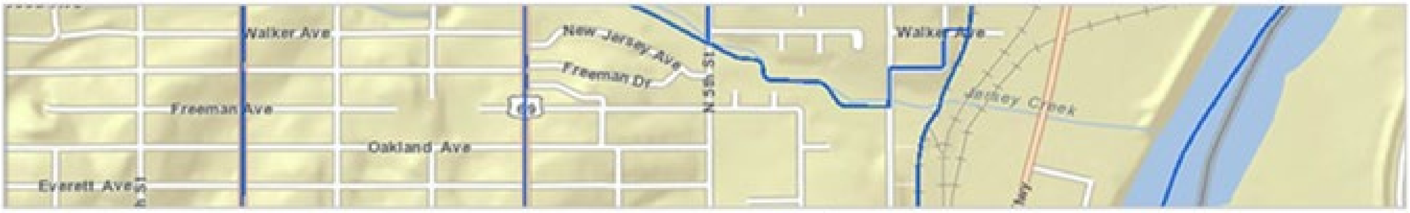

EJSCREEN

Environmental Justice Mapping and Screening Tool

EJSCREEN Technical Documentation

September 2019

EJSCREEN
Environmental Justice Mapping and Screening Tool

EJSCREEN Technical Documentation

September 2019

U.S. Environmental Protection Agency
Office of Policy
Washington, D.C. 20460

Suggested citation:

U.S. Environmental Protection Agency (EPA), 2019. EJSCREEN Technical Documentation.

For more information:
www.epa.gov/ejscreen

ACKNOWLEDGEMENTS

This document was prepared by the United States Environmental Protection Agency's Office of Policy (OP) and Office of Environmental Justice (OEJ) with drafting, editing and technical support from Abt Associates, ESRI and SAIC, working with the Office of Environmental Information (OEI), obtained and processed datasets obtained from the Census Bureau, US Department of Transportation (DOT), and various EPA offices including the Office of Air (OAR), Office of Research and Development (ORD), Office of Water (OW), and Office of Land and Emergency Management (OLEM). Ongoing input, review, and outreach have been provided by all of these offices as well as EPA's Regional offices, Office of Chemical Safety and Pollution Prevention (OCSPP), Office of International and Tribal Affairs (OITA), Office of Enforcement and Compliance Assurance (OECA), and Office of General Counsel (OGC). External expert peer reviewers provided comments on the methods and documentation as well. Specific acknowledgement is given to the contributions of Mark Corrales, Bridgid Curry, William Nickerson, Tai Lung and Fred Talcott in the design of EJSCREEN and the preparation of this document.

Contents

Appendices

Tables

EJSCREEN Environmental Justice Mapping and Screening Tool

Technical Documentation

1 INTRODUCTION

The United States Environmental Protection Agency (EPA) is charged with protecting human health and the environment for all Americans. In order to better meet the Agency's responsibilities related to the protection of public health and the environment, EPA has developed an environmental justice (EJ) screening tool, called EJSCREEN. In some ways, EJSCREEN is similar to prior screening or mapping tools. As a newer tool, however, it offers improvements such as easy web-based access to powerful mapping and data reporting tools, a wide range of updated demographic information, environmental indicators addressing more topics, and higher resolution maps covering the entire nation. EJSCREEN also provides standard reports that bring together environmental and demographic data in the form of EJ indexes. These are summarized as percentiles to put the information in perspective and facilitate comparisons between locations.

EJ screening tools may be used to explore one location using a data report, or to look across a wide area using maps. EJ tools have been used in a wide variety of circumstances, and EJSCREEN can support a similarly broad range of applications. EJSCREEN provides useful data and indicators, and highlights places that may be candidates for further review, including additional consideration, analysis or outreach.

This document describes EJSCREEN within the context of EPA's EJ program, and provides details on the data and methods used to create the indicators and indexes in EJSCREEN. The Appendices in this document provide additional detail on data and methods for interested users.

Environmental Justice at EPA

Since EJ mapping and screening is just one aspect of EPA's ongoing commitment to environmental justice, it is helpful to understand the broad, historical context of EPA's EJ work.

EPA has defined "environmental justice" as follows:

> Environmental Justice is the fair treatment and meaningful involvement of all people regardless of race, color, national origin, or income with respect to the development, implementation, and enforcement of environmental laws, regulations, and policies. … Fair treatment means that *no group of people should bear a disproportionate share of the negative environmental*

> *consequences* resulting from industrial, governmental and commercial operations or policies.[1] (italics added)

EPA's efforts to understand EJ concerns date back at least to EPA's 1992 report on Environmental Equity (U.S. EPA, 1992). The 1992 report documented health and exposure disparities associated with race/ethnicity and income. To address such disparities, in 1994, Executive Order 12898 (EO 12898) mandated that each covered federal agency make achieving environmental justice part of its mission by identifying and addressing, as appropriate, any disproportionately high and adverse human health or environmental effects of its programs, policies and activities on minority, low-income, tribal and indigenous populations.[2] These early activities provided a foundation for EPA's continued commitment to environmental justice.

EPA has been engaged in a variety of research and analytic efforts related to environmental justice, to support both regulatory analysis and screening efforts. The original EJSCREEN efforts were guided by Plan EJ 2014, which was released in September 2011.[3] EPA's Office of Environmental Justice (OEJ), in conjunction with the rest of the Agency, works to protect human health and the environment in communities overburdened by environmental pollution by integrating environmental justice into all EPA programs, policies and activities.

Environmental Justice Mapping and Screening at EPA

Mapping tools as well as screening-level applications have a substantial history at EPA, and EJ is one area in which maps and screening can be useful. Several EPA Regional offices have used basic screening tools that map demographic information and allow staff to overlay selected environmental data such as facility locations. In connection with EJSEAT, a screening tool that was developed by EPA's Office of Enforcement and Compliance Assurance (OECA), the National Environmental Justice Advisory Council (NEJAC) provided recommendations to EPA about how to design an EJ screening tool in its May 2010 report, "Nationally Consistent Environmental Justice Screening Approaches" (NEJAC, 2010).

These various early screening tools have been used for internal EPA purposes only, and generally were not available to the public. EPA's main publicly available EJ mapping tool until 2015 was EJVIEW,[4] a web-based tool that displayed selected demographic and environmental data, and allowed users to overlay these data on maps of a community or wider area.

Development of EJSCREEEN

Plan EJ 2014 included a commitment to develop a nationally consistent environmental justice screening and mapping tool in order to better meet the Agency's responsibilities related to the protection of public health and the environment in a manner that is consistent with EO 12898 and the goals of the plan. This commitment was the impetus behind the development of EJSCREEN. This new effort provided

---

[1] https://www.epa.gov/environmentaljustice/learn-about-environmental-justice , accessed 10/16/2014.
[2] http://www.archives.gov/federal-register/executive-orders/pdf/12898.pdf
[3] https://www.epa.gov/environmentaljustice/ej-2020-action-agenda
[4] https://ejscreen.epa.gov/mapper/

an opportunity to reassess and build upon prior efforts, while considering new data, new scientific findings, new analytic methods and a variety of policy considerations.

The goal in developing EJSCREEN has been to take account of this prior progress and learning, and provide a new, user-friendly screening tool that addresses policy questions and stakeholder concerns in an informative manner. An important part of this effort has been to ensure the screening tool reflects an appropriate balance between simple, feasible, screening-level information on the one hand, and high-quality data and strong science on the other.

Development of EJSCREEN began in late 2010, and EPA staff began using an early version in 2012, as prior tools such as EJSEAT were phased out. EJSCREEN was peer reviewed in early 2014 through a letter review (see Appendix G), and updated with newer data and an improved interface in 2014 and 2015. EJSCREEN was released to the public in 2015, replacing EJVIEW as EPA's public-facing EJ mapping tool. EPA released new versions of EJSCREEN with updated data and an improved interface in 2016, 2017, 2018 and 2019.

Purposes and Uses of EJSCREEN

EJ mapping and screening tools combine environmental and demographic indicators in maps and reports. This information can help to highlight geographic areas and the extent to which they may be candidates for further review, including additional consideration, analysis or outreach. The tools also allow users to explore locations at a detailed geographic level, across broad areas or across the entire nation. Environmental indicators typically are direct or proxy estimates of risk, pollution levels or potential exposure (e.g., due to nearby facilities). Demographic indicators are often used as proxies for a community's health status and potential susceptibility to pollution. Environmental and demographic data and indicators may be viewed separately or in combination.

This type of screening information may be of interest to communities as well as many other stakeholders, and also can support a wide range of research and policy goals. In general, EPA's efforts are more effective and efficient if they are informed by an understanding of where the impacts of existing pollution may be greatest. Screening tools can also help ensure that such areas are not overlooked, and receive appropriate consideration, analysis or outreach.

Screening tools can be appropriately put toward a wide variety of uses. Since its release in 2015, the public has used EJSCREEN in a variety of ways. EPA has used EJSCREEN in aspects of enforcement, compliance, the Superfund program, permitting, and voluntary programs. EJSCREEN has also been used in developing retrospective reports, and to enhance geographically based initiatives.

EJSCREEN should be used for a "screening-level" look. Screening is a useful first step in understanding or highlighting locations that may be candidates for further review. However, it is essential to remember that screening-level results do not provide a complete assessment of risk, and have significant limitations.

Caveats and Limitations of EJSCREEN

EJSCREEN is a pre-decisional screening tool, and was not designed to be the basis for agency decision-making or determinations regarding the existence or absence of EJ concerns. It also should not be used to identify or label an area as an "EJ Community." Instead, EJSCREEN is designed as a starting point, to highlight the extent to which certain locations may be candidates for further review or outreach. EJSCREEN's initial results should be supplemented with additional information and local knowledge whenever appropriate, for a more complete picture of a location. Additional considerations and data, such as national, regional, or local information and concerns, along with appropriate analysis, should form the basis for any decisions.

EJSCREEN, as a screening tool, is more limited than a detailed analysis in two key ways. First, it has data on only some of the relevant issues, and second, there is uncertainty in the data it does have. It is important to understand each of these limitations.

The first limitation arises because a screening tool cannot capture all the relevant issues that should be considered (e.g., other local environmental concerns). Any national screening tool must balance a desire for data quality and national coverage against the goal of including as many important environmental indicators as feasible given resource constraints. Many environmental concerns are not yet included in comprehensive, nationwide databases. For example, data on environmental indicators such as local drinking water quality and indoor air quality were not available with adequate quality, coverage and/or resolution to be included in this national screening tool. EJSCREEN cannot provide data on every environmental impact and demographic factor that may be important to any location.

The second important limitation is that EJSCREEN relies on demographic and environmental estimates that involve substantial uncertainty. This is especially true when looking at a small geographic area, such as a single Census block group. A single block group is often small and has uncertain estimates. A buffer that is roughly the same size as a block group or smaller will introduce additional uncertainty because it has to approximate the locations of residences. Therefore, it is typically very useful to summarize EJSCREEN data for a larger area, covering several block groups, in what is called a "buffer" report, as explained later in this document. There is a tradeoff between resolution and precision: Detailed maps at high resolution can suggest the presence of a local "hotpot," but are uncertain. Estimates based on larger areas will provide more confidence and precision, but may overlook local "hotpots" if not supplemented with detailed maps.

The demographic uncertainty combined with uncertainty in environmental data means EJ index values are often quite uncertain for a single block group. Therefore, modest differences in percentile scores between block groups or small buffers should not be interpreted as meaningful because of the uncertainties in demographic and environmental data at the block group level. We do not have a high degree of confidence when comparing or ranking places with only modest differences in estimated percentile. For this reason, it is critical that EJSCREEN results be interpreted carefully, particularly for individual block groups, and that additional information be used to supplement or follow up on screening, where appropriate.

