficients, bioconcentration factor, water solubility, melting and boiling point, vapor pressure, Henry's constant, acid dissociation constant, lake/stream volatilization rate, and atmospheric half-life are available. Data sets are related to environmental characteristics (climate, soil, rivers, groundwater, vegetation), source releases (POTWs and industrial water discharges, Census business patterns, RCRA permit sites), and receptors (population and household estimates for 1970, '80, '90, and '95 by small area census district; drinking water facilities).

Recent analysis of several existing models, however, suggests that there may be significant disparity among models in their ability to predict effects. For these models the disparities are mainly due to differences in the number of chemicals being considered, the estimate of the chemical's toxicity and the inclusion or exclusion of exposure pathways in the analysis. Models may be very useful, however, as long as their correct application and their limitations are understood and considered with respect to the objectives of the study.

The most desirable approach to assessing disproportionate effect would be to use paired epidemiological and environmental exposure studies that measure the amount of pollutants as well as their effect on human health. The problems with this approach are

many, such as time, cost, and the inability to measure effects. An example of a paired study is the Michigan study which examines the presence and effects of PCBs in mother's milk and associated learning disabilities.

5.9.1 Additional Measures of Vulnerability

In estimating disproportionate effects, the Executive Order recommends consideration of cumulative and synergistic effects. In an attempt to capture sub-populations that may be susceptible, several additional characteristics have been incorporated in the Vulnerability Index used in Region III. The Vulnerability Index concept was originated in Region VI and has been further developed by Region III to assist decision-makers in consideration of cumulative and synergistic effects.

Demographic characteristics included in assessing cumulative and synergistic effects may include age, educational attainment, employment status, as well as special attributes of the householder and the household itself. It is generally accepted that the young and the elderly are sub-populations which are more susceptible to health effects than the general population. Although the reasons for susceptibility differ for elderly and children (see Snodgrass, 1992; Calabrese, 1978; Calabrese, 1986). Infants and children have different anatomical and physiological functions than do adults, including a greater relative brain size, brain growth and blood flow, increased breathing rates and an increased surface area to body ratio. Children also have a reduced ability to biochemically detoxify toxicants as compared to adults. Each of these attributes may alter an individual's response to an exposure to toxic materials. While these differences are most pronounced in infants and toddlers, 10 year old children have also been shown to have approximately 40 percent greater air intake than adults (Plunkett et al., 1992).

In particular, children afflicted with Pica are extremely susceptible to certain adverse health effects, e.g., lead poisoning. Pica is an eating disorder in which non-nutritive substances such as dirt, clay, ice, starch, or debris are compulsively consumed. The exact cause of Pica is unknown. This behavior has led to severe lead poisoning, from eating dust and paint chips in older homes. Soil eating behavior continues to be a problem especially around older homes or buildings with lead paint, heavy-traffic areas, or factories where soil may contain dangerous levels of lead or other contaminants.

In addition, the activity patterns of older children include increased outdoor time that may make them more vulnerable to contaminants in outdoor air, soil and surface water as compared to adults. The EPA has performed a preliminary review of the activity pattern studies currently available in the literature and these data indicate that differences in indoor and outdoor activity patterns may range from 30-70 percent depending on the outdoor activity (US EPA 19??). In an effort to distinguish between effects on infants and younger children, the Region III Vulnerability Index permits identification of 0-1 year

old children as well as 1-14 year old children as separate demographic groups. Region VIII defines at-risk age groups as less than five (based on indoor/outdoor activity patterns), and greater than seventy (RBT Meeting, 1997). The Exposures Factors Handbook contains more useful information on activity patterns of children.

It is generally accepted that pregnant women and their unborn fetuses are also sub-populations that are more susceptible to health insults than the general population. This increased vulnerability is due to additional physiological stress on the mother as well as the accelerated growth rate of the fetus (Klaassen et al., 1986). However, because the Census does not provide a category which measures pregnancy, the Region III Vulnerability Index model assumes that if a one year old child is present within the household, a female in the same household was pregnant during the prior year of the Census count. This category does not correct for multiple births or adoptions; however, these errors may only constitute less than two percent of the total count (Forman, 19??, Matthews, 19??).

According to several studies, educational attainment (or, rather, a lack of) might be the single best predictor of vulnerability (see Winkleby et al., 1995; Guralnik et al., 1993; Winkleby et al., 1992; Wells and Horm, 1992; Liberatos, et al., 1998). One plausible reason noted by the authors is that less-educated individuals may be more susceptible to disease due to life-style behaviors, problem solving abilities, and value systems. Winkleby et al. suggest that attaining positive social, psychological and monetary assets and skills may depend on a person's educational level. Attainment of these skills may also influence an individual's range of job choices, limiting them to the more hazardous jobs, where there may be a higher incidence of on-the-job injuries. In this regard, higher educational attainment may provide some protection from adverse influences. To measure educational attainment, the Region III Vulnerability Index uses counts of persons 18 years or older with less than a 12^{th} grade education.

A recent paper published in the American Journal of Public Health (19??) the demographic category "female heads of households with no husband present" and "children younger than 18 years old" were highlighted as having the greatest potential for lead poisoned children (Sargent et al., 1995). In addition, this category (with and without children) is cited by the Bureau of Census as comprising an average of 35 percent of poor people during any month in 1990-1991 (Bureau of the Census, 1994). A further study performed by the Virginia Commonwealth examined the ability of the Sargent model to predict blood lead levels (Fox and Maynard, 1996). This study used actual blood & lead data collected within the state and adjusted the model to improve its predictability. The authors noted that this demographic category, along with minority status and age of housing had a significant influence on the ability of the model to predict blood lead. They retained all three demographic categories in their analysis and the final model demonstrated good ability to predict blood lead levels in the selected communities (a 72 percent correlation). Thus, using the "number of female heads of household, no husband

present with children less than 18 years" variable may assist in identifying those populations which are more vulnerable to environmental affects than the general population, and Region III has incorporated this characteristic in the Vulnerability Index model.

An additional household characteristic for measuring vulnerability is the age of housing. Housing built prior to 1950 has a high probability of containing lead-based paint and HUD has derived a formula for estimating the number of affected housing units (HUD, 1990). Furthermore, ingestion of lead paint chips due has been recognized as one of the leading causes of lead poisoning in children.

Indoor air may also be a useful indicator of vulnerability since it provides some description of the living environment and may offer some additional information regarding potential exposures. Indicators that a household may be more vulnerable to environmental affects resulting from poor indoor air quality may include homes that use coal, coke or wood for heating fuel. The Census provides information regarding household use of coal, coke or wood for home heating and the EPA has recognized the health effects of several of the contaminants which are produced during the combustion of coal, coke or wood. In an effort to improve the quality of indoor air, the EPA has promulgated standards with particular reference to residential wood burning and other home heating combustion units (US EPA, 1991; US EPA 1987; US EPA, 1984).

Based on these considerations, the Region III Vulnerability Index uses counts of households using coal, coke or wood for heating fuel as well as counts of households with the potential for lead-based paint.

The percentage or number of people within a community that rely on subsistence hunting and fishing is also an important indicator of greater potential vulnerability. This group may include Native Americans, as well as poor; both in urban and rural areas. The Executive Order states that Federal agencies should collect data on "consumption patterns of populations who principally rely on fish and/or wildlife for subsistence." Pollutants discharged into the airways and waterways near these populations may pose a greater health risk to those communities. The EPA's Exposure Factors Handbook provides a good source of regional activity pattern and consumption information that may be useful in defining this sub-population.

5.9.2 Recommendation for Measuring Vulnerability in Environmental Justice

Though the Executive Order does not explicitly state that the Agency needs to consider all of the criteria discussed above, these criteria do indicate some degree of vulnerability, and may be important when considering cumulative and synergistic effects. As a result, these indicators may serve as "red flags" which alert us to the potential Environmental

Justice status. Moreover, Environmental Justice communities identified only on the basis of minority status and income, but which also possess these additional attributes may be at even greater risk than other Environmental Justice communities. For example, decision-makers should focus attention on an EJ area that has a significantly higher percentage of young children, or a significantly higher percentage of adults who cannot read.

5.9.3 Integrating Disproportionate Effect Data

This section addresses the integration of different pieces of "environmental" information (e.g., proximity to sites and actual effect single-media data vs. multi-media data) in order to identify potential areas of disproportionate effect or burden. To identify a community with a disproportionately high environmental effect, both cumulative and synergistic effects of exposure should be considered. The types of data currently available may provide information as to the location and type of facility or source, such as data on emissions, and some ambient levels of pollutants. Much of the emission data available are media-specific.