The demographic estimates, such as percent low-income, come from the American Community Survey (ACS) from the United States Census Bureau. The ACS is comprised of surveys, not a full census of all households. This means the Census Bureau may estimate that a block group is 30% low-income, for example, but it might actually be 20% or 40% in some cases (see Appendix B for a discussion of uncertainty in demographics).

Uncertainties are also discussed in section 2 (with regard to buffer reports), and Appendix B (in discussions of buffering details and demographic data).

Related to the issue of uncertainty is that fact that the environmental indicators are only screening-level proxies for actual exposures or health risks. This is particularly true for the proximity indicators, for example. Even for the indicators that directly estimate risks or hazards, as with the air toxics cancer risk indicator, estimates have substantial uncertainty because emissions, ambient levels in the air, exposure of individuals, and toxicity are uncertain. Section 3 provides technical details on each environmental indicator.

The inclusion of a dataset in EJSCREEN does not imply it is the newest, best, or primary estimate of actual conditions or risks. Estimates are based on historical data and may not reflect current or future conditions. The vintage of environmental indicators varies and is not the same as the vintage of the demographic data. The NATA air toxics indicators and the $PM_{2.5}$ and ozone indicators in particular should be viewed with this in mind, because emissions related to $PM_{2.5}$, ozone, and air toxics generally have decreased in recent years. This version of EJSCREEN incorporated the most recent data that were available at the time of each indicator's development. Every attempt will be made to use the most recent appropriate data available in future updates of EJSCREEN. There is always a delay between the release of raw data and their eventual incorporation into any models, tools, or maps. It is also useful to note that although the raw numbers for some indicators do not represent current conditions, the percentiles are much more likely to be reasonably representative of today's conditions in most locations. This is because even if emissions have been significantly reduced overall, for example, the differences between various locations are unlikely to have changed as dramatically, especially when the reductions have come from national regulations and other trends affecting entire industries or sectors in many locations. For this reason, the percentiles may be more representative of current conditions than the raw values of the indicators. Finally, some supplementary maps and local information can complement the EJSCREEN indicators to provide more recent information. In particular, EJSCREEN also provides updated maps of $PM_{2.5}$ and ozone nonattainment areas (areas not meeting national ambient air quality standards).

There are also some limitations in geographic coverage -- EJSCREEN lacks data in some locations for some indicators, such as in Alaska, Hawaii, and Puerto Rico.

In short, as with any screening tool, the indicators in EJSCREEN cannot address all the considerations that may be relevant to a given situation, they are often only a screening-level proxy for a given issue, and in any case, there is significant uncertainty involved, particularly for a single block group. For these types of reasons, among others, it generally is not appropriate to rely on any screening tool as the basis for a key decision. It is often very useful to obtain information on other issues not included in EJSCREEN, updated information when available, as well as local knowledge, data, and concerns.

Advances in EJ Screening Provided by EJSCREEN

EJSCREEN offers a variety of enhancements relative to previous approaches to EJ mapping and screening or analysis. For example, the tool includes updated demographic information (from the ACS) rather than relying on the Census that is conducted every ten years. It also provides several new environmental indicators, covering a wider range of issues such as traffic volume and proximity. EJSCREEN includes a suite of EJ indexes that quantify the combination of environmental and demographic indicators. It includes high resolution maps, and a new geospatial software and data system that improves access to new tools with a simple, browser-based interface and centralized, consistent data. The buffer reports are calculated using detailed Census block data for more accurate estimates of where residents are located. EJSCREEN provides access to a great deal of data, and presents standardized reports. These reports use summary metrics and percentiles to facilitate national, regional or state-level perspectives and a better understanding of EJ issues.

EJSCREEN can help explore the environmental, demographic and EJ characteristics of a block group or buffer area. It provides numerical estimates for each place, for both environmental and demographic data, such as the traffic proximity indicator, or the percentage of local residents who are racial/ethnic minorities.

EJSCREEN also presents multiple “EJ Indexes” for each place. An EJ Index is a way of combining, in a quantitative way rather than only visually on a map, the environmental indicator and the demographic information for a location. A separate EJ Index is provided for each environmental indicator in EJSCREEN, for each block group in the US. The EJ Index goes beyond a simple visual overlay of maps of environment and demographics, to actually quantify the extent to which these two factors co-occur.

In EJSCREEN, the basic level of geographic resolution is the Census block group. Each block group is defined by the U.S. Census Bureau, with a logical and unambiguous numbering scheme, and associated digital shape files that permit mapping with modern geographical information system (GIS) software. Block group data are widely used by researchers and others. Block groups also provide a relatively stable framework; for instance, block groups are not subject to frequent boundary definition changes that political jurisdictions and postal ZIP codes may experience.

Estimates in EJSCREEN are compiled by block group, and that is the most detailed level at which results can be viewed. However, demographic estimates for a single block group are often based on a small sample of the local population, and are uncertain. Similarly, some environmental indicator estimates are derived from lower-resolution data, and all involve uncertainty. Therefore, it is typically very useful and advisable to summarize EJSCREEN data within a larger area that covers several block groups, in what is called a “buffer” report. An EJSCREEN user can specify or draw buffers of custom sizes and shapes as needed. For example, a buffer could include all residents within 1 mile of a certain location. When a buffer covers several block groups, it provides an estimate that has less uncertainty than a single block group or smaller buffer would. EJSCREEN summarizes data for all residents within some distance from a selected point, using a circular buffer, or within a user-defined buffer of any shape, using Census *blocks* (not just block groups) to refine estimates of how many residents are inside the buffer, as explained in Appendix B.

2 OVERVIEW OF DATA AND METHODS IN EJSCREEN

This section describes the environmental and demographic data used in the tool, as well as the methods used to combine them and produce EJ Indexes. As of 2017, EJSCREEN contains 11 environmental indicators, which range from estimates of human health risk to proxies for potential exposure such as proximity to hazardous waste sites. The tool also contains six demographic indicators and a demographic index. The EJSCREEN demographic index uses the average of two indicators, low-income and minority and is combined with an environmental indicator to create the associated EJ Index. The environmental, demographic, and EJ indicators and indexes are all calculated for each block group, and can be summarized within a defined buffer area. The sections immediately below summarize the environmental data, demographic data, and EJ indexes. Section 3 provides more detail on each environmental indicator.

Environmental Indicators Selected for EJSCREEN

Some environmental indicators used in EJSCREEN quantify proximity to and the numbers of certain types of potential sources of exposure to environmental pollutants, such as nearby hazardous waste sites or traffic. The lead paint indicator indicates the presence of older housing, which often, but not always, indicates the presence of lead paint, and therefore the possibility of exposure. In some cases, the term "exposure" is used very broadly here to refer to the potential for exposure. Others indicators in EJSCREEN are estimates of ambient levels of air pollutants, such as PM2.5, ozone, and diesel particulate matter. Still others are actual estimates of air toxics-related cancer risk, or a hazard index, which summarizes the ratios of ambient air toxics levels to health-based reference concentrations. In other words, these environmental indicators vary widely in what they indicate, as discussed further below.

A variety of considerations has informed the selection of these environmental indicators; in general, the selected indicators exhibit the following characteristics:

- Resolution: Screening level data are available (or could be readily developed) at the block group level (or at least close to this resolution).
- Coverage: Screening level data are available (or could be readily developed) for the entire United States (or with nearly complete coverage).
- Relevance to EJ: Pollutants or impacts are relevant to EJ (e.g., differences between groups have been indicated in exposures, susceptibility, or health endpoints associated with the exposures)
- Public health significance: Pollutants or impacts are potentially important in the United States (e.g., notable impacts estimated or significant concerns have been expressed, at least locally, or exposure has been linked to health endpoints with substantial impacts nationwide).

EPA selected environmental indicators after a review of data availability (including the criteria and review of data availability for an Environmental Quality Index, now provided by county in Messer, Jagai, Rappazzo, & Lobdell (2014), and described in Lobdell, Jagai, Rappazzo, & Messer (2011)); health disparity information (e.g., CDC's major 2011 report on health disparities (Centers for Disease Control and Prevention, 2011a)); risk ranking studies (e.g., Unfinished Business (U.S. EPA, 1987), Reducing Risk (U.S. EPA, 1992), and related reports); and risk estimates from major studies (e.g., related to $PM_{2.5}$ ambient standards). EPA also reviewed data from the CDC (Centers for Disease Control and Prevention, 2011c) and other sources in the federal government (Fedstats, 2007), and consulted EPA Regions and program offices that are responsible for data collection and analysis under EPA's key environmental statutes. The State of California's work on CalEnviroScreen[5] was also tracked throughout its development. Other internal EPA data tools were also examined, including EJSEAT, EJVIEW, C-FERST, and tools used in EPA Regional Offices.

After review, EPA selected the following environmental indicators for use in the public version of EJSCREEN:

- Air pollution:
    - $PM_{2.5}$ level in air.
    - Ozone level in air.
    - NATA air toxics:
        - Diesel particulate matter level in air.
        - Air toxics cancer risk.
        - Air toxics respiratory hazard index.
- Traffic proximity and volume: Amount of vehicular traffic nearby, and distance from roads.
- Lead paint indicator: Percentage of housing units built before 1960, as an indicator of potential exposure to lead.
- Proximity to waste and hazardous chemical facilities or sites: Number of significant industrial facilities and/or hazardous waste sites nearby, and distance from those:
    - National Priorities List (NPL) sites.
    - Risk Management Plan (RMP) Facilities.
    - Hazardous waste Treatment, Storage and Disposal Facilities (TSDFs).
- Wastewater discharge indicator: Proximity to toxicity-weighted wastewater discharges

Each of these environmental indicators is explained in detail in section 3, and Appendix D provides summary statistics for the indicators. Again, it is important to understand that these indicators vary in how relevant they are to actual estimated risks to health or welfare, and how significant those impacts may be. These indicators represent a spectrum in terms of the quality of information about potential impacts, ranging from direct estimates of risk to rough indicators of proximity or exposure to pollution or other environmental hazards. Table 1 provides more detail on how closely each environmental indicator in EJSCREEN approximates actual estimated risk.

[5] http://oehha.ca.gov/ej/ces2.html at http://oehha.ca.gov/ej

Table 1. Types of Environmental Indicators Included in EJSCREEN

Indicator	Place on Exposure–Risk Continuum	Key Medium
NATA Air Toxics Cancer Risk Lifetime inhalation cancer risk	Risk/Hazard	Air
NATA Respiratory Hazard Index Ratio of exposure concentration to RfC		
NATA Diesel PM (DPM) (µg/m^3)	Potential Exposure	
Particulate Matter ($PM_{2.5}$) Annual average (µg/m^3)		
Ozone Summer seasonal average of daily maximum 8-hour concentration in air (ppb)		
Lead Paint Percentage of housing units built before 1960		Dust/ Lead Paint
Traffic Proximity and Volume Count of vehicles (average annual daily traffic) at major roads within 500 meters (or nearest neighbor outside 500 meters), divided by distance in kilometers (km)	Proximity/ Quantity	Air/ Other
Proximity to RMP Sites Count of facilities within 5 km (or nearest neighbor outside 5 km), divided by distance		Waste/ Water/ Air
Proximity to TSDFs Count of major TSDFs within 5 km (or nearest neighbor outside 5 km), divided by distance		
Proximity to NPL Sites Count of proposed and listed NPL sites within 5 km (or nearest neighbor outside 5 km), divided by distance[6]		
Wastewater Discharge Toxicity weighted stream concentrations divided by distance in kilometers (km)		Water

Abbreviations:

NATA — National Air Toxics Assessment
NPL — National Priorities List, Superfund program
RMP — Risk Management Plan
TSDFs — Hazardous waste Treatment, Storage, and Disposal Facilities
RfC — Reference concentration from EPA's Integrated Risk Information System
$PM_{2.5}$ — Particulate matter (PM) composed of particles smaller than 2.5 microns
µg/m^3 — micrograms of $PM_{2.5}$ per cubic meter of air
ppb — parts per billion, of ozone in air

[6] Count of NPL sites excludes deleted sites and sites in U.S. territories.