Environmental Justice research has used various approaches toward data integration. However, most of the major studies done to date have used either abandoned hazardous waste sites (CERCLIS) (United Church of Christ, 1987); manufacturers that release toxic chemicals into the environment (TRI) (EPA, 1996d); or treatment, storage, and disposal facilities needing a hazardous waste permit (RCRIS) (US GAO, 1983). One national study used several sources of data including a wide range of manufacturing plants, commercial hazardous waste disposal facilities, Superfund sites, electric generating plants, chemical weapons storage sites, radiation-related research facilities, and disposal and industrial facilities (Nieves, 1992). However, the other studies, which use several categories of facilities including petrochemical plants, incinerators, landfills and other locally, undesired land uses (LULU's) limit their studies to smaller geographic areas like counties or cities. A 1994 study looked at wide variety of sources of pollution including motor vehicle air pollution, sewer overflows, lead exposure, CERLIS sites, incinerators, and hazardous waste generators in the District of Columbia (US EPA, 1996d). None of the major studies have integrated health data or actual (versus potential) exposure to pollutants.

Using large databases with several types of information available (e.g., RCRIS, TRI, or CERCLIS) in combination with Census data allows us to obtain information about a given community's proximity to a site, with many of the caveats mentioned earlier.

An interim Region VIII policy for inspection targeting from a universe of permit holders used Census demographic information and the number of TRI facilities to develop a surrogate for disproportionate burden (US EPA, 1996d). However, these data give us little information other than the facility's manufacturing or waste emissions to assess

environmental effect. To determine communities with disproportionate environmental effects, the analysts made assumptions using the available data. For example, the assumption that a high concentration of a particular type of facility (e.g., incinerators) is a disproportionate risk. TRI data have the potential to be used for broader analyses because they contain quantified releases of specific chemicals. TRI data can be used to assess environmental effect, from at least those sources included in the database (e.g., manufacturing facilities).

Single media studies may be appropriate if the purpose is to target a specific issue or problem. Single-media studies have a narrower focus than multi-media studies and are usually easier to design and obtain quality data. As a result, single-media studies are more common. A problem with using media-specific data is that those communities that have a disproportionate effect due to one specific media (e.g., air) might not be disproportionately affected when all media or routes of exposure are considered (e.g., motor vehicles, waste facilities, indoor air, etc.). The results of single media studies should have the caveat "disproportionate effects found based on comparison of relative TRI emission" for example.

If the EJ study is to support significant issues/decisions and is seeking to determine disproportionate risk, the study should attempt to include all sources of exposure. The information needed to do such a study will not be found in any one standardized database, and tend to be developed as needed. The available data, calculations, and estimations must be researched, developed, analyzed and authenticated by technical specialists who can assess their validity, their associated level of confidence, and their relevance to the situation at hand.

Environmental burden/effect can be related to ambient conditions, a specific source or sources, cumulative or area-wide sources, or uneven application of government authorities. While comprehensive geographically-based data on chemical releases and potential exposures are can difficult or impossible to assemble, valuable information may be generated from currently available data sources (US EPA, 1996c). What is most important is having a clear understanding of the limitations of the data used in the analyses, and how these factors affect the conclusions drawn from the study.

5.10 Consideration of Urban Areas vs. Rural Areas

EJ identification methods typically evaluate populations close to environmental pollution sources. Population data can assist toxicologists and economists in determination of magnitude of effect (i.e., number of people affected). For example, economists report that net benefits of Federal air pollution control policy are substantially higher for residents of large urban areas than for rural residents. Health professionals use population information to evaluate location or cultural related exposures (mobile and

industrial source air pollution, shellfish consumption, subsistence fishing) or factors related to economics (e.g., deficient diet, minimal preventive health care, occupational exposures).

Most, if not all, EPA EJ identification methodologies have use of total population as a significant variable in their assessment algorithms. Total populations and percentage of minorities residing within given distances from Superfund (CERCLA) National Priorities List sites were demographic variables used in several studies. It is important to note, however, that EJ identification analyses based on population density and industrialization are biased toward urban areas. Studies using percentage of minority or low-income residents around a site regardless of total population have less of an urban bias, but significant bias remains when radii or set distances from the pollution source are used for study areas. It is possible to have significant environmental justice effects with few or no residents in the areas of interest. Affected land can be owned by low-income or minority persons who do not actually live in the area or these same low population areas can be of significant cultural historical value. A situation where the magnitude of EJ concern is not a function of population is common. Recent sites of regulatory interest in Region VI EPA include: farmland in Louisiana being considered as a building site for a plastic manufacturing facility; a rural community in Texas strongly resisting operations of a nearby landfill; and river waters in Oklahoma adjacent to Native American owned lands and a uranium enrichment facility. Each location is sparsely populated and near minority owned land and/or communities.[6] The EJ movement goes back to the siting of a landfill in Warren County, N.C., another example of environmental justice concerns in rural, lesser-populated areas.

Wording in the National Environmental Policy Act (NEPA) requires Federal agencies to consider a variety of environmental effects "...aesthetic, historic, cultural, economic, social, or health, whether direct, indirect, or cumulative." These effects often have EJ relevance but do not always have urban association. Calculations relying heavily upon population density can easily under-estimate Environmental Justice effects to the community as well as the ecological resources in low population areas.

A separate methodology for rural areas is possible, but would result in significant methodological inconsistencies. Consistency of analytical approach is needed for clear understanding and communication of socio-economic issues. Although less technical, an effective approach can be the development of an implementation process that fosters understanding of specific EJ analysis limitations. Such a process would address rural versus urban calculations.

6. Recommendations and Conclusions

The Executive Order requires that "....each Federal agency, whenever practicable and appropriate, shall collect, maintain, and analyze information assessing and comparing environmental and human health risks borne by populations identified by race, national origin, or income." In order to meet this requirement, EJ analytical tools, data, and study results should be easily accessible within or outside the Agency, and analysts must be able to readily understand and replicate these methods and results. This document includes not only a discussion of available resources, tools, and approaches to use in EJ analyses, but also makes specific recommendations to facilitate more consistent, conceptually sound, and methodologically defensible assessments.

There are a variety of methodological options available for conducting an EJ assessment. To assist the EJ analyst in developing an analytical approach, this document recommends several definitions as well as quantitative and qualitative methods. These definitions and methods are presented below along with a cite for each from the body of the report. It is important to note that these definitions and methods are only recommendations. It is up to the analyst to ensure that the definition or methods chosen fit the particular assessment at hand.

I. Recommendation for Developing a Conceptual Framework (Page 7) - The complexity of socio-economic and health risk scenarios and the need to assess the analytical demands of a study requires developing a conceptual framework or analytical strategy before performing an EJ assessment. It is the responsibility of the EJ analyst to develop such a strategy, with the input of those requesting the study.

In creating a conceptual framework, it is important to first establish the purpose and significance of the decision-making activity that the assessment is expected to support (e.g., how will the outcome of the analysis affect the community of concern, the potential for legal challenge, the public's interest, or the direct and indirect economic factors of investment and development issues). Based on the significance of that decision-making activity, a 'level of uncertainty' should be determined for the assessment. The level of uncertainty concept refers to the confidence an analyst has in the accuracy of the assessment leading to the eventual decision. A level of uncertainty is dependent on factors influencing the validity of that assessment, such as the accuracy of data, the rigorousness of the methodology, or the level of effort. The level of effort required for a particular assessment may be limited by available resources (e.g., time, money, data availability, required hardware and software, etc.).

II. Recommended Definition of Minority (Page 15) - The sum of the populations identified as members of *American Indian or Alaskan Native, Asian or Pacific Islander,*

Black, and *Other races* together with *Hispanics* with a listed race of *White*. This is equivalent to an alternate calculation of total population minus *White non-Hispanics*.

III. Recommended Definition of Low-Income (Page 21) - This document recommends a two-tiered approach which includes the following definitions for low-income and very low-income populations:

- *Low-income* is defined as income below two times the poverty threshold, the STF3A $25,000 income range, or 60% of AMGI values (imbedded in Qualified Census Tracts);
- *Very low-income* is defined as income below the poverty threshold or the STF3A $15,000 income range.