It is also important to note that each proximity indicator focuses on one category of facility or site (e.g., NPL), but the category's facilities or sites vary in the degree to which they could actually pose risks. They vary in the amount of emissions (if any), the possibility of exposure to any pollutants released, the size of the facility or site, and toxicity of the pollutants or severity of the impacts that might occur. As a screening tool, EJSCREEN generally does not distinguish based on these factors in proximity indicators (although the NATA and Wastewater Discharge indicators do account for such information). Any closer review of a particular location would have to consider these important differences.

All of these indicators are focused on potential impact at residential locations (e.g., proximity of residence to traffic), and therefore only address some of the exposures that individuals may face. Data are generally insufficient to readily estimate exposures away from the home, particularly in a screening tool. Exposures that occur away from the home, such as at work, at school or during a commute, are not captured in EJSCREEN unless those exposures are near the home or in other locations that happen to have the same level of exposure.[7]

Demographics in EJSCREEN

This section describes why demographic indicators are included in EJSCREEN, which specific demographic indicators were selected, and what data are used to derive the demographic indicators.

Using Demographics as Proxies for Potential Susceptibility

EJSCREEN has been designed in the context of Executive Order 12898[8] which ordered the following:

> 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[9]

EJSCREEN was also designed in the context of EPA's EJ policies, including EPA's Guidance on Considering Environmental Justice During the Development of an Action (U.S. EPA, 2010). That guidance document explained EPA's focus on demographics as an indicator of potential susceptibility or vulnerability to environmental pollution:

> To help achieve EPA's goal for EJ (i.e., the fair treatment and meaningful involvement of *all* people), EPA places particular emphasis on the public health of and environmental conditions affecting minority, low-income, and indigenous populations. In recognizing that these populations frequently bear a disproportionate burden of environmental harms and risks … EPA works to protect them from adverse public health and environmental effects of its programs.

[7] A partial exception is the data from NATA, which make some attempt to include some nonresidential exposures, as explained in NATA's technical documentation (http://epa.gov/nata/).
[8] https://www.epa.gov/laws-regulations/summary-executive-order-12898-federal-actions-address-environmental-justice
[9] http://www.archives.gov/federal-register/executive-orders/pdf/12898.pdf

> EPA should pay particular attention to the vulnerabilities of these populations because they have historically been exposed to a combination of physical, chemical, biological, social, and cultural factors that have imposed greater environmental burdens on them than those imposed on the general population. (U.S. EPA, 2010, p. 4)

EJSCREEN uses demographic indicators as very general indicators of a community's potential susceptibility to the types of environmental exposures included in this screening tool. Impacts of pollutants depend on a combination of exposure and susceptibility to those exposures. Demographic factors may be related to both of these. Therefore, it is very useful to distinguish between 1) the fact that some demographics are associated with higher exposure, and 2) the fact that demographics are useful in predicting susceptibility to those exposures. To indicate potential exposures, EJSCREEN uses environmental indicators, not demographics. EJSCREEN uses demographics to indicate potential susceptibility. EJSCREEN then combines the exposure and susceptibility indicators in the form of an EJ Index.

The demographic indicators in EJSCREEN are a way to indicate which communities may be more susceptible to a given level of exposure to environmental pollutants. For example, individuals may be more susceptible when they are already in poor health, have reduced access to care, lack resources or language skills or education that would help them avoid exposures or obtain treatment, or are at susceptible life stages. Nationwide direct measures of health status are not available for all block groups or even tracts – such data are typically compiled by county in national databases. Demographics, however, are available for every block group, and are correlated with health status and these other susceptibility factors, making them useful screening-level indicators of potential susceptibility at the local level.

Note that this report uses the term susceptibility in a qualitative, general sense, to refer to what various authors have called susceptibility and/or vulnerability. Susceptibility in this report means greater "impact" for a given environmental indicator value. The terms vulnerability and susceptibility sometimes are used interchangeably, although various other reports and programs have made distinctions between these terms.[10]

The relationships between demographics, exposure, and susceptibility are complex. For example, demographics may be associated with susceptibility to pollutants in any of the following ways:

- Greater personal exposure despite the same ambient level of pollutant. For example, children have higher breathing rates or ingest more lead dust than adults (U.S. EPA, 2011a),

[10] For example, EPA's 2009 National Ambient Air Quality Standards (NAAQS) documents (U.S. EPA, 2009b) and also EPA's Regional Vulnerability Assessment program treat susceptibility and vulnerability as essentially identical. Other EPA definitions have addressed particular contexts, such as in EPA glossaries (http://www.epa.gov/OCEPAterms/vterms.html, http://www.epa.gov/OCEPAterms/sterms.html), and a report on vulnerability to climate change (U.S. EPA, 2009a). One National Academies report (Science and Decisions) distinguished between the two terms (National Research Council, 2009). In other contexts, the terms have varied uses – see Villagrán de Léon (2006) for a detailed comparison of various definitions of vulnerability in the context of natural disasters, or work on vulnerability indexes for developing countries by Briguglio (1997).

and certain groups may tend to encounter or be less able to avoid certain exposures due to limited resources, language barriers, education, cultural practices, or lack of information.

- Susceptibility because of a greater percentage increase in health risk for a given exposure, e.g., “effect modification” or “multiplicative interaction” may occur. An example would be where cumulative previous exposure means a group is more likely to be closer to a threshold for adverse effects, or where greater stress/allostatic load increases susceptibility through inflammatory or other pathways. Several examples of effect modification relevant to EJ and $PM_{2.5}$ are referenced by Bell & Ebisu, 2012 and in a review of subgroups susceptible to ozone (Bell, Zanobetti, & Dominici, 2014). A growing body of research has documented interactions of psychosocial stress and environmental exposures.
- Susceptibility because of higher baseline risk or rates of pre-existing diseases. The same percent increase in mortality risk has a larger impact on absolute risk if baseline risk is higher.
- Susceptibility because of increased overall burden resulting from an initial health risk (e.g., because of less ability to recover due to lack of health care or resources). For example, low-income or minority individuals, or those with less than a high school education, are far less likely to have health insurance (Cohen & Martinez, 2011).

One reason for EJSCREEN to focus on potentially susceptible demographic groups is that a large body of research has documented health disparities between demographic groups in the United States, such as differences in mortality and morbidity associated with factors that include race/ethnicity, income and educational attainment (e.g., Centers for Disease Control and Prevention, 2011a; Galea, Tracy, & Hoggatt, 2011). For example:

- About two thirds (65%) of non-Hispanic white adults reported excellent or very good health in 2009. In contrast, less than half (49%) of non-Hispanic black adults and 52% of Hispanic adults reported excellent or very good health.[11]
- Residents with lower income report fewer average healthy days than others (Centers for Disease Control and Prevention, 2011a), and report worse health overall (Centers for Disease Control and Prevention, 2010).
- Both lower income and minority race/ethnicity have independent associations with higher asthma rates, particularly among children, and diabetes rates differ greatly by race/ethnicity (Centers for Disease Control and Prevention, 2011a).
- Mortality rates for cancer and heart disease vary somewhat by race/ethnicity (Centers for Disease Control and Prevention, 2011b). Coronary heart disease and stroke are elevated among black individuals but generally not in other minority subgroups (Centers for Disease Control and Prevention, 2011a).

[11] This survey, the National Health Interview Survey, represents the U.S. non-institutionalized civilian population, and presents age-adjusted estimates based on household interviews (Centers for Disease Control and Prevention, 2011d).

- Infant mortality is higher among non-Hispanic black, American Indian/Alaska Native, and Puerto Rican (but not other Hispanic) populations (Centers for Disease Control and Prevention, 2011a).

While some health disparities are due to differences in health care, diet, activities, psychosocial stress or even genetics, it is possible that some portion of certain disparities may be related to differences in environmental exposures. Some of these differences in exposure are associated with residential location, and could be considered in EJSCREEN (while others cannot be considered in EJSCREEN, such as those related to use of consumer products or diet, for which high-resolution geographic data are not available). Various environmental exposures have been shown to vary by race/ethnicity, income and other demographic factors (Liu, 2001; Maantay, Chakraborty, & Brender, 2010; U.S. EPA, 2006a), but EJSCREEN is not predicated on an assumption of such a correlation.

In addition to Executive Order 12898's call, perhaps the most important reason to focus on key demographic groups in EJSCREEN is that a growing body of research has shown that demographic factors are associated with susceptibility – certain groups are more impacted by a given level of exposure to certain pollutants. Various groups have shown increased susceptibility to certain pollutants, but further evidence is still emerging in this area and data are limited. Evidence currently available includes the following:

- Certain demographic groups, such as those with lower educational attainment, children, the elderly and those with low socio-economic status (SES), appear to be more susceptible to a given exposure to particulate matter (U.S. EPA, 2009b).
- Blood lead's association with cardiovascular outcomes appears to be stronger among Mexican Americans and non-Hispanic blacks than non-Hispanic whites (U.S. EPA 2011c).
- Some but not all studies suggest lead has a greater impact on IQ among low SES than high SES individuals (U.S. EPA 2011c).

EJSCREEN is not designed to explore the root causes of differences in exposure. The demographic factors included in EJSCREEN are not necessarily causes of a given community's increased exposure or risk. This does not limit their usefulness for the limited purposes of the screening tool, however – these demographic factors are still useful as indicators of potential susceptibility to the environmental indicators in EJSCREEN. They may be associated with susceptibility, whether or not they are causal, and can be used as proxies for other harder-to-measure factors that would better describe or determine susceptibility but for which nationally consistent data are not available. EJSCREEN screens geographic areas for increased potential for exposure and increased potential for susceptibility to exposures. Additional analysis is always needed to explore any underlying reasons for differences in susceptibility, exposure or health.

Some studies have begun to quantify the degree of susceptibility to specific pollutants in particular demographic groups, such as the work on educational attainment as an effect modifier for risks associated with $PM_{2.5}$ exposure, but such emerging knowledge is still limited to a handful of pollutants and demographic factors (U.S. EPA, 2009b). EJSCREEN, as a screening tool, does not attempt to use this type of emerging quantitative information.

Demographic Indicators Included in EJSCREEN

A wide range of demographic descriptors have been used by researchers and in EJ screening tools to represent the "social vulnerability" characteristics of a disadvantaged population (for example, see deFur et al., 2007, and Bell & Ebisu, 2008).

Executive Order 12898, addressing EJ issues, refers to low-income and minority populations. We define these two core factors as:

- Low-Income: The number or percent of a block group's population in households where the household income is less than or equal to twice the federal "poverty level."[12]
- Minority: The number or percent of individuals in a block group who list their racial status as a race other than white alone and/or list their ethnicity as Hispanic or Latino. That is, all people other than non-Hispanic white-alone individuals. The word "alone" in this case indicates that the person is of a single race, since multiracial individuals are tabulated in another category – a non-Hispanic individual who is half white and half American Indian would be counted as a minority by this definition.[13]

Based on a review of other factors used in various EPA EJ screening tools, the four other factors most commonly used by EPA Headquarters and Regions for EJ analyses are also included in EJSCREEN. The other four factors are:

- Less than high school education: The number or percent of people age 25 or older in a block group whose education is short of a high school diploma.
- Linguistic isolation: The number or percent of people in a block group living in linguistically isolated households. A household in which all members age 14 years and over speak a non-English language and also speak English less than "very well" (have difficulty with English) is linguistically isolated.
- Individuals under age 5: The number or percent of people in a block group under the age of 5.
- Individuals over age 64: The number or percent of people in a block group over the age of 64.

---

[12] More precisely, percent low-income is calculated as a percentage of those for whom the poverty ratio was known, as reported by the Census Bureau, which may be less than the full population in some block groups. More information on the federally-defined poverty threshold is available at http://www.census.gov. See Appendix B for details on using twice the poverty threshold.