IV. Recommended Methods for Determining the Significance of Minority Populations (Page 53) - The methods for determining whether or not a community is considered a minority population, in order of higher to lower levels of allowable uncertainty are: absolute thresholds; relative thresholds; and the comparison method. Recommended absolute thresholds are 50% or the national average (25%) and relative thresholds are the state average plus 20% (state average times 1.2). When using the comparison approach (most rigorous) an analyst should screen using a fixed percentage cut-off, then employ statistics to determine if the difference is significant.

V. Recommended Methods for Determining the Significance of Low-Income Populations (Page 55) - The methods for determining whether or not a community is considered a low-income population, in order of higher to lower levels of allowable uncertainty, are: absolute thresholds; relative thresholds; and the comparison method. Recommended moderate absolute thresholds are 20% or more of the population below the selected low-income benchmark (i.e., poverty threshold, 2 times poverty threshold, STF3A $15,000 income range, or the STF3A $25,000 income range). Extreme absolute thresholds are 40% or more of the population below the selected low-income benchmark (i.e., poverty threshold, 2 times poverty threshold, STF3A $15,000 income range, or the STF3A $25,000 income range). Relative thresholds are the percent of persons below the selected threshold greater than or equal to the related state percent. When using the comparison approach (most rigorous) an analyst screen using a fixed percentage cut-off, then employ statistics to determine if the difference is significant. The specific percentage cut-off recommended in this guidance is 20%.

VI. Recommendations for Considering Vulnerability (Page 68) - Although the Executive Order does not explicitly state that the Agency needs to consider the degree of vulnerability in a community, it may be important when considering effects. Persons who are most susceptible to adverse effects include: children; the elderly; pregnant women and their unborn fetuses; persons lacking in education; persons who rely on subsistence hunting and fishing; persons who reside in older housing; and single mothers with

children under 18 years. In addition, EO 12898 directs Federal agencies conducting health research to ".... whenever practicable and appropriate, shall include diverse segments of the population in epidemiological and clinical studies..." The purpose of this directive is to create the data for dealing with differences in vulnerability.

Effects that are widely recognized throughout the EJ community for use in EJ analyses, are: proximity to a source or sources; cumulative effect of multiple sources; and Agency action or inaction (e.g., alleged delays in Agency clean-ups in minority and/or low-income areas when compared with clean-ups in other areas). Still, this document recommends that EJ analysts look beyond the relative pounds of emissions as a source of human health effects only, and consider *other* adverse effects, (e.g., effects on religious practices or social structures, etc.). Some *other* sources of adverse effects are contained in the following examples:

- *Urban Pesticide Abuse* - The illegal diversion of restricted-use pesticides, particularly methyl and ethyl parathion, from the agribusiness to minority an low-income communities has become an emerging problem in recent years. As such, EPA's Office of Enforcement and Compliance Assurance and the Office of Prevention, Pesticide, and Toxic Substances have joined forces to address this new EJ related problem.

- *Concentrated Animal Feedlot Operations (CAFO's)* - There are a growing number of animal feedlot operations (e.g., poultry and hog producers, etc.), being located in rural low-income and minority communities. These have created serious environmental problems ranging from water contamination and fish kills to severe odor problems. Currently, there are no Federal effluent standards for these facilities.

- *West Oakland Pilot Project* - The West Oakland community residents are concerned about recent zoning, contamination from past and present industrial activities, truck traffic, and impacts of the freeway construction. EPA is looking at a full range of environmental and related issues affecting the West Oakland community; and working with residents as well as local and State government agencies and other parties to identify and prioritize those issues and develop means for resolving them.

- *Mattaponi Tribe vs King William County Reservoir* - In Williamsburg, (King William County) Virginia, the decision to build a dam, reservoir, and pumping station on the Mattaponi and Pamunkey Rivers impacts the social structure and sense of community of the Mattaponi and Pamunkey Tribes as well as raising historic preservation issues. The Mattaponi Tribe has raised environmental justice issues under Federal Indian Law, Environmental and Civil Rights laws, and a 350 year old treaty with Charles II to which all parties agreed and is still in effect.

- *Homer, Louisiana and the Louisiana Energy Services Case Study* - Louisiana Energy Services (LES) , a consortium of private energy power companies, was applying to be licensed by the Nuclear Regulatory Commission (NRC) to site a uranium enrichment facility in Claiborne Parish, Louisiana. NRC required a revision of the Environmental Impact Study (EIS) for the project, because it had not considered the socio-economic adverse effects on two small, low-income, African- American towns most adjacent to the site. The proposed site fell between the two towns and impacted the access to a church and a food store shared by the two towns.

- *South Lawrence Trafficway* - The Federal Highway Administration (FHWA) proposed a trafficway in Lawrence, KS at the Haskell Indian Nations University. The proposed road expansion, immediately adjacent to "sweat lodges" and a "medicine wheel" was proposed. Haskell Indian Nations University informed the FHWA that the EIS didn't adequately consider the impacts to the cultural and spiritual community at Haskell. As a result, Haskell requested that a Supplemental EIS be undertaken to respond to Haskell's concerns. While the noise from the trafficway would not be considered significant under normal circumstances, it would in this case, significantly *effect the religious practices* carried out at this site.

Unfortunately, the data and methods for dealing with these sources and effects are not as fully developed and consistently applied as in the more established environmental and human health effect studies. Still, they are considered an integral part of the EJ problem and as such will develop with the continual exploration of EJ concepts and methods.

This document is a work in progress and as such, the concepts and methods presented may change over time. In particular, those areas in need of further research such as cumulative and synergistic effects, rural vs metropolitan area considerations, or other sources of adverse effects, are expected to evolve into more thoroughly developed and consistently applied approaches. Optimally this guidance will standardize and promulgate the basic concepts and methods of an EJ assessment, as well as lay the groundwork for the development of more rigorous and technically advanced methods and tools.

Appendix A. Text of Executive Order 12898: Federal Actions to Address Environmental Justice in Minority Populations and Low-Income Populations

By the authority vested in me as President by the Constitution and the laws of the United States of America, it is hereby ordered as follows:

Sec. 1-1. Implementation

1-101. Agency Responsibilities

To the greatest extent practicable and permitted by law, and consistent with the principles set forth in the report on the National Performance Review, each Federal agency shall make achieving environmental justice part of its mission by identifying and addressing, as appropriate, disproportionately high and adverse human health or environmental effects of its programs, policies, and activities on minority populations and low-income populations in the United States and its territories and possessions, the District of Columbia, the Commonwealth of Puerto Rico, and the Commonwealth of the Mariana Islands.

1-102. Creation of an Interagency Working Group on Environmental Justice

a. Within 3 months of the date of this order, the Administrator of the Environmental Protection Agency ("Administrator") or the Administrator's designee shall convene an interagency Federal Working Group on Environmental Justice ("Working Group"). The Working Group shall comprise the heads of the following executive agencies and offices, or their designees: Department of Defense; Department of Health and Human Services; Department of Housing and Urban Development; Department of Labor; department of Agriculture; department of Transportation; Department of Justice; Department of the Interior; Department of Commerce; Department of Energy; Environmental Protection Agency; Office of Management and Budget; Office of Science and Technology Policy; Office of the Deputy Assistant to the President for Environmental Policy; Office of the Assistant to the President for Domestic Policy; National Economic Council; Council of Economic Advisers; and such other Government officials as the President may designate.

The Working Group shall report to the President through the Deputy Assistant to the President for Environmental Policy and the Assistant to the President for Domestic Policy.

a. The Working Group shall: provide guidance to Federal agencies on criteria for identifying disproportionately high and adverse human health or environmental effects on minority populations and low-income populations; coordinate with, provide

guidance to, and serve as a clearinghouse for, each Federal agency as it develops an environmental justice strategy as required by section 1-103 of this order, in order to ensure that the administration, interpretation and enforcement of programs, activities and policies are undertaken in a consistent manner; assist in coordinating research by, and stimulating cooperation among, the Environmental Protection Agency, the Department of Health and Human Services, the Department of Housing and Urban Development, and other agencies conducting research or other activities in accordance with section 3-3 of this order; assist in coordinating data collection, required by this order; examine existing data and studies on environmental justice; hold public meetings as required in section 5-502(d) of this order; and develop interagency model projects on environmental justice that evidence cooperation among Federal agencies.