[13] Census definitions of race/ethnicity are available at: http://www.census.gov/topics/population/race/about.html

The source of all demographic data used in EJSCREEN is the American Community Survey (ACS) five-year summary file, which the U.S. Census Bureau compiles yearly.

Appendix D provides summary statistics for the demographic indicators.

The supplementary reports and maps provided by EJSCREEN also include an extensive list of additional demographic variables, including statistics on race/ethnicity subgroups (e.g., percent Hispanic or Latino), languages spoken (e.g., % speaking Vietnamese), income (% in poverty), and many other factors. This supplementary information may be very useful. For example, subgroups within the broad category of “minority” can differ greatly in their baseline health, exposures, geographic locations, and other factors.

Demographic Indexes in EJSCREEN

The Demographic Index in EJSCREEN is created using the two demographic indicators that were explicitly named in EO 12898, low-income and minority. For each Census block group, these two indicators are simply averaged together.

Demographic Index = (% minority + % low-income) / 2

Users can also view each demographic indicator separately in EJSCREEN, in reports or in maps.

The Demographic Indexes count each indicator as adding to overall potential susceptibility of the population in a block group, and assumes the demographic indicator have equal and additive impacts. The current lack of available data precludes any attempt to disentangle the different influences of the individual demographic indicators, or quantify the degree of overlap or potential synergy between them.

The demographic groups in EJSCREEN overlap to some extent, because some individuals are both low-income and minority, for example. In fact, these indicators are correlated at the block group level, because minorities are more likely to be low-income than non-minorities. Appendix D has information on the correlations between these variables. These correlations do not affect the indicator’s ability to account for susceptibility, if the assumption of additive effects on susceptibility is appropriate. As more data becomes available in the future, some of these complexities can be reexamined.

Additional information about the demographic data used in EJSCREEN is available in Appendix B.

Environmental Justice Indexes in EJSCREEN

The EJ index is a combination of environmental and demographic information. The environmental portion of the EJ index is drawn directly from the environmental indicators described above, and the demographic information is also taken from the demographic indicators above.

How the EJ Index Works

To calculate a single EJ Index, EJSCREEN combines a single environmental indicator with demographic information. It considers the extent to which the local demographics are above the national average. It does this by looking at the difference between the demographic composition of the block group, as measured by the Demographic Index, and the national average (which is approximately 35%). It also considers the population of the block group.

Mathematically, the EJ Index is constructed as the product of three items, multiplied together as follows:

EJ Index =

(Environmental Indicator)
X (Demographic Index for Block Group –Demographic Index for US)
X (Population count for Block Group)

The demographic portions of the EJ Index can be thought of as the additional number of susceptible individuals in the block group, beyond what you would expect for a block group with this size total population.

"Susceptible" or "potentially susceptible individuals" are used informally in these examples, as a way to think of the Demographic Index times the population count in a block group, which is essentially the average of the count of minorities and count of low-income individuals.[14] It is easiest to think of the average of these counts as "the susceptible individuals" in these examples.

The number of potentially susceptible individuals (Demographic Index times population count) of course is typically less than the actual number who are minority, low-income, or both. The demographic breakdown is not reported by block group –the ACS does not provide that level of resolution on the

[14] To be precise, the percent low-income times population is not always exactly the same as the count of low-income residents. The percent low-income is calculated as a fraction of those for whom poverty status could be determined, which is less than the full population in some block groups. For simplicity, these examples omit that detail.

overlaps.[15] For example, suppose that in a certain block group of 1000 people, 350 (35%) are minority and 350 (35%) low-income. There might be 200 (20%) who are low-income but not minority, and 200 (20%) who are minority but not low-income. In that case, there would be 150 (15%) who are both, and 450 (45%) who are neither. Therefore, there actually would be 550 (55%) who were either minority, low-income, or both. The Demographic Index would use 35% in this case, which falls between the 15% who were both minority and low-income, and the 55% who were in at least one of these groups. These detailed numbers cannot be obtained from the ACS by block group. Therefore, to represent both groups in a simple way, the average is used.

An extreme example shows another situation: Suppose a block group has 1000 people but is 0% minority and 100% low-income. The demographic index would be 50%, or the equivalent of 500 "potentially susceptible individuals" in this case. The same would be true in a block group that was 100% minority but 0% low-income – it would be treated as having the equivalent of 50% (500) "potentially susceptible" for the sake of these examples.

The EJ Index uses the concept of "excess risk" by looking at how far above the national average the block group demographics are. For example, assume a block group with 1000 people in it. In that block group, one would expect 350 potentially susceptible individuals (1000 people here x US average of 35%). However, if the Demographic Index for that block group is 75%, well above the US average, then there are the equivalent of 750 potentially susceptible people in that block group, or 400 more than expected for a block group with a population of 1000. The EJ Index would be 400 times the environmental indicator in this case.

This formula for the EJ Index is useful because for each environmental indicator it finds the block groups that contribute the most toward the national disparity in that environmental indicator. By "disparity" in this case we mean the difference between the environmental indicator's average value among certain demographic groups and the average in the US population.

Minority and low-income individuals live in older housing more often than the rest of the US population, for example. The EJ Index for lead paint (pre-1960 housing) tells us how much each block group contributes toward this "excess population risk" or "excess number" of people in older housing, for potentially susceptible individuals. "Excess" in this context simply means the number of potentially susceptible individuals in older housing nationwide is above what it would be if they were in older housing at the same rate as the rest of the U.S. population. Locally, it also means the number is above what it would be if the block group had the same demographic percentages as the U.S. overall.

Analysis of the EJSCREEN data for minority, or for low income, individuals (roughly one third of the US population in either case), shows they have a higher environmental indicator value on average than the rest of the U.S. population, for 10 of the 11 environmental indicators (ozone is the exception).

[15] The closest available data would be table B17001 and related tables, which provide tract resolution cross-tabulations of race/ethnic groups by poverty status, but this is not available for block groups and does not provide the income to poverty ratio data needed to calculated "low-income" as defined in EJSCREEN.

Note that the EJ Index raw value itself is not reported in EJSCREEN reports– it is reported in percentile terms, to make the results easier to interpret. If one is calculating the actual raw values using the formula, it is clear that the EJ Index value can be a positive or negative number. A positive number occurs where the local Demographic Index is above the US average, and this means the location adds to any excess in environmental indicator values among the specified populations (minority and low-income) nationwide. A negative value occurs where the local Demographic Index is below the US average, and it means the location offsets the other locations, reducing any excess in nationwide average environmental indicator values among minority and low-income populations relative to others. Most EJSCREEN users will not work directly with EJ Index raw values, however, and positive raw values for an EJ Index will be presented as higher percentiles and negative raw values will appear as lower percentiles.

EJ Indexes, Population Density, and Rural Areas

It is very important to understand that the population count per block group is not the same as population density, so the population weighting of the EJ Index has nothing to do with whether a place is high density, or rural versus urban. In fact, there is almost no correlation between population count per block group and population density (population count per square mile covered by the block group), because in low density areas each block group covers a larger area, keeping the population per block group fairly consistent. This means population weighting in the EJ Index does not emphasize urban or high density locations – it is neutral with regard to population density or urbanization. Furthermore, the vast majority of block groups in the US have similar population counts, so the population weighting in the EJ Index has a strong influence only in a tiny fraction of locations. For example, about 90% of block groups had a population between 500 and 2500 in 2010-2014, and only about 1% had a population over 4000.

It is true that many of the EJ Indexes have higher values in urban, high-density areas, but this not the result of population weighting. Differences in environmental indicator values (and to some extent percentage demographics) are generally the drivers of higher EJ Indexes in urban or high-density block groups. NATA indicators and the lead paint indicator, in particular, are strongly correlated with population density, as are the $PM_{2.5}$ and traffic indicators to some extent. The proximity indicators are also positively (but weakly) correlated with population density. In other words, these environmental indicators appear to be lower in rural areas in general, and combined with some demographic differences, this tends to make the EJ Indexes lower in those areas. Relative to those factors, the population weighting (or choice of EJ Index formula) has a very small influence on whether urban or rural areas are highlighted.

Why the EJ Indexes are Not all Combined

For each environmental indicator, one standard EJ index is available in EJSCREEN. At this time, there is not a single composite EJ index that combines all the environmental indicators. Although it would be useful if a simple metric could summarize all of the information in EJSCREEN as a single number, there is no widely-accepted, objective way to combine the differing environmental concerns into one number. This is because of the value judgments and scientific challenges inherent in deciding how much weight or importance should be given to each of the environmental indicators. They are very difficult to compare, in terms of public health importance, public concerns, and the many other important considerations that could be weighed. This topic has been covered extensively elsewhere[16], but a very brief explanation may be useful here.

First, a so-called "equal weighting" does not exist, because it would just be an artifact of the units (scaling) and aggregation method one chose, which would carry implicit value judgments about how to weight and combine the factors, even if it seemed simple at first glance. Putting equal weight on each percentile, for example, would implicitly equate very low risks and much higher risks (e.g., $PM_{2.5}$ levels well above a health based standard).

Furthermore, while the use of percentiles provides useful perspective by putting the 11 EJ indexes in common units, it would be a mistake to assume the 80th percentile, for example, has the same "importance" for one index or indicator as for another. If two indexes are at the same percentile, it simply means those two scores are equally common (or equally rare) in the United States. It does not mean the risks are comparable. It is therefore critical when interpreting EJSCREEN percentiles to also look at the actual raw numbers for the environmental and demographic indicators.

The challenge is compounded by the fact that rankings of block groups using a composite environmental index would be quite sensitive to the method chosen to combine the environmental indicators, based on EPA's analysis of the data. This is also the case, albeit to a lesser degree, for any composite EJ index. This is a result of the environmental indicators not being highly correlated with each other. The locations with the highest $PM_{2.5}$ levels are not usually the same as the ones with the highest NPL proximities, for example, as suggested in Appendix D. If the NPL indicator is treated as more important, different block groups would be highlighted than if the $PM_{2.5}$ indicator were given more weight. Again, it is important to acknowledge that there is no objective version of "equal weighting."

For these reasons, the environmental indicators are not combined as a single number, and must be understood individually for a complete picture. However, they can be viewed all at the same time in a single tabular report, and this facilitates a broad perspective on all the factors at one time. Those using EJSCREEN and considering aggregating the data as a single summary metric are strongly urged to carefully consider these pitfalls associated with doing so. A thorough understanding of each indicator and the ability to view all of them in a report provides a far better picture of the screening results than any single number or map is capable of.

[16] See, for example, OECD, 2008, or Finkel & Golding, 1994.

Percentiles

What a Percentile Means

EJSCREEN puts each indicator or index value in perspective by reporting the value as a percentile. For example, an area may show 60% of housing was built prior to 1960. It may not be obvious whether this is a relatively high or low value, compared to the rest of the nation or in the state. Therefore, EJSCREEN also reports that 60% pre-1960 puts this area at the 80th percentile nationwide. For a place at the 80th percentile nationwide, that means 20% of the US population has a higher value.

A percentile in EJSCREEN tells us roughly what percent of the US population lives in a block group that has a lower value (or in some cases, a tied value). This means that 100 minus the percentile tells us roughly what percent of the US population has a higher value. This is generally a reasonable interpretation because for most indicators there are not many exact ties between places and not many places with missing data.

More precisely, the exact percentile for a given raw indicator value is calculated as the number of US residents of block groups with that value or lower, divided by the total population with known indicator values. This is typically the same as or almost exactly the same as dividing by the total US population, but for some indicators some locations do not have an indicator value. For example, the NATA indicators missing for only about one twentieth of 1% of the US population in the 2017 version of EJSCREEN. The calculated percentile would change by much, much less than 1 percentile point if calculated as a fraction of the total population instead of as a fraction of those with valid indicator values.

All percentiles in EJSCREEN are population percentiles, meaning they describe the distribution of block group indicator scores across the population. Note that a population percentile may be slightly different than the unweighted percentile (the percent of block groups, not people, with lower or tied values), because not all block groups have the same population size. In practice they are very similar because very few block groups diverge very much from the average in population size.

Color-coded High Percentile Bins

Locations at least at the 80th percentile but less than the 90th are shown in yellow on EJSCREEN maps, while those at the 90th percentile but less than 95th percentile are orange on the maps, and those at the 95th percentile or above are shown in red on maps and reports. These colors call attention to certain locations as a very simple way to communicate relative screening results. There is no official policy significance assigned to each individual color on the maps, but the choice of these categories or “bins” is noteworthy because it signifies that certain ranges of percentiles may merit closer attention.

Percentiles at or above the 95th percentile are shown in red on the EJSCREEN standard report. This is a way to call particular attention to those cases where the value is in the top 5% of the nation (or region or state). Indicator or index values in the top 5% tend to be much higher than those in the next 5-10%, so they may merit close attention. This is especially true for the indicators with highly skewed distributions, such as the traffic proximity indicator (see Appendix D, Table 6 and Figure 1). For example, block groups in the top 5% (shown in red on maps and reports) have traffic, NPL, and TSDF proximity indicators on average that are about three times as high as in the next 5% (shown in orange on the maps). These differences are far less extreme in the cases of $PM_{2.5}$ and lead paint indicators, which don't vary as much across block groups. In general, though, indicator or index values above the 95th percentile represent much higher demographic, environmental, or EJ Index values than those at lower percentiles.

The maps also identify areas in the 90th to 95th percentiles as orange, and those at the 80th to 90th as yellow. These additional categories highlight larger groups of locations that have indicator or index values well above the national mean or median for the given indicator or index. The actual values are lower than those in the top 5%, typically much lower, but they are still in the top 10 to 20% of values for the US population overall.

A relatively high percentile means the value is relatively uncommon. However, a high percentile is not necessarily a real concern from a health or legal perspective. To understand the actual health or other implications of any screening results requires looking at the actual data and the indicator represents, and also looking at other relevant data if available. Besides the percentile, other important considerations in interpreting any screening results include the following:

1. whether and to what extent the environmental data shows values above any relevant health-based or legal threshold,
2. the significance of any such thresholds, or the magnitude and severity of the health or other impacts of the given environmental concern, nationally or locally, and
3. the degree of any disparity between various groups, in exposures to the relevant environmental pollutants.

In maps, EJSCREEN focuses on the US percentiles, as a way to visualize all results in common units.

The US percentile uses the US population as the basis of comparison. The state or regional percentile was calculated based on the population in a given state (or DC) or one of EPA's 10 regions.[17] The national or state or regional mean value was calculated as the population weighted average of the block groups with data for that indicator, within the respective geographic scope.

Note that the US and state percentiles both will rank block groups in exactly the same rank order within the given state. If the goal is just to rank or compare locations within a single state, it does not matter whether the US or state percentile is used. The difference between state and US percentiles becomes

[17] Regions in the 2019 version of EJSCREEN were defined only by State. A small number of block groups on Tribal Lands along the borders of Nevada and Arizona are actually part of EPA Region 9, even though they are within the States OR/ID (Region 10), UT/CO (Region 8), or NM (Region 6). EJSCREEN's percentiles and reports processed those block groups as if they were in the Region that corresponds to their State. Regional and Tribal Lands maps should be consulted when viewing those locations and reported Regional percentiles should not be used in those locations.

apparent mainly in two situations: when comparing places across states, or when comparing results to some pre-determined, specific reference percentile (e.g., 80th percentile).

The advantage of US percentiles for an EJ Index, for example, is that a higher percentile in place A versus place B clearly indicates that the combination of the environmental indicator and demographic index is greater in place A than place B. In a sense, the US percentile indicates how uncommon it is to have such a high level for an indicator or index.

State or regional percentiles cannot be compared across states or regions as easily. If two places A and B, in two different states, happen to both be at the 80th percentile for the traffic EJ Index, for example, it is not clear which actually has the higher index value. It just means that A's index is just as uncommon within that state as B's is in B's state. However, this may be useful information because an EJSCREEN user may want to know how high the indicator is relative to the rest of that state.

The state and US percentiles will be very similar if the state and US average indicator values are very similar. However, if the state average is very low compared to the US, the state percentile shown will be higher than US percentile shown, for a given raw value of an indicator. If the state average is much higher than the US average, for an indicator like the traffic indicator, then a traffic score that would normally be considered fairly high nationwide, such as the 90th percentile in the US, would not be considered very unusual within that state, so the state percentile would be lower, and might be only 78th percentile, for example. The state percentile being lower than the US percentile does not mean the indicator value is lower in the given place, it just means the state average is higher than the US average.

Appendix B provides details on how percentiles are calculated, and rounded when displayed.

Buffer Reports

What the Buffer Report Calculates

EJSCREEN allows a user to define a buffer, such as the circle that includes everything within 1 mile of a specific point. Non-circular, user-defined shapes also can be defined to represent buffers of any shape.

The summary within a buffer represents the average resident within the buffer. A report summarizes the demographics of residents within this buffer, as well as the environmental indicators and EJ index values within the buffer. It also provides an estimate of the total population residing in the buffer.

Note that this means one cannot compare two buffers of very different population counts without understanding what each set of results represents. Each represents the average person in that buffer. It does not represent the absolute total amount the buffer contributes to overall disparity in indicator scores nationwide or statewide. Even if the two sets of scores are identical other than in population counts, the buffer with a larger population will contribute more to any national or overall disparity in indicator scores. In general, however, this situation does not tend to arise because most buffers that a user creates in practice will be at least roughly similar in size and population. Even if they are not, a user simply needs to acknowledge that some buffers have larger populations, in which case those percentile results represent more people.

Appendix B provides the details of buffer calculations.

Choosing a Buffer Size versus Rationale for Distance in Proximity Indicators

An EJSCREEN user's choice of distance for a circular buffer is important, and the considerations need to be understood.

EJSCREEN is not able to report on buffers that are too large (e.g., ten or more miles in radius) due to computational limits. It also is not able to report on buffers that are too small (i.e., they do not intersect the internal points of any Census blocks).

In addition to those limits, as a rule of thumb, it is important to know that a buffer that is as small as the block group it centers on will result in estimates with substantially higher uncertainty than one which covers several block groups. A buffer covering five or more block groups, for example, will provide much more confidence in demographic estimates (because of sampling uncertainty as well as the challenge of estimating where residents are located within block groups intersected by the edges of the buffer). Examining patterns the size of a block group or smaller requires using maps of block groups and Census blocks, rather than attempting to draw buffers in EJSCREEN.

Uncertainties are also discussed in section 1 (as general caveats), and Appendix B (in discussions of buffering details and demographic data).

It is important not to confuse two different distances:

1) the distance a user selects for a circular buffer radius (e.g., by default 1 mile, which is 1.6 km), and
2) the distance EJSCREEN used in proximity score calculations (i.e., 5 km for facilities and sites, or 500 meters for traffic).

These two are very different, as explained here, because proximity scores use a pre-determined cut-off distance (e.g., 5 km) and then inverse distance weighting, while for circular buffers a user may wish to specify can vary in distance (e.g., 1 mile or 1.6 km), but be at least as large as one or more local block groups) because a buffer report does not use distance weighting.

The buffer analysis provides a summary of the average resident inside the buffer. It gives equal weight to each resident, regardless of whether they are closer or further from the center of the buffer. There is no distance weighting in a buffer report.

Proximity scores were created very differently than buffer reports are calculated. The proximity scores for each block group were calculated for each residential location using distance weighting to give more weight to closer facilities, sites, or traffic. Because the proximity score uses distance weighting to focus less on the more distant points, it was designed to look at a large area using a large radius, or distance (5 km for facilities or NPL sites). By distance weighting, the proximity score can examine this large area and still provide a useful summary of all the facilities or sites in that wide area.

The proximity score for traffic, for example, looks within a search radius of 500 meters (or further if none is found in that radius). This distance, or scope, was selected to be large enough to capture the great majority of road segments (with traffic data) that could have a significant impact on the local residents, balanced against the need to limit the scope due to computational constraints. Within this relatively wide zone, the closest traffic is given more weight, and the distant traffic given less weight, through inverse distance weighting. The water discharge proximity scores are calculated using the same method as the traffic. The difference being the use of stream segments instead of highways and toxic concentrations instead of traffic counts. The same approach was also used for the facility or NPL site proximity scores. A distance of 5 km was chosen to capture the great majority of facilities or sites that could have a significant impact on local residents. The fact that impacts may be very small for distances of 4 or 5 km is handled by the use of inverse distance weighting in the proximity score formula.

By contrast, a buffer report, again, averages together all residents in the buffer, treating them all equally regardless of their distance from the buffer center. Therefore, many EJSCREEN users may wish to define a modest buffer distance that focuses on those residents who may be “most affected” and "similarly affected" by a single facility or site of interest at the center of the circular buffer.

A very large buffer (e.g., over three miles in radius) could provide misleading results if the goal is to describe the "affected" population, because people in a large buffer could be extremely varied in the extent of their exposure to some source at the center of such a large buffer. If impacts decline with distance, a large buffer would mix many relatively unimpacted, distant residents, with fewer residents who are closer and impacted more, giving a diluted result that fails to describe those most impacted.

At the other extreme, a very small buffer (e.g., the size of a local block group or smaller) is problematic because it could fail to include some significantly affected residents, and also because estimates are more uncertain for smaller geographic areas due to sampling error in the ACS and spatial error in estimating which residents are inside the buffer. Some EJSCREEN users may wish to define a large buffer distance when they know a local facility or site covers a wide area (for example some NPL sites can be very large). Some users may wish to run and compare separate buffer reports for two or more choices of distance, or define hand-drawn buffers to look at zones at various distances from some point.

3 DETAILS ON THE ENVIRONMENTAL INDICATORS IN EJSCREEN

Environmental Indicators Not Included

As described above, EJSCREEN contains 11 environmental indicators, which were selected after a review of available data, other EJ tools and analyses, and the data selection criteria discussed above.

A number of possible factors were identified in the review, which were not ultimately included in EJSCREEN due to various limitations. These limitations were almost always a lack of high resolution data (e.g., only available at county level), and/or lack of geographic coverage (e.g., only available in selected or sampled locations). In some cases an indicator was not added because of a high degree of overlap and double-counting with existing indicators, or resource constraints and practical considerations. One or more of these indicators may be included in future versions of EJSCREEN as more data become available.

Other EPA resources also have more information on many of these issues, such as

- EPA's website (www.epa.gov)
- Envirofacts (https://enviro.epa.gov/)
- A county-level US Environmental Quality Index (EQI)[18]
- EnviroAtlas (https://www.epa.gov/enviroatlas)

EJSCREEN also provides the ability to view some of these issues directly within the EJSCREEN maps, such as impaired water bodies, criteria air pollutant nonattainment areas, TRI facilities, and others. Furthermore, users can import and view, within EJSCREEN, other maps available on the internet, for more of these environmental issues.

Users of EJSCREEN are also encouraged to consider these issues, where appropriate, to the extent they have relevant local information.