1-103. Development of Agency Strategies

a. Except as provided in section 6-605 of this order, each Federal agency shall develop an agency-wide environmental justice strategy, as set forth in subsections (b)-(e) of this section that identifies and addresses disproportionately high adverse human health or environmental effects of its programs, policies, and activities on minority populations and low-income populations. The environmental justice strategy shall list programs, policies, planning and public participation processes, enforcement, and/or rule makings related to human health or the environment that should be revised to, at a minimum: promote enforcement of all health and environmental statutes in areas with minority populations and low-income populations; ensure greater public participation; improve research and data collection relating to the health of and environment of minority populations and low-income populations; and identify differential patterns of consumption of natural resources among minority populations and low-income populations.

In addition, the environmental justice strategy shall include, where appropriate, a timetable for undertaking identified revisions and consideration of economic and social implications of the revisions.

a. Within 4 months of the date of this order, each Federal agency shall identify an internal administrative process for developing its environmental justice strategy, and shall inform the Working Group of the process.

b. Within 6 months of the date of this order, each Federal agency shall provide the Working Group with an outline of its proposed environmental justice strategy.

c. Within 10 months of the date of this order, each Federal agency shall provide the Working Group with its proposed environmental justice strategy.

d. Within 12 months of the date of this order, each Federal agency shall finalize its environmental justice strategy and provide a copy and written description of its strategy to the Working Group. During the 12 month period from the date of this

order, each Federal agency, as part of its environmental justice strategy, shall identify several specific projects that can be promptly undertaken to address particular concerns identified during the development of the proposed environmental justice strategy, and a schedule for implementing those projects.

e. Within 24 months of the date of this order, each Federal agency shall report to the Working Group on its progress in implementing its agency-wide environmental justice strategy.

f. Federal agencies shall provide additional periodic reports to the Working Group as requested by the Working Group.

1-104. Reports to the President

Within 14 months of the date of this order, the Working Group shall submit to the President, through the Office of the Deputy Assistant to the President for Environmental Policy and the Office of the Assistant to the President for Domestic Policy, a report that describes the implementation of this order, and includes the final environmental justice strategies described in section 1-103(e) of this order.

Sec. 2-2. Federal Agency Responsibilities for Federal Programs

Each Federal agency shall conduct its programs, policies, and activities that substantially affect human health or the environment, in a manner that ensures that such programs, policies, and activities do not have the effect of excluding persons (including populations) from participation in, denying persons (including populations) the benefits of, or subjecting persons (including populations) to discrimination under, such programs, policies, and activities, because of their race, color, or national origin.

Sec. 3-3. Research, Data Collection, and Analysis

3-301. Human Health and Environmental Research and Analysis

a. Environmental human health research, whenever practicable and appropriate, shall include diverse segments of the population in epidemiological and clinical studies, including segments at high risk from environmental hazards, such as minority populations, low-income populations and workers who may be exposed to substantial environmental hazards.

b. Environmental human health analyses, whenever practicable and appropriate, shall identify multiple and cumulative exposures.

c. Federal agencies shall provide minority populations and low-income populations the opportunity to comment on the development and design of research strategies undertaken pursuant to this order.

3-302. Human Health and Environmental Data Collection and Analysis

To the extent permitted by existing law, including the Privacy Act, as amended (5 U.S.C. section 552a): each Federal agency, whenever practicable and appropriate, shall collect, maintain, and analyze information assessing and comparing environmental and human health risks borne by populations identified by race, national origin, or income.

To the extent practical and appropriate, Federal agencies shall use this information to determine whether their programs, policies, and activities have disproportionately high and adverse human health or environmental effects on minority populations and low-income populations;

a. In connection with the development and implementation of agency strategies in section 1-103 of this order, each Federal agency, whenever practicable and appropriate, shall collect, maintain and analyze information on the race, national origin, income level, and other readily accessible and appropriate information for areas surrounding facilities or sites expected to have a substantial environmental, human health, or economic effect on the surrounding populations, when such facilities or sites become the subject of a substantial Federal environmental administrative or judicial action.

Such information shall be made available to the public, unless prohibited by law; and

a. Each Federal agency, whenever practicable and appropriate, shall collect, maintain, and analyze information on the race, national origin, income level, and other readily accessible and appropriate information for areas surrounding Federal facilities that are: subject to the reporting requirements under the Emergency Planning and Community Right-to-Know Act, 42 U.S.C. section 11001-11050 as mandated in Executive Order No. 12856; and expected to have a substantial environmental, human health, or economic effect on surrounding populations. Such information shall be made available to the public, unless prohibited by law.

b. In carrying out the responsibilities in this section, each Federal agency, whenever practicable and appropriate, shall share information and eliminate unnecessary duplication of efforts through the use of existing data systems and cooperative agreements among Federal agencies and with State, local, and tribal governments.

Sec. 4-4. Subsistence Consumption of Fish and Wildlife

4-401. Consumption Patterns

In order to assist in identifying the need for ensuring protection of populations with differential patterns of subsistence consumption of fish and wildlife, Federal agencies, whenever practicable and appropriate, shall collect, maintain, and analyze information on the consumption patterns of populations who principally rely on fish and/or wildlife for subsistence. Federal agencies shall communicate to the public the risks of those consumption patterns.

4-402. Guidance

Federal agencies, whenever practicable and appropriate, shall work in a coordinated manner to publish guidance reflecting the latest scientific information available concerning methods for evaluating the human health risks associated with the consumption of pollutant-bearing fish or wildlife. Agencies shall consider such guidance in developing their policies and rules.

Sec. 5-5. Public Participation and Access to Information

a. The public may submit recommendations to Federal agencies relating to the incorporation of environmental justice principles into Federal agency programs or policies. Each Federal agency shall convey such recommendations to the working Group.
b. Each Federal agency may, whenever practicable and appropriate, translate crucial public documents, notices, and hearings relating to human health or the environment for limited English speaking populations.
c. Each Federal agency shall work to ensure that public documents, notices, and hearings relating to human health or the environment are concise, understandable, and readily accessible to the public.
d. The Working Group shall hold public meetings, as appropriate, for the purpose of fact-finding, receiving public comments, and conducting inquiries concerning environmental justice. The Working Group shall prepare for public review a summary of the comments and recommendations discussed at the public meetings.

Sec. 6-6. General Provisions

6-601. Responsibility for Agency Implementation

The head of each Federal agency shall be responsible for ensuring compliance with this order. Each Federal agency shall conduct internal reviews and take such other steps as may be necessary to monitor compliance with this order.

6-602. Executive Order No. 12250

This Executive order is intended to supplement but not supersede Executive Order No. 12250, which requires consistent and effective implementation of various laws prohibiting discriminatory practices in programs receiving Federal financial assistance. Nothing herein shall limit the effect or mandate of Executive Order No. 12250.

6-603. Executive Order No. 12875

This Executive order is not intended to limit the effect or mandate of Executive Order No. 12875.

6-604. Scope

For purposes of this order, Federal agency means any agency on the Working Group, and such other agencies as may be designated by the President, that conducts any Federal program or activity that substantially affects human health or the environment. Independent agencies are requested to comply with the provisions of this order.

6-605. Petitions for Exemptions

The head of a Federal agency may petition the President for an exemption from the requirements of this order on the grounds that all or some of the petitioning agency's programs or activities should not be subject to the requirements of this order.

6-606. Native American Programs

Each Federal agency responsibility set forth under this order shall apply equally to Native American programs. In addition, the Department of the Interior, in coordination with the Working Group, and, after consultation with tribal leaders, shall coordinate steps to be taken pursuant to this order that address Federally-recognized Indian Tribes.

6-607. Costs

Unless otherwise provided by law, Federal agencies shall assume the financial costs of complying with this order.

6-608. General

Federal agencies shall implement this order consistent with, and to the extent permitted by, existing law.

6-609. Judicial Review

This order is intended only to improve the internal management of the executive branch and is not intended to, nor does it create any right, benefit, or trust responsibility, substantive or procedural, enforceable at law or equity by a party against the United States, its agencies, its officers, or any person. This order shall not be construed to create any right to judicial review involving the compliance or noncompliance of the United States, its agencies, its officers, or any other person with this order.

William Clinton
The White House,
February 11, 1994.

Executive Order No. 12898
59 FR 7629
1994 WL 43891 (Pres.)