Factors not currently used in EJSCREEN include:

- Health data (e.g., overall mortality rate) – Note this is not an environmental factor, but is sometimes of interest in this context. Relevant data were found to be available only at county level resolution.[19]
- Drinking water (from private wells or public water supplies) and surface water quality (other than through the potential relevance of the NPDES proximity indicator in EJSCREEN)[20]

[18] Messer, Jagai, Rappazzo, & Lobdell (2014)
[19] Useful resources include http://www.countyhealthrankings.org, http://www.americashealthrankings.org/, and http://statehealthcompare.shadac.org/content/2/About
[20] See http://www2.epa.gov/learn-issues/water-resources and http://water.epa.gov/scitech/datait/tools/waters/tools/index.cfm

- Contaminated fish/ seafood (other than through the potential relevance of the NPDES proximity indicator in EJSCREEN)[21]
- Beach closures due to pathogens (other than through the potential relevance of the NPDES proximity indicator in EJSCREEN)[22]
- Radon gas exposure[23] or indoor air pollutants other than radon[24]
- Criteria air pollutants other than $PM_{2.5}$ and ozone (Pb, CO, SO_x and NO_x)[25] (Not included due to resource constraints, and more limited modeling/coverage, frequency of nonattainment, or health impact than for $PM_{2.5}$ and ozone).
- Proximity to other point or area sources not already accounted for by EJSCREEN's proximity indicators, NATA indicators, or other air quality indicators. This can include some TRI reporting facilities that do not emit HAPs to air, for example. TRI facilities are a small but important fraction of all regulated facilities. Those emitting HAPs to air are already considered through the NATA indicators, and many others are included in the RMP indicator.[26]
- Exposures to short episodes of elevated releases of air pollutants during startup, shutdown, malfunction, etc. (data gaps in coverage and resolution)
- Exposures to undocumented emissions caused by leaks
- Exposures related to oil and gas extraction, such as hydraulic fracking[27]
- Mining (e.g., uranium mining, etc.)[28]
- Coal ash ponds
- Combined animal feeding operations (CAFOs)[29]
- Leaking underground storage tanks or other contaminated sites other than National Priorities List (NPL) sites, Risk Management Plan (RMP) facilities, and Treatment, Storage and Disposal Facilities (TSDFs)[30]
- Pesticide exposures from spray drift or other sources, or pesticide exposures from residential and other non-agricultural uses[31]
- Noise pollution and odors not already accounted for in other data[32]
- Occupational exposures

---

[21] See http://www2.epa.gov/learn-issues/water-resources
[22] See http://www2.epa.gov/learn-issues/water-resources
[23] See http://www.epa.gov/iaq/ia-intro.html
[24] See http://www.epa.gov/iaq/ia-intro.html
[25] See https://www.epa.gov/criteria-air-pollutants
[26] For examples of other efforts to consider a wide range of facilities, see the various TRI mapping tools such as (http://www2.epa.gov/toxics-release-inventory-tri-program/tri-data-and-tools), and see EPA's RSEI tool: http://www.epa.gov/opptintr/rsei/index.html
[27] See http://www2.epa.gov/hydraulicfracturing
[28] See http://water.epa.gov/polwaste/npdes/Mining.cfm
[29] See http://water.epa.gov/polwaste/npdes/afo/index.cfm
[30] See http://www.epa.gov/swerust1/overview.htm
[31] See http://www2.epa.gov/safepestcontrol and http://www2.epa.gov/science-and-technology/pesticides-science
[32] See http://www.epa.gov/air/noise.html

- Exposures related to imported or domestic consumer products, foods and beverages, or other sources of exposure where we lack detailed geographic data
- Ecosystem services[33]

EJSCREEN is designed to be a nationally consistent screening tool, with results calculated and displayed at the Census block group level. Data inputs must be from publicly available sources, available and consistent across the entire country, and with sufficient spatial resolution. There must be some plausible means of quantifying an adverse effect or a proxy for an adverse effect on residential populations. These requirements set a high bar for including environmental data. Currently, none of the potential environmental indicators listed above meet those criteria.

Environmental Indicators in EJSCREEN

Each of the 11 environmental indicators included in EJSCREEN can be viewed separately. Each environmental indicator is also combined with demographic indexes to form the EJ indexes outlined above. Table 2 summarizes the environmental indicators in EJSCREEN, and the following sections describe each environmental indicator in more detail. The sections above address criteria for selecting which environmental indicators to include in this version of EJSCREEN. Appendix D provides summary statistics for each indicator, including mean and percentile values.

Table 2. Summary Table of Environmental Indicators and Sources

Key Medium	Indicator	Details	Source	Data Year
Air	NATA air toxics cancer risk	Lifetime cancer risk from inhalation of air toxics	EPA NATA, retrieved 2019 https://www.epa.gov/national-air-toxics-assessment/2014-nata-assessment-results	2014
Air	NATA respiratory hazard index	Air toxics respiratory hazard index (ratio of exposure concentration to health-based reference concentration)	EPA NATA, retrieved 2019 https://www.epa.gov/national-air-toxics-assessment/2014-nata-assessment-results	2014
Air	NATA diesel PM	Diesel particulate matter level in air, µg/m^3	EPA NATA, retrieved 2019 https://www.epa.gov/national-air-toxics-assessment/2014-nata-assessment-results	2014
Air	Particulate matter	$PM_{2.5}$ levels in air, µg/m^3 annual avg. (2016)	EPA, OAR (fusion of model and monitor data). For methods, see EPA Report EPA-454/S-15-001 https://www.epa.gov/green-book/green-book-pm-25-2012-area-information	2016

[33] See http://enviroatlas.epa.gov/enviroatlas/

Air	Ozone	Ozone summer seasonal avg. of daily maximum 8-hour concentration in air in parts per billion (2016)	EPA, OAR (fusion of model and monitor data). For methods, see EPA Report EPA-454/S-15-001 https://www.epa.gov/green-book/green-book-8-hour-ozone-2015-area-information	2016
Air/other	Traffic proximity and volume	Count of vehicles (AADT, avg. annual daily traffic) at major roads within 500 meters, divided by distance in meters (not km)	Calculated from 2017 U.S. DOT traffic data, retrieved 2019 https://www.fhwa.dot.gov/policyinformation/hpms/shapefiles.cfm	2017
Dust/ lead paint	Lead paint indicator	Percent of housing units built pre-1960, as indicator of potential lead paint exposure	Calculated based on Census/ACS data, retrieved 2019 https://www.census.gov/programs-surveys/acs/data/summary-file.html	2013-2017
Waste/ air/ water	Proximity to RMP sites	Count of RMP (potential chemical accident management plan) facilities within 5 km (or nearest one beyond 5 km), each divided by distance in kilometers	Calculated from EPA RMP database, retrieved 06/2019 https://www.epa.gov/rmp/risk-management-plan-rmp-rule-overview	2019
Waste/ air/ water	Proximity to TSDFs	Count of TSDFs (hazardous waste management facilities) within 5 km (or nearest beyond 5 km), each divided by distance in kilometers	Calculated from EPA RCRAInfo database, retrieved 07/2019 https://www.epa.gov/hwpermitting/reference-document-hazardous-waste-treatment-storage-and-disposal-facilities	2019
Waste/ air/ water	Proximity to NPL sites	Count of proposed and listed NPL sites[34] within 5 km (or nearest one beyond 5 km), each divided by distance in kilometers	Calculated from EPA CERCLIS database, retrieved 07/2019 http://cumulis.epa.gov/supercpad/cursites/srchsites.cfm	2019
Water	Wastewater discharge	Toxicity-weighted stream concentrations at stream segments within 500 meters, divided by distance in kilometers (km)	Calculated from RSEI modeled toxicity-weighted stream concentrations, created 05/2019 https://www.epa.gov/rsei	2017
Note: EJSCREEN's EJ Indexes also include demographic information that is obtained from the U.S. Census Bureau's American Community Survey (ACS). The 2019 version of EJSCREEN includes 2013-2017 ACS 5-year summary file data, which is based on 2017 Census boundaries.				

[34] Count of NPL sites excludes deleted sites, sites in U.S. territories, and other sites that could not be included.

NATA Air Toxics and NATA Diesel PM

Air toxics, often referred to as hazardous air pollutants (HAPs), are pollutants that are known or suspected to cause cancer or other serious health effects, such as reproductive effects or birth defects, or adverse environmental effects. EPA regulates 187 chemicals under its HAP program (U.S. EPA, 2009d). Most air toxics originate from transportation and industry, including motor vehicles, industrial facilities and power plants.

Indicator	EPA's National Air Toxics Assessment (NATA) provides the following indicators that are used in EJSCREEN: • Estimated lifetime inhalation cancer risk from the analyzed carcinogens in ambient outdoor air. • Hazard index for respiratory effects. • Diesel particulate matter concentration.
Rationale for Inclusion	A chemical's listing as a HAP is based on evidence of cancer or other adverse health effects or environmental effects associated with exposure to the chemical, as determined by EPA and the initial list in the Clean Air Act Amendments of 1990. EPA's Integrated Risk Information System (IRIS) program documents the health risks associated with these chemicals and serves as a basis for the analysis of health implications (U.S. EPA, 2012c). Air toxics cancer risk and noncancer impacts have been included in other EPA EJ screening tools. HAPs are emitted from a wide variety of sources and disperse around the sources, especially downwind. In some cases, these substances react with other constituents in the atmosphere or break down to other chemicals, and most are eventually removed through precipitation or other atmospheric processes. People are exposed in their daily activities in and around their homes, at school or work, and while moving about the area. They inhale the substances, exhale or excrete some portion of them, and have the potential for incurring adverse effects from the portion that stays in the body.
More Information	More information is available at the air toxics website (https://www.epa.gov/haps), the NATA website (https://www.epa.gov/national-air-toxics-assessment), and the IRIS website (www.epa.gov/iris).
Relevant Studies	A comprehensive list of EJ studies using the NATA database can be found in Chakraborty et al., (2011). Some examples of EJ studies of chemicals listed as

HAPs include Morello-Frosch & Jesdale (2006), and other studies reviewed by Liu (2001) and Brender et al., (2011). Diesel particulate matter has also been the subject of EJ analysis (Rosenbaum, Hartley, & Holder, 2011).

Data Source

EJSCREEN uses the most recent data from EPA's National-Scale Air Toxics Assessment (NATA). NATA estimates cancer risk or noncancer implications of many of the 187 air pollutants classified as HAPs, as well as diesel particulate matter (DPM). NATA uses emissions estimates from the National Emissions Inventory (NEI), which is updated every three years. The NEI includes all of the Toxics Release Inventory (TRI) reporting facilities that release hazardous air pollutants, along with many other sources of air pollutants, such as motor vehicles.

Note that the publicly-available NATA, $PM_{2.5}$, and ozone estimates are at tract resolution, and tract level is the resolution used for EJSCREEN, unlike with proximity indicators, for example. Each block group was assigned the NATA or PM or ozone score of the tract containing it. All indicators or statistics then were calculated using block group data, whether or not those block group scores had been assigned based on tracts.

Data Version

The 2019 version of EJSCREEN uses 2014 NATA data, which is based on NEI emissions estimates for 2014 (U.S. EPA, 2015b). This version of NATA estimated ambient concentrations of 180 HAPs plus DPM, and then estimated health implications for 138 of these HAPs.

Data from recent years may no longer be as representative of current conditions as they were at the time the data was collected. The NATA-based indicators in particular should be viewed with this in mind, because emissions of air toxics generally have decreased in the intervening years. This version of EJSCREEN incorporated the most recent data that were available at the time of indicator development. Every attempt will be made to use the most recent appropriate data available in future updates of EJSCREEN. There is always a delay between the release of raw data and their eventual incorporation into any models, tools, or maps. It is also useful to note that although the raw numbers for some indicators do not represent current conditions, the percentiles are much more likely to be reasonably representative of today's conditions in most locations. This is because even if emissions have been significantly reduced overall, for example, the differences between various locations are unlikely to have changed as dramatically, especially when the reductions have come from national regulations and other trends affecting entire industries or sectors in many locations. For this reason, the percentiles may be more representative of current conditions than the raw values of the indicators.