Appendix B. List of EPA Environmental Justice Contacts

United States Environmental Protection Agency Headquarters
401 M Street, SW
Washington, DC 20460

Please use Mail Code (MC) for each Office

Office of the Administrator
Mail Stop: MC-1101
Phone: (202) 260-8852
Contact: Angela Chung
E-Mail: chung.angela@epamail.epa.gov
Fax: (202) 962-6215

Office of Environmental Justice
Mail Stop: MC-2201A
Phone: (202) 564-2515
(800) 962-6215
Contact: Dr. Robert Knox
E-Mail: knox.robert@epamail.epa.gov
Fax: (202) 962-6215

American Indian Environmental Office
Mail Stop: MC-4104
Phone: (202) 260-8106
Contact: Elizabeth Bell
E-Mail: bell.elizabeth@epamail.epa.gov
Fax: (202) 260-7509

Office of Air and Radiation
Mail Stop: MD-52
Phone: (202) 260-5574
Contact: Dr. William Wilson
E-Mail: wilson.william@epamail.epa.gov
Fax: (202) 260-0253

Office of Civil Rights
Mail Stop: MC-1201
Phone: (202) 260-4567
Contact: Rosezella Canty-Letsome
E-Mail: canty-letsome@epamail.epa.gov
Fax: (202) 260-4580

Office of Communication, Education, and Public Affairs
Mail Stop: MC-1702
Phone: (202) 260-3534
Contact: Doretta Reaves
E-Mail: reaves.doretta@epamail.epa.gov
Fax: (202) 260-0130

Office of Enforcement and Compliance Assurance
Mail Stop: MC-2201A
Phone: (202) 564-2619
Contact: Sherry Milan
E-Mail: Milan.sherry.epamail.epa.gov
Fax: (202) 501-0284

Office of General Counsel
Mail Stop: MC-2322
Phone: (202) 260-1487
Contact: Mary O'Lone
E-Mail: olone.mary.epamail.epa.gov
Fax: (202) 260-8393

Office of Policy, Planning, and Evaluation
Mail Stop: MC-2127
Phone: (202) 260-2730
Contact: Janice Bryant
E-Mail: bryant.janice@epamail.epa.gov
Fax: (202) 260-0174

Office of Prevention, Pesticides, and Toxic Substances
Mail Stop: MC-7408
Phone: (202) 260-2301
Contact: Carol Christensen
E-Mail: christensen.carol@epamail.epa.gov
Fax: (202) 260-8850

Office of Regional Operations and State/Local Relations
Mail Stop: MC-1502
Phone: (202) 260-6188
Contact: Janice Berry-Chen
E-Mail: berry-chen@epamail.epa.gov
Fax: (202) 260-9365

Office of Research and Development
Mail Stop: MC-8103R
Phone: (202) 564-6497
Contact: Lawrence Martin
E-Mail: martin.lawrence@epamail.epa.gov
Fax: (202) 260-0507

Office of Solid Waste and Emergency Response
Mail Stop: MC-5101
Phone: (202) 260-2822
Contact: Kent Benjamin
E-Mail: benjamin.kent@epamail.epa.gov
Fax: (202) 260-6606

Office of Water
Mail Stop: MC-1803
Phone: (202) 260-5410
Contact: Sandra Germann
E-Mail: germann.sandy@epamail.epa.gov
Fax: (202) 260-7923

US EPA Regional Contacts

US EPA Region I
One Congress Street, 10th Floor
Boston, MA 02203-0001
Contact: Rhona Julien
Phone: (617) 565-9454
E-Mail: julien.rhona@epamail.epa.gov
Fax: (617) 565-3415

US EPA Region II
290 Broadway, 26th floor
New York, NY 10007
Contact: Melva Hayden
Phone: (212) 637-5027
E-Mail: hayden.melva@epamail.epa.gov
Fax: (212) 637-5024

US EPA Region III
841 Chestnut Building
Philadelphia, PA 19107
Contact: Reginald Harris E-Mail: harris.reginald@epamail.epa.gov
Phone: (215) 566-2988 Fax: (215) 566-2901

US EPA Region IV
345 Courtland Street, NE
Atlanta, GA 30365
Contact: Connie Raines E-Mail: raines.connie@epamail.epa.gov
Phone: (404) 562-9671 ext. 6764 Fax: (404) 562-9664

US EPA Region V
77 West Jackson Blvd., MC T-165
Chicago, IL 60604-3507
Contact: Karla Johnson E-Mail: johnson.karla@epamail.epa.gov
Phone: (312) 886-5993 Fax: (312) 886-2737

US EPA Region VI
1445 Ross Ave., 12th Floor, Suite 1200
Dallas, TX 75202-2733
Contact: Shirley Augurson E-Mail: augurson.shirley@epamail.epa.gov
Phone: (214) 665-7401 Fax: (214) 665-2146

US EPA Region VII
726 Minnesota Avenue, CNSLRGAG
Kansas City, KS 66101
Contact: Rupert Thomas E-Mail: thomas.rupert@epamail.epa.gov
Phone: (913) 551-7282 Fax: (912) 551-7925

US EPA Region VIII
999 18th Street, Suite 500 (8ENF-EJ)
Denver, CO 80202-2405
Contact: Elisabeth Evans E-Mail: evans.elisabeth@epamail.epa.gov
Phone: (303) 312-6053 Fax: (303) 312-6558

US EPA Region IX
75 Hawthorn Street
San Francisco, CA 94105
Contact: Lori Lewis E-Mail: lewis.lori@epamail.epa.gov
Phone: (415) 744-1542 Fax: (415) 744-1605

US EPA Region X
1200 Sixth Avenue
Seattle, WA 98101
Contact: Joyce Crosson-Kelly
Phone: (206) 553-4029
E-Mail: crosson-kelly@epamail.epa.gov
Fax: (206) 553-8338

Appendix C. Glossary of Terms

Aesthetic Effect	An environmental effect based on desirability in appearance, taste or odor, but not associated with adverse effects on health or welfare.
AFS	US EPA AIRS Facility Subsystem
Agent	The substance that is responsible for an effect on health or welfare.
Aggregation	The grouping of a set of observations on a variable within a geographical boundary. For example, calculation of the sum (the total number) of persons of low-income residing within a county administrative area.
AIRS	US EPA Aerometric Information Retrieval System, the database containing the Agency's air-related data.
Ambient Condition	The meteorological or atmospheric state in a specific location.
Area of Concern	A potentially affected Target Area that has some disproportionate environmental effects and a significant minority and/or low-income population.
Bias	A systematic or subjective distortion of statistics as a result of the sampling procedure, method choice, or interpretation.
Block	Census blocks are small areas bounded on all sides by visible features such as streets, roads, streams, and railroad tracks, and by invisible boundaries such as city, town, township, and county limits, property lines, and short, imaginary extensions of streets and roads.
Block Group	A unit for Census data reporting formed by a cluster of Census blocks. Census block groups generally contain between 250 and 550 housing units.
Boundary	A line between two or more geographical areas defining administrative and property boundaries.

CD-ROM	Compact Disc - Read Only Memory
CENDATA	US Census Bureau online database
Decennial Census	The 10-year official enumeration (counting) of the US population, with details as to gender, age, sex, race, income, and other information.
CERCLIS	Comprehensive Environmental Response, Compensation and Liability Information System, a database containing information on Superfund sites.
Community	A geographical area defined by social, historical, cultural, or political boundaries.
Community Input	Information provided by representatives of an affected community on neighborhood boundaries, health concerns etc.
Contiguous	Bordering or adjoining (as in neighboring communities, or administrative areas).
Criteria	Established standards for environmental contaminants.
Cumulative Exposure	Total exposure to environmental contaminants, including exposures originating from multiple sources.
Decision-Criterion	An established test used for determining whether a Target Community meets a specific EJ factor (such as low-income, and minority).
Demographic	Data on population characteristics, such as total population, births, deaths, etc.
Disproportionate Environmental Burden	The adverse human health or environmental effect on a particular community or segment of the population (the Target Community) that is out of proportion to the level of the same effect felt in a Reference Community, and can be either actual or potential.
Enumeration Unit	A geographical area where data has been collected and published, which includes a Census block, block group, or other summary level available from the decennial Census.