Discussion

EPA's NATA website has extensive documentation of all of the data and methods used in developing the NATA indicators, as well as discussions of uncertainty, caveats, and limitations in the NATA estimates. That information is not repeated here, but it is important that anyone using NATA data understand these issues, so anyone using EJSCREEN should consult the NATA documentation (https://www.epa.gov/national-air-toxics-assessment).

Very briefly, the air pollutants in NATA include likely or known carcinogens such as formaldehyde, benzene, chloroprene, and coke oven emissions, as well as important sources of noncancer impact such as acrolein, DPM and chlorine. The cancer risk in NATA is aggregated as a cumulative risk for the combination of all analyzed HAPs, and this total is used in EJSCREEN.

NATA calculates a hazard quotient, which is the ratio of ambient air concentration to a chemical's health-based RfC. No adverse health effects are expected from exposure if the hazard quotient is less than one. A hazard index is the sum of hazard quotients for chemicals that cause adverse effects through the same toxic mechanism. NATA includes a hazard index for respiratory effects, which is included as an environmental indicator in EJSCREEN. It represents the cumulative impacts of all the relevant air toxics for which respiratory effects were the key health effect.

The NATA website provides more detailed data than EJSCREEN – Tables and maps on individual HAPs or specific types of sources (e.g., mobile sources only) can be generated by GIS practitioners using data from the NATA website, for those requiring more detail than is provided by the data in EJSCREEN.

The reports in EJSCREEN present the environmental indicators from NATA using ranges of percentiles such as 90-95 or 95-100 rather than as the numbers 1-100 (for Regional and US percentiles). This is done in recognition of the uncertainties inherent in comparing NATA estimates across States that may have different approaches in emissions inventories.

Particulate Matter ($PM_{2.5}$)

$PM_{2.5}$ is particulate matter that is 2.5 microns or less in diameter. Common sources of $PM_{2.5}$ emissions include power plants and industrial facilities. Secondary $PM_{2.5}$ can form from gases, such as oxides of nitrogen (NO_x) or sulfur dioxide (SO_2), reacting in the atmosphere. EPA set the first $PM_{2.5}$ National Ambient Air Quality Standards (NAAQS) in 1997, and revised the standards in 2006 and 2012.

Indicator	Annual average $PM_{2.5}$ concentration in micrograms per cubic meter ($\mu g/m^3$).
Rationale for Inclusion	EPA's work associated with the PM NAAQS has documented the health effects associated with exposure to $PM_{2.5}$, including elevated risk of premature mortality from cardiovascular diseases or lung cancer, and increased health problems such as asthma attacks (U.S. EPA, 2009b). $PM_{2.5}$ concentrations at different levels are found in all parts of the United States, so residents are exposed via inhalation to varying degrees. A 2012 EPA report found that the majority of the U.S. population lived in areas in nonattainment of one or more of the NAAQS in effect at that time—a total population of 159 million people (based on 2010 Census data for those locations) (U.S. EPA, 2012g). Several studies relevant to EJ and PM2.5, including those discussing susceptible subgroups, are referenced by Bell & Ebisu (2012). $PM_{2.5}$ has been included in other EPA EJ screening tools.
More Information	More information is available at the PM2.5 website (https://www.epa.gov/pm-pollution).
Relevant Studies	Some examples of studies focused on disparities in exposure to $PM_{2.5}$ have been reviewed in Liu (2001), and more recent studies include Bell & Ebisu (2012); Fann et al. (2011); Post, Belova, & Huang (2011); Miranda, Edwards, Keating, & Paul (2011); Brochu et al. (2011); and Levy, Wilson, & Zwack (2007). A very recent study found disparities in exposure to NO_X, a precursor to $PM_{2.5}$. (Clark, Millet, & Marshall, 2014).
Data Source	EJSCREEN's $PM_{2.5}$ data are estimated from a combination of monitoring data and air quality modeling. Ambient $PM_{2.5}$ concentration is estimated by EPA's Office of Research and Development using a Bayesian space–time downscaling fusion model approach. This approach is described in a series

of three published journal articles (Berrocal, Gelfand, & Holland, 2010a, 2010b, 2011).[35]

$PM_{2.5}$ and ozone estimates were not available for Alaska or Hawaii for use in the 2019 version of EJSCREEN, due to a lack of CMAQ modeling. EPA may be able to include estimates in a future version of EJSCREEN.

Data Version

The 2019 version of EJSCREEN uses $PM_{2.5}$ data that are based on 2016 monitoring and modeling estimates (U.S. EPA, 2015a).

Data from several years ago may no longer be as representative of current conditions as they were at the time the data was collected. The $PM_{2.5}$ and ozone indicators in particular should be viewed with this in mind, because emissions related to $PM_{2.5}$ and ozone generally have decreased in the intervening years. This version of EJSCREEN incorporated the most recent data that were available at the time of indicator development. Every attempt will be made to use the most recent appropriate data available in future updates of EJSCREEN. There is always a delay between the release of raw data and their eventual incorporation into any models, tools, or maps. It is also useful to note that although the raw numbers for some indicators do not represent current conditions, the percentiles are much more likely to be reasonably representative of today's conditions in most locations. This is because even if emissions have been significantly reduced overall, for example, the differences between various locations are unlikely to have changed as dramatically, especially when the reductions have come from national regulations and other trends affecting entire industries or sectors in many locations. For this reason, the percentiles may be more representative of current conditions than the raw values of the indicators. Finally, some supplementary maps and local information can complement the EJSCREEN indicators to provide more recent information. In particular, EJSCREEN also provides updated maps of $PM_{2.5}$ and ozone nonattainment areas (areas not meeting national ambient air quality standards).

Resolution

High-resolution estimates of $PM_{2.5}$ are very difficult to develop for the entire United States. Block groups vary widely in geographic area—some are larger than 100 square kilometers, but a substantial fraction are smaller than 1 sq. km in area. This makes it challenging to develop relevant spatial data.

[35] Detailed documentation of the data fusion approach is provided in EPA Report EPA-454/S-15-001 (https://www3.epa.gov/ttn/scram/11thmodconf/Draft_Guidance_SingleSource_SecondarilyFormed-07152015.pdf via https://www3.epa.gov/ttn/scram/11thmodconf.htm).

Some small areas have used the high-resolution AERMOD model to estimate $PM_{2.5}$ levels for EJ analysis (Maroko, 2012), but such modeling is not feasible for the entire United States at this time.

In the past (prior to approximately 2007), CMAQ used a grid size of 36x36 km in the Western United States and 12x12 km in the Eastern United States, or 36 km in general, as in a recent EJ analysis of the heavy-duty diesel emissions rule finalized in 2001 (Post et al., 2011). After approximately 2007, CMAQ modeling has divided the nation into a grid of cells that are each roughly 12 km by 12 km, and has estimated the $PM_{2.5}$ concentration in each cell.

In a 2010 study, satellite data have been used to estimate $PM_{2.5}$ levels with a spatial resolution of roughly 10 km by 10 km, showing reasonably good agreement with monitoring data (van Donkelaar et al., 2010).

Land use regression (LUR) has also been proposed, and can provide better resolution. Nationwide LUR-based estimates have been developed for NO_x but not $PM_{2.5}$.

The downscaler method was selected for EJSCREEN partly because it is particularly useful for this application, in that it estimates concentration at a specified point, rather than for the average of a large grid cell. The downscaler algorithms combine information from nearby monitors and CMAQ grid cell estimates. This provides an estimate based on more information than models alone or monitors alone could provide. It is important to note that the downscaler and indicators here are not attempting to describe *all* of the local variations in ambient air concentrations. They are merely capturing *some additional variation* that is not seen when relying on models or monitors alone.

Discussion

The downscaling fusion model uses both air quality monitoring data from NAMS/SLAMS (data collected by EPA, state, local and tribal air pollution control agencies at more than 600 hundred monitors nationwide) and numerical output from the Models-3/CMAQ model. The CMAQ model is used extensively by EPA and has been described in detail elsewhere (Byun & Schere, 2006).

This downscaling approach is designed to provide daily, predictive $PM_{2.5}$ (daily average) and O_3 (daily 8-hour maximum) surfaces for a given year, such as 2011, at specified points.

For EJSCREEN, the downscaling method was applied to a grid, and the centroid of each Census tract was determined. EPA's Office of Air and

Radiation generated an estimate for each tract, and then assigned the same tract value to every block group within the given tract. Daily estimates from the downscaling method were averaged for the whole year in the case of $PM_{2.5}$ and for the ozone season (May–September) in the case of ozone.

Again, it is important to note that the downscaler and indicators here are not attempting to describe *all* of the local variations in ambient air concentrations. They are merely capturing *some additional variation* that is not seen when relying on models or monitors alone.

Several data sources have been used elsewhere and were considered for inclusion in EJSCREEN. For instance, EPA's regulatory impact analyses (RIAs) for recent rules and the $PM_{2.5}$ NAAQS have used estimates of $PM_{2.5}$ that combine modeling and monitoring in a different way (U.S. EPA, 2009b). These estimates also start with CMAQ air quality modeling results, but then locally adjust those estimates up or down based on local monitoring data using MATS (monitor attainment test software), which provides an enhanced Voronoi neighbor averaging interpolation technique. Published analyses of $PM_{2.5}$ health impacts have used similarly fused estimates (Fann et al., 2012). Before the downscaler method was developed, a different Bayesian modeling approach was also used, as described in McMillan, Holland, Morara, & Fang (2010). Other efforts have used interpolation between monitors without air quality modeling, such as through basic Voronoi neighbor averaging (Fann & Risley, 2011), or simply the average of monitors within a county (Bravo, Fuentes, Zhang, Burr, & Bell, 2012). Monitors provide reliable estimates where they are located, but suitable $PM_{2.5}$ data are available at fewer than 900 monitors in the United States. While urban areas tend to have $PM_{2.5}$ monitors, more than two-thirds of U.S. counties lack any monitoring data, so modeling is an important complement to monitoring. Methods based on CMAQ alone, monitors alone, CMAQ-MATS and the downscaling approach all provide somewhat different estimates.

Note that the EJSCREEN value does not directly indicate nonattainment of the NAAQS standard because the indicator in EJSCREEN is based on estimates from a combination of modeling and monitoring for a single year, while nonattainment is determined for a large area (often a county) based on three years of monitoring data.

Ozone

Ozone (O_3) is not usually emitted directly into the air, but is created at ground level by a chemical reaction between oxides of nitrogen (NO_x) and volatile organic compounds (VOCs) in the presence of sunlight. These ozone precursors are emitted by motor vehicles, industrial facilities and power plants as well as natural sources. Ground-level ozone is the primary constituent of smog.

Indicator

The May–September (summer/ ozone season) average of daily-maximum 8-hour-average ozone concentrations, in parts per billion (ppb).

Rationale for Inclusion

Toxicological and epidemiological studies have established an association between exposure to ambient ozone and a variety of health outcomes, including reduction in lung function, increased inflammation and increased hospital admissions and mortality (U.S. EPA, 2006b). In the 2006 Air Quality Criteria Document for Ozone, a comprehensive review of the clinical and epidemiological evidence was inconclusive about a possible threshold for ozone-induced health effects. EPA concluded that if a population threshold level exists, it is near the lower limit of ambient ozone concentrations in the United States (U.S. EPA, 2006b). Several subpopulations may experience susceptibility to ozone-induced health effects. These subpopulations include older adults, children, individuals with preexisting pulmonary disease and those with higher exposure levels such as outdoor workers (U.S. EPA, 2006b). A recent review of studies identifying subgroups susceptible to ozone found the strongest evidence for greater sensitivity among the elderly and also the unemployed (Bell, Zanobetti, & Dominici, 2014).

A 2012 EPA report found that the majority of the U.S. population lived in areas in nonattainment of one or more of the NAAQS—a total population of 159 million people (based on 2010 Census data for those locations) (U.S. EPA, 2012g). As standards are updated, nonattainment areas are redefined along with the number of people living in redefined nonattainment areas. Ozone concentrations at different levels are found in all parts of the United States, so residents are exposed via inhalation to varying degrees.