Environmental Burden	The adverse human health or environmental effect on a particular community or segment of the population related to a specific source or sources, resulting from cumulative or area-wide sources, and/or resulting from uneven application of government authorities.
Environmental Impact	Degree/intensity of Environmental Burden
Environmental Justice	The fair treatment of people of all races and cultures, and those with lower incomes, with respect to the development, implementation, and enforcement of environmental laws, regulations and policies.
Ethnic Group	A group of persons of the same race or nationality who share a common and distinctive culture.
Ethnicity Data	Describes information pertaining to a population's *religious*, *national*, *racial*, or *cultural* composition.
Exposure	Subject to the action or influence of environmental contaminants through ingestion, inhalation, or skin contact.
Facility	A factory, plant, industry, utility, or commercial establishment that is a potential or known source of environmental contamination or degradation.
Federal Register	Published by the US Government. It lists government announcements, rules and regulations.
Geographic Information System	A computer system designed to efficiently capture, store, analyze and display forms of geographically referenced information. A GIS is used to produce thematic maps, allowing for convenient visual analysis of information.
Government Intervention	Judgment, action or command taken by a local, state, or Federal governing body to address a specific environmental issue.
HHS	US Department of Health and Human Services
Hispanic	Persons who classify themselves as Mexican, Puerto Rican, or Cuban, as well as those who indicate that they are of other Spanish/Hispanic origins.

Housing Value	A data category in the US Census that represents the attributed worth of the homes in a designated area; often used as a surrogate for income in analyses.
HUD	US Department of Housing and Urban Development
Indexes	Formulae that express some ratio calculation between at least two variables.
Large Quantity Generator	A facility that produces greater than a threshold quantity of a toxic substance annually (i.e., greater than 1000 pounds of a chemical substance).
Mean	The average value of a group of values (the sum of observation on a variable divided by the number of observations).
Median	The middle observation in a ranked group of values.
Meta-Data Sample File	A database from the Bureau of the Census that includes records for unidentified individuals, households, and housing units.
Mode	The value within group of values that occurs with the greatest frequency.
Minority Data	Data characterizing a population group that is smaller in number when compared to the *majority* population group that resides within an population areal unit.
NEPA	National Environmental Policy Act
NPL	National Priorities List
OEJ	US EPA Office of Environmental Justice
PCS	US EPA Permit Compliance System, a database of direct discharge of wastewater.
Percentage	A rate or proportion per hundred.
Political Boundary	The line dividing two areas with separate governing bodies, i.e., cities from rural areas, one county from the next.

Population Density	An index assessing the number of people contained with in a unit area, i.e., persons per square mile.
Proportion	The comparative relation between things or magnitudes; a ratio.
Proximate	Adjacent, or very near to.
Quartile	In a statistical frequency distribution, one of the values of a variable that divides the distribution into four groups having equal frequencies.
Query	To extract, to retrieve information from a database. Query languages (such as Structured Query Language (SQL)) enable a user to define a request for geographic data and attribute data (often stored as tables or spreadsheets).
Quintile	In a statistical frequency distribution, one of the values of a variable that divides the distribution into five groups having equal frequencies.
Racial Group	A group of persons related by common descent/genetic heredity.
Range	The difference between the smallest and largest values in a statistical distribution (a group of observations).
RCRA	Resource Conservation and Recovery Act
RCRIS	US EPA Resource Conservation and Recovery Information System
Reference Area	An area that is used as a standard of comparison when determining whether a Target Area suffers from disproportionate environmental burdens, and/or whether the target area has a significantly greater minority or low-income population.
Regulatory Effect	A potential form of environmental injustice characterized by bias in the administration of government programs.
Screening Analysis	An initial Environmental Justice analysis for identifying areas that may have Environmental Justice concerns.

Sensitivity Analysis	A statistical procedure conducted to identify the factors in an analysis that have the greatest bearing on the outcome.
Site-Specific Analysis	An Environmental Justice analysis intended to determine whether a specific identified area (the Target Area) has Environmental Justice concerns.
Source	A site, facility, or area from which one (or a suite of) environmental contaminants originate (i.e., a factory, or incinerator).
STF/Summary Tape File	US Bureau of the Census Summary Tape Files. The STF files are a commonly used source of information for EJ analyses.
STORET	US EPA Storage and Retrieval of Water-Related Data System
Summary Statistic	A calculation used to characterize a set of observations (for example: mean, variance, standard deviation statistics).
Superfund	US EPA's uncontrolled hazardous waste site program created by the Comprehensive Environmental Response and Liability Act.
Surrogate Measure	A demographic or other factor assumed to be representative of a second factor for which data are unavailable (i.e., house value is often used as a surrogate measure for income).
Target Area	A geographical area that is potentially burdened by physical and/or chemical effects on the environment and the health of those residing there. The materials are created and dispersed as a result of natural and anthropogenic processes.
Target Population	Part and/or full time residents of a Target Area. May include transient residents, such as migrant workers, commuters, and visitors.
TIGER	TIGER is the acronym for the Bureau of the Census's digital (GIS-readable) geographic database that helps automate mapping and related activities in a GIS. The

	Topologically Integrated Geographic Encoding and Referencing data format is commonly readable by a GIS.
TRI/Toxic Release Inventory	The US EPA program which requires large-quantity generators of hazardous materials to report the nature and quantity of their annual emissions into the environment.
TRIS	US EPA Toxic Release Inventory System. The TRIS database is a major source of contaminant release information for EJ analyses.
Tract	An area of land used in the data collection efforts of the Bureau of the Census. Tracts have boundaries.
Transient	A person who is not expected to remain in a given location for an extended period of time.
Variance	The natural heterogeneity (variability) within a population.

Appendix D. Environmental Justice Bibliography

African American Environmentalist Association (1994). "Our Unfair Share: A Survey of Pollution Sources in Our Nation's Capital." African American Environmentalist Association.

Anderton, D., Anderson, A., Rossi, P., et al. (1994). "Hazardous Waste Facilities: Environmental Equity Issues in Metropolitan Areas." Evaluation Review, 18: 123.

Boerner, C. and Lambert, T. (1995). "Environmental Justice in the City of St. Louis: The Economics of Siting Industrial and Water Facilities, St. Louis, MO." Washington University Center for the Study of American Business.

Calabrese, E.J. (1978). "Developmental Processes." In Pollutants and High Risk Groups. John Wiley & Sons, NY.

Calabrese, E.J. (1986). Age and Susceptibility to Toxic Substances. John Wiley & Sons, NY.

Council on Environmental Quality (1978). Regulations for Implementing the Procedural Provisions of the National Environmental Policy Act, 40 CFR Parts 1508.8. Executive Office of the President, Washington, DC.

Council on Environmental Quality (1996). Guidance for Federal Agencies for Key Terms in Executive Order 12898. Executive Office of the President, Washington, DC.

Cutter, S. (1995). "Race, Class and Environmental Justice." Progress in Human Geography 19:107-18.

Fox, C.A. (1996). "Risk-Based Targeting: A New Approach in Environmental Protection." Presented at the Society of Environmental Toxicology and Chemistry Meetings, Washington, DC.

Fox, C.A. and Maynard, P. (1996). "Virginia Childhood Lead Poisoning Estimation Model: Field Test Report." Virginia Department of Health, Division of Child and Adolescent Health.

Goldman, B.A. (1993). "Not Just Prosperity: Achieving Sustainability with Environmental Justice (Draft Report)." National Wildlife Federation Corporate Conservation Council.

Goldman, B.A. and Fitton, L. (1994). "Toxic Wastes and Race Revisited." Center for Policy Alternatives, the National Association for the Advancement of Colored People, and the United Church of Christ Commission for Racial Justice. Washington, DC.

Guralnik J.M., Land K.C., Blazer, D. et al. (1993). "Educational Status and Active Life Expectancy Among Older Blacks and Whites." The New England Journal of Medicine, 329: 110-116.

Hall, L. (1997). Personal Communication between Fox, C. of EPA OECA and Hall, L. of EPA OPPT.

Hamilton, J.T. (1995). "Testing for Environmental Racism: Prejudice, Profits, Political Power?." Journal of Policy Analysis and Management, 14:107.

Heitgerd, J., Burg, J., and Strickland, H. (1995). "A Geographic Information Systems Approach to Estimating and Assessing National Priorities List Site Demographics: Racial and Hispanic Origin Composition." International Journal of Occupational Medicine and Toxicology, Vol. 4, No. 3: 343.

Klaassen, C.D., Amdur, M.O. and Doull, J. [eds.] (1986). Casarett and Doull's Toxicology, the Basic Science of Poisons (3rd Edition). Macmillan, NY.

Lake, R. W. (1996). "Volunteers, NIMBYs, and Environmental Justice: Dilemmas of Democratic Practice." Antipode, Vol. 28, No. 2: 160-174.

Liberatos P., Link B.G., and Kelsey J.L. (1988). "The Measurement of Social Class in Epidemiology." Epidemiology Reviews, 10: 87-121.

Matthews, T.J. (19??). Personal Communication. Bureau of the Census, Population Division. Washington, DC.

Nieves (1992). "Not in Whose Backyard? Minority Population Concentrations and Noxious Facility Sites." Presented at the American Academy for the Advancement of Science Meetings, Chicago, IL.

Nieves and Nieves (1992). "Regional Differences in the Potential Exposure of US Minority Populations to Hazardous Facilities." Presented at the Regional Science Association Meetings, Chicago, IL.

Plunkett L.M. et al. (1992). "Differences between Adults and Children Affecting Exposure Assessment." In Similarities and Differences between Children and Adults, Implications for Risk Assessment, Guzelman, P.S., Henry C.J, and Olin S.S. [eds.]. ILSI Press, Washington, DC.

Sargent, J.D., Brown, M.J., Freeman, J.L. et al. (1995). "Childhood Lead Poisoning in Massachusetts Communities: Its Association with Socio-Demographic and Housing Characteristics." American Journal of Public Health, 85: 528-534.

Snodgrass, W.R. (1992). "Physiological and Biochemical Differences between Children and Adults as Determinants of Toxic Responses to Environmental Pollutants." In Similarities and Differences between Children and Adults, Implications for Risk Assessment, Guzelman, P.S., Henry C.J, and Olin S.S. [eds.]. ILSI Press, Washington, DC.

Tietenberg, T. (1996). "Environmental Justice." In Environmental and Natural Resource Economics (4th Edition). Harper Collins Publishers.

United Church of Christ (1987). "Toxic Wastes and Race in the United States: A National Report on the Racial and Socio-Economic Characteristics of Communities with Hazardous Waste Sites." Commission for Racial Justice of the United Church of Christ. New York, NY.

US Department of Commerce, Economics and Statistics Administration, Bureau of the Census (1992). "Census of Population and Housing (1990). Guide, Part A. [CPH-R-1A]." Washington, DC.

US Department of Commerce, Economics and Statistics Administration, Bureau of the Census (1992). "Census of Population and Housing (1990). Summary Tape File 3 on CD-ROM." Washington, DC.

US Department of Commerce, Economics and Statistics Administration, Bureau of the Census (1994). "Statistical Brief on Poverty - Long and Short Term [SB/94-34]." Washington, DC.

US Department of Housing and Urban Development (1990). "Comprehensive and Workable Plan for the Abatement of Lead-Based Paint in Privately-Owned Housing: A Report to Congress." Washington, DC.

US Department of Housing and Urban Development (1990). "Statutorily Mandated Designation of Qualified Census Tracts and Difficult Development Areas for Section 42 of the Internal Revenue Code of 1986." Washington, DC.

US Environmental Protection Agency (19??). "Exposures Factors Handbook, Review Draft Volume III, Activity Factors." (See Tables 14-102, 14-118, 14-130). EPA/600/P-95?002BC, ORD. Washington, DC.

US Environmental Protection Agency (1984). "Residential Wood and Coal Combustion." OAQPS, RTP, EPA/450/5-83-005a,b,c. Washington, DC.

US Environmental Protection Agency (1987). "Residential Wood Heaters - Background Information for Promulgated Standards." OAQPS, RTP, EPA/450/3-87-025. Washington, DC.

US Environmental Protection Agency, the US Public Health Service, and the National Environmental Health Association (1991). "Introduction to Indoor Air Quality, A Reference Manual." EPA/400-3-91/003. Washington, DC.

US Environmental Protection Agency (1994a). "Guidance for Federal Agencies on Key Terms in Executive Order 12898: Text of Executive Order 12898, Federal Actions to Address Environmental Justice in Minority Populations and Low-Income Populations." Washington, DC.

US Environmental Protection Agency (1994a). "Computer Assisted Environmental Justice Index Methodology." Office of Planning and Analysis, Region VI. Dallas, TX.

US Environmental Protection Agency (1995). "Catalog of Demographic Variables for Environmental Justice Analysis." Office of Pollution Prevention and Toxic Substances. Washington, DC.

US Environmental Protection Agency (1996a). "Environmental Justice Implementation Plan." Office of Environmental Justice. EPA 300-R-96-004. Washington, DC.

US Environmental Protection Agency (1996b). "Review Draft Guidance for Incorporating Environmental Justice Concerns in EPA's NEPA Compliance Analyses." Washington, DC.

US Environmental Protection Agency (1996c). "Preliminary Draft Interim Region II Policy on Identifying EJ Areas." New York, NY.

US Environmental Protection Agency (1996d). "Preliminary Draft Interim Environmental Justice Inspection Protocols for Region VIII." Denver, CO.

US General Accounting Office (1983). "Siting of Hazardous Waste Landfills and their Correlation with Racial and Economic Status of Surrounding Communities." GAO/RCED-83-168. Washington, DC.

US Office of Management and Budget (1978). "Statistical Policy Directive 15. Race and Ethnic Standards for Federal Statistics and Administrative Reporting." In

Statistical Policy Handbook, Department of Commerce, Office of Federal Statistical Policy and Standards. Washington, DC.

Wells, B.L. and Horm J.W. (1992). "Stage at Diagnosis in Breast Cancer: Race and Socioeconomic Factors." American Journal of Public Health, 82 (10): 1383-1385.

Winkleby, M.A., Jatulis D.E., Frank E, and Fortmann S.P. (1992). "Socioeconomic Status and Health: How Education, Income and Occupation Contribute to Risk Factors for Cardiovascular Disease." American Journal of Public Health, 82 (6): 816-820.

Winkelby, M.A., Schooler, C., Kraemer H.C., et al. (1995). "Hispanic Versus White Smoking Patterns by Sex and Level of Education." American Journal of Epidemiology, 142 (4): 410-418.

Zimmerman, R. (1993). "Social Equity and Environmental Risk." Risk Analysis, Vol. 13, No. 6: 649.

Zimmerman, R. (1994a). "An Environmental Equity Study for Inactive Hazardous Waste Sites." US EPA Region II Emergency and Remedial Response Division. New York, NY.

Zimmerman, R. (1994b). "Issues of Classification in Environmental Equity: How We Manage It and How We Measure It." Fordham Urban Law Journal, Vol. 21, No. 3: 633-670.

Appendix E. Access to Environmental and Human Health Effect Databases

The Environmental Protection Agency has extended considerable resources in order to make environmental information more available to the staff of the Agency and to the public. The nature, information content, and access methods through which such persons can access EPA databases are described in this Appendix. The methods by which such persons can access the data are:

1. Online Resources

1.1 EPA's Gopher Server, FTP Server, and World Wide Web (WWW) Server

The data available through different EPA servers are identical in content. The addresses of the EPA servers are:

EPA Gopher Server	Address: gopher.epa.gov
EPA FTP Server	Address: ftp.epa.gov
EPA Web Site Server	Address: www.epa.gov

For EJ applications the following information (at least) are available:

TRI Public Data Release Documentation and data tables as ASCII text format and Lotus spreadsheet format files.

TRI ARC/INFO Coverages For Geographical Information System (GIS) applications. GIS files available for all 50 states. TRI facility and attribute data exported as ARC/INFO coverages.

1.1.2 ENVIROFACTS Database

ENVIROFACTS is a relational database maintained by the EPA that integrates data from four major EPA program systems, as well as location information in the Facility Indexing System (FINDS). ENVIROFACTS allows the user to perform queries that integrate facility data from the four databases based on their FINDS ID (Identification/Primary Key) but it does not allow the user to down-load information to perform separate analyses. ENVIROFACTS includes monthly-updated data made available under the Freedom of Information Act. No enforcement or budget-sensitive information is contained in ENVIROFACTS. The databases accessible are:

AIRS	Aerometric Information Retrieval System
CERCLIS	Comprehensive Environmental Response Compensation and Liability Information System. CERCLIS tracks information collected under CERCLA and contains Superfund data on hazardous waste site assessment and remediation, including data on active sites from point of discovery to listing on the National Priorities List through completion of remedial and response actions. (U.S. EPA, 1995f?)
EMCI	ENVIROFACTS Master Chemical Integrator. This is a cross reference index of chemical data reported in the program systems.
FINDS	Facility Indexing System
TRIS	Toxic Release Inventory System
PCS	Permit Compliance System
RCRIS	Resource Conservation and Recovery Information System. This is used primarily to track entities regulated under RCRA Subtitle C (hazardous waste handlers). RCRIS includes data on general handler information, permit or closure status, compliance with federal and state regulations, and cleanup activities. (US EPA, 1995f)

To access ENVIROFACTS users must have Internet access. ENVIROFACTS is accessible through the EPA's Web Site at *http://www.epa.gov/enviro/*. Direct Telnet to ENVIROFACTS server is not available to the public. Online query forms are available to access selected data from ENVIROFACTS. Access to the data is available by connecting to the database directly using SQL Net software.

1.1.3 Aerometric Information Retrieval System (AIRS) Executive

The EPA's Web Site provides access to release and transfer databases, including the AIRS Executive. It is a software package designed for easy access and presentation of some of the most frequently used data in AIRS. The software package and the monthly updates may be down loaded from the EPA Web Site at *http://www.epa.gov/airs/aexec.html*. However, it does not contain the extensive air pollution data found in AIRS proper, and is not a substitute for AIRS.

1.1.4 Biennial Reporting System (BRS) Hazardous Waste Reports and Data

The BRS data are contained in self-extracting zipped flat files and may be downloaded from the EPA Web Site at *http://www.epa.gov/docs/OSWRCRA/hazwaste/data/*. At the time of writing, the current version of the report is based on the 1993 BRS data. The data files themselves are also available on the Web Site, although expanding a year's worth of BRS data requires over 200 MB of disk space. In addition, a software package such as SAS is necessary for manageable data manipulation. Hard copies of the National Biennial RCRA Hazardous Waste Report are also available from the NTIS.

1.1.5 Emergency Response Notification System (ERNS)

ERNS data can be downloaded in a flat file format from the EPA Web Site at *http://www.epa.gov/ERNS/* for off-line manipulation and analysis. In addition, the ERNS home page offers a search form to request a customized search of the ERNS database for users that do not have the expertise or tools to download and manage ERNS data off-line. At the time of this writing, a query system is under development to allow some searches of the data to be done online. Data are updated weekly. BRS and ERNS, while accessible from the EPA Web Site, are not currently linked to ENVIROFACTS.

1.2 National Library of Medicine (NLM) Toxicology Data Network (TOXNET)

The EPA utilizes the resources of the National Library of Medicine (NLM) to make information available to the public via TOXNET.

TOXNET is a system offering state-of-the-art online searching of the Toxic Release Inventory (TRI) database. It contains the complete national TRI data for all reporting years and provides access to other toxicological databases, such as EPA's Integrated Risk Information System (IRIS). TOXNET is oriented to information on hazardous effects and the toxicology of various chemical substances. It was designed to be easy-to-use by persons with limited computer experience and is currently managed by NLM's Toxicology Information Program.

Users must register, by contacting the NLM, and acquire a password for access. Costs vary according to the time a user is connected and the amount of information viewed and/or downloaded. TOXNET offers extensive online user assistance and is available 24 hours/day, 7 days/week, with the exception of a brief daily maintenance period. TOXNET may be accessed using standard telecommunications software packages, which include: Grateful-Med, CompuServe, TELENET, INFONET, or the TYMNET telecommunication networks. Internet access is also available.

TOXNET offers many modules for building and reviewing records and for search and retrievals, including free text searching. In addition, it is a gateway to other information in the Medical Literature Analysis and Retrieval System (MEDLARS) online computer system.

1.3 Right-to-Know Network (RTK NET)

RTK NET is an online service operated by the OMB Watch and the Unison Institute. RTK NET provides free public access to TRI data as well as to other environmental and governmental databases. RTK NET can be reached via Telnet, dial-up, or the Internet. Telnet and dial-up users must register with RTK NET before they can access the databases. Internet users can access RTK NET without registering (and are encouraged to

do so). RTK NET offers access to more than thirteen different health and environmental databases, including:

BRS	Biennial Reporting System
CERCLIS	CERCLA Information System (RTK NET reports 20 fields of hundreds).
CUS	Chemical Update System. Because roughly half of the database is Confidential Business Information, RTK NET asks EPA to blank out that information, but to leave the fields intact with flags.
DOCKET	Criminal and Enforcement Dockets
ERNS	Emergency Response Notification System (RTK NET updates ERNS data from the EPA Web Site every six months. Users seeking more up-to-date data should access the EPA Internet site directly, a site updated weekly).
FINDS	Facility Indexing System
PCS	Permit Compliance System
TRI	Toxic Release Inventory

Data within RTK Net is only that which is publicly accessible by other means. Most databases are complete, but note the above exceptions.

2. Digital Media

2.1 Toxic Release Inventory (TRI) CD-ROM

The resource contains the complete national coverage of the TRI, starting with the first inventory in 1987. Chemical Fact Sheets (formerly TRI-FACTS) containing reference material on the health and ecological effects of the regulated substances are also available on the same CD-ROM. Contacts for gaining a copy of the product: NTIS, Government Printing Office (GPO), Federal Depository Libraries, and EPA Regional Offices.

2.2 National Economics, Social and Environmental Data Bank (NESE-DB) CD-ROM/Diskettes

The resource is the National Economics, Social and Environmental Data Bank. It database includes the TRI state data and the national public data file on a CD-ROM. The product is produced quarterly by the Department of Commerce and provides access to socio-economic, as well as environmental statistics and other pertinent information. The data are gathered from over 15 Federal Agencies. Contacts for gaining a copy of the product: Department of Commerce, selected Federal Depository Libraries. Persons requesting diskette products can select either 5.25 or 3.5 inch diskettes by state or for the entire US in DBASE III PLUS or Lotus 1-2-3 (Version 2.0) formats. Diskettes are

accompanied by documentation. Contacts for gaining a copy of the product: NTIS and the Government Printing Office (GPO)

3. Printed Media

TRI Reports: EPA assembles several detailed annual reports providing summaries, analyses, and comparison of TRI data by year. The reports summarize data on total releases and transfers of TRI chemicals; geographic distribution of TRI releases and transfers; industrial patterns of releases and transfers; the interstate and intrastate transport of wastes and other kinds of analyses. Contact: EPCRA Information Hot Line.

4. Freedom of Information Act (FOIA)

The FOIA is another mechanism that can be utilized by the public to request information, including data, from EPA. FOIA allows a person to submit a written request for records held or believed to be held by EPA, where a record is defined as any existing document, memorandum, report, photograph, sound or magnetic recording, computer tape, drawing, or other medium in which information has been preserved. FOIA requests are usually utilized when there are no direct methods, such as Hot Lines or clearinghouses, available through which the public can obtain the information. EPA will release the requested information unless it falls under one of the following nine exemptions:

- Matters of national defense or foreign policy
- Internal Agency rules
- Information exempted by other statutes
- Trade secrets, commercial or financial information, Confidential Business Information
- Privileged inter- or intra-Agency memoranda
- Personal privacy
- Records or information compiled for law enforcement purposes
- Records of financial institutions
- Geographical or geophysical information and data concerning wells.

The FOIA authorizes EPA to charge requesters the direct cost for any searching, reviewing, and duplication required to respond to the request if these costs exceed $25.

5. National Technical Information Service (NTIS)

NTIS is a component of the Department of Commerce and is used extensively by many EPA system managers to make information available to the public. NTIS, by law, is self-supporting and sells its products and services to users on a cost-recovery basis. NTIS reproduces and sells material created by EPA which includes publications, diskettes, CD ROMs, magnetic tapes, and video tapes. To search online for environmental reports, NTIS's Bibliographic Database is offered through various commercial services. In addition, NTIS publishes a twice a month the periodical "NTIS Alert on Environmental Pollution and Control". This periodical provides summaries of newly released environmental reports and studies released by Government agencies. NTIS also offers several EPA databases on computer tape and CD-ROM. Some of the databases are text only, requiring the user to provide the search and retrieval software.

Available databases include:

CERCLIS	CERCLA Information System (Text Only)
ERNS	Emergency Response Notification System (Text Only)
FINDS	Facility Indexing System (Text Only)
IRIS	Integrated Risk Information System
ISI	Information Systems Inventory (An Inventory of Inventories!)
PCS	Permit Compliance System
RCRIS	RCRA Information System (Extracts)
TRI	Toxic Release Inventory (CD-ROM)
TSCA	Toxic Substances Control Act Chemical Substances Inventory