Ozone has been included in other EPA EJ screening tools.

More Information

More information is available at the ground-level ozone website (https://www.epa.gov/ground-level-ozone-pollution).

Relevant Studies

Some examples of studies that have focused on disparities in exposure to ozone include Fann et al. (2011), Grineski (2007), Grineski et al. (2007), and those reviewed in Liu (2001).

Data Source

EJSCREEN's ozone data are estimated by EPA from a combination of monitoring data and CMAQ air quality modeling. Ozone was estimated with the same approach as $PM_{2.5}$, and the methodology is described above. Ozone faces similar limitations, in that a limited number of U.S. monitors have suitable data, so modeling is an important complement to monitoring data.

$PM_{2.5}$ and ozone estimates were not available for Alaska or Hawaii for use in the 2019 version of EJSCREEN, due to a lack of CMAQ modeling. EPA may be able to include estimates in a future version of EJSCREEN.

Data Version

The 2019 version of EJSCREEN uses ozone data that are based on 2016 monitoring and modeling estimates (U.S. EPA, 2015a). Tract estimates were assigned to block groups as described for the $PM_{2.5}$ indicator.

Resolution

See "Resolution" section under the previous environmental indicator, $PM_{2.5}$.

Discussion

See "Discussion" section under the previous environmental indicator, $PM_{2.5}$.

Lead Paint Indicator

A key source of exposure to lead for many Americans is through lead paint and lead-containing dust that accumulates indoors, in homes or in other buildings where lead paint was used. Exterior structures painted with lead-based paint are also a source of ambient lead through chipping exterior paint. Elevated short-term lead dust loadings have also been observed following demolition of old buildings (U.S. EPA, 2011c). Lead-based paint was banned in the United States by the Consumer Product Safety Commission in 1978, but lead-based paint used in housing before the ban remains a significant source of exposure to lead for children and adults. Lead paint and contaminated dust and soil are considered the leading cause of high lead levels in U.S. children (Levin et al., 2008).

Indicator

The percentage of occupied housing units built before 1960 was selected as an indicator of the likelihood of having significant lead-based paint hazards in the home.

Rationale for Inclusion

Elevated blood lead levels are a well-documented public health concern of particular interest to EJ stakeholders, and represent an important environmental health issue (U.S. EPA, 2006a, 2011c).

Certain demographic groups may be more susceptible to lead exposure. For example, blood lead's association with cardiovascular outcomes appears to be stronger among Mexican Americans and non-Hispanic blacks than non-Hispanic whites (U.S. EPA 2011c). Also, some but not all studies suggest lead has a greater impact on IQ among those of low socioeconomic status (U.S. EPA 2011c).

Despite significant reductions in ambient levels of lead from the phase-out of leaded gasoline and the 1978 ban on lead-based paint, lead remains a significant hazard for children. Recent research has demonstrated that children can experience neurological damage even at low levels of exposure to lead. In May 2012, the Centers for Disease Control and Prevention (CDC) agreed to adopt the recommendations of the CDC Advisory Committee for Childhood Lead Poisoning Prevention (ACCLPP) for defining elevated blood-lead levels (BLLs) among children. The ACCLPP recommended that CDC use a childhood BLL reference value based on the 97.5th percentile of the population BLL in children under age 6 to identify children and environments associated with lead-exposure hazards (Centers for Disease Control and Prevention, 2012). The 97.5th percentile value is currently 5 µg/dL.

Surfaces originally covered with lead-based paint may chip, flake or develop a chalky surface. The lead in these pieces or particles may be moved about the interior or exterior of the painted structures, and be moved from inside to outside and vice versa. Through direct contact with the painted surfaces or through contact with the released particles, lead may adhere to hands and other parts of the residents' bodies, and people may ingest some portion of the lead. If the painted surfaces are disturbed through renovation or other actions, some lead-based paint particles may be temporarily suspended in the air, and particles on surfaces within the structures may be re-suspended during the residents' activities. The suspended particles may be inhaled or may fall on food and be ingested. Children playing inside or outside and exposed to particles of lead-based paint may ingest some of the lead through hand–mouth actions.

An analysis of data collected during the 1999–2004 National Health and Nutrition Examination Survey (NHANES) showed that children living in older housing stock (built before 1950) were significantly more likely to have blood lead levels greater than 5 µg/dL than children living in housing built after 1978 (Jones et al., 2009). Jones et al. estimated that 7.4% of children under age 6 had blood lead levels greater than 5 µg/dL during NHANES 1999–2004. For children under age 6 living in the highest risk housing (built before 1950), Jones et al. observed that 15.1% had blood lead levels above 5 µg/dL. For children under age 6 living in the lowest risk housing (built in 1978 or later), 2.1% had BLLs above 5 µg/dL.

EPA EJ screening tools in the past generally have not included proxies for lead exposure.

More Information

More information is available at EPA's lead website (http://www.epa.gov/lead).

Relevant Studies

Several examples of EJ studies of exposure from lead paint exist, including Gaitens et al., 2009 and others.

Data Source

The data were collected at the block group level from the ACS estimates from the Census Bureau. The indicator was calculated by dividing the count of occupied housing units built prior to 1960 by the total number of built housing units in the block group (ACS table B25034, see Appendix B for details).

Data Version

The block-group level data for the 2019 version of EJSCREEN were collected from the 2013-2017 ACS (U.S. Census Bureau). Approximately 40% of occupied, non-institutional housing units in the United States were built prior to 1960, as of 1999.

Discussion

Lead paint was used extensively in the United States prior to the 1978 ban on lead in new residential paint, and a home built prior to 1960 is far more likely to have lead hazards than one built more recently (Gaitens et al., 2009; Jacobs et al., 2002). In 2002, Jacobs et al. reported that approximately 40 million homes in the United States still had lead-based paint hazards, based on a nationally representative survey conducted in 1998–2000. The likelihood of such hazards was found to have changed dramatically for housing built in 1960–1977 compared to pre-1960 housing (Table 3).

Based on Jacobs et al. (2002), EPA calculated the following likelihoods of significant lead-based paint hazards:

- Pre-1960 vs. all others: 54% vs. 6% (9 times as likely).
- Pre-1960 vs. all others, among those with children under age 6: 68% vs. 4% (16 times as likely).

Data and analysis published in 2009 confirmed prior conclusions that potential exposure to lead is associated with housing age, providing more information on lead concentrations in household dust as a function of housing age (Gaitens et al., 2009). Some of the models presented by Gaitens et al. (2009) suggest that the largest decreases in lead dust levels are seen between housing built prior to 1940 and after 1940, with more modest contrasts seen for housing built after 1960 and after 1977. A cutoff of 1960, however, is consistent with the data from Jacobs et al. (2002), and the window sill lead dust models from Gaitens et al. (2009).

It is important to note that older housing alone may not represent any actual risk or even exposure.

Table 3. Likelihood of Lead-Based Paint Hazards by Housing Construction Date

Year Built	Share of Housing with Significant Lead-Based Paint Hazards (and 95% Confidence Interval)
Post-1960:	
1978–1998	3% (1–6%)
1960–1977	8% (6–12%)
Pre-1960:	
1940–1959	43% (32–51%)
Before 1940	68% (56–75%)

Source: Jacobs et al. (2002). A "significant lead-based paint hazard" is defined as "a lead-based paint hazard above de minimis levels as defined in U.S. EPA and U.S. Department of Housing and Urban Development (HUD) regulations."[36]

[36] The de minimis levels for paint deterioration are ≤ 20 ft^2 (exterior) or ≤ 2 ft^2 (interior) of lead-based paint on large surface area components (walls, doors), or damage to ≤ 10% of the total surface area of interior small surface area component types (windowsills, baseboards, trim) (40 C.F.R. § 745.65).

Traffic Proximity

A substantial fraction of the U.S. population lives in close proximity to traffic, and the number of vehicle-miles traveled has grown 40% from 1990 to 2010 (U.S. EPA, 2012d). Proximity to motor vehicle traffic is associated with increased exposures to ambient noise, toxic gases and particulate matter including diesel particulates. Technical details about the methodology used to determine traffic proximity are provided in Appendix C.

Indicator

The count of vehicles per day within 500 meters of a block centroid, divided by distance in meters, presented as the population-weighted average of blocks in each block group. Adjustments are made so that the minimum distance used is reasonable when very small.

Rationale for Inclusion

A 2011 literature review identified several studies that "found that living near hazardous waste sites, industrial sites, cropland with pesticide applications, highly trafficked roads, nuclear plants, and gas stations or repair shops is related to an increased risk of adverse health outcomes" (Brender et al., 2011, p. S37). This indicator is the first of five that relate to this category of concern.

Is should first be noted that there are both positive and negative aspects to living near major roads. Proximity to roads can provide access to jobs, health care, food, recreational opportunities, and other benefits. The indicator of traffic proximity and volume is designed to screen for the negative aspects, so it uses distance weighting and volume weighting to focus on the extremes of very close proximity to very high volumes of traffic, such as living closer than 50-100 meters from a multi-lane highway, as explained below.

Residential proximity to traffic has been associated with various health impacts, particularly asthma exacerbation and possibly onset of asthma, as well as mortality rates (Baumann et al., 2011; Health Effects Institute, 2010). Proximity to traffic has also been associated with subclinical atherosclerosis (a key pathology underlying cardiovascular disease (CVD)), prevalence of CVD and coronary heart disease (CHD), incidence of myocardial infarction, and CVD mortality (Hoffman et al., 2009).

Vehicle-related emissions of various pollutants—ultrafine and other components of $PM_{2.5}$, lead and other metals, and mobile source air toxics such as benzene, nitrogen oxides (NO_x), hydrocarbons and carbon monoxide (CO)—are believed to contribute to these health effects. Vehicles also emit precursors that add to ambient ozone and $PM_{2.5}$. Additionally, EPA's 2005 NATA estimated that mobile emissions accounted for about 30% of average cancer risk from the pollutants in NATA, mainly from benzene (U.S. EPA, 2009c). However, the spatial accuracy of NATA's mobile source impacts is limited, because local estimates are based on countywide total mobile source emissions roughly allocated to each part of the county based on presence of major roads. The traffic indicator in EJSCREEN provides a more detailed analysis of the volume and location of traffic than was used in NATA. Also, NATA captures only some of the impacts associated with traffic, so the traffic indicator is a useful complement.

Traffic proximity is also associated with noise, which is a risk factor for various health problems. Workplace and transportation-related noise have been associated with release of stress hormones; sleep disturbance; hypertension; altered heart rate; ischemic heart disease; myocardial infarction; and, among the elderly, risk of stroke (Sørensen et al., 2011). In one study, for example, among those older than 64.5 years of age, the stroke incidence rate ratio was 1.27 per 10 dB more road traffic noise (Sørensen et al., 2011).

Whether noise or other factors account for it, local traffic volume is a predictor of stress (which itself is associated with significant health risks). In 2010, Yang & Matthews concluded that, "[a]t the neighborhood level, the presence of hazardous waste sites and traffic volume were determinants of self-rated stress even after controlling for other individual characteristics" (2010, p. 803).

Any indicator of residential proximity addresses exposures relevant to the residences within a block group, and would not capture most exposures that occur away from the home, such as at work, at school or during a commute.

In the past, EJ screening tools at EPA have not included traffic proximity.

More Information

More information is available at the mobile source pollution website (https://www.epa.gov/mobile-source-pollution).

Relevant Studies

Some examples of studies of disparities in proximity to traffic include Tian et al., 2013; Rowangould, 2013; Brender et al., 2011; Chakraborty (2006); and Liu, 2001.

Data Source

Measures of traffic proximity in EJSCREEN are based on average annual daily traffic (AADT) estimates in the Highway Performance Monitoring System (HPMS) dataset in the Department of Transportation (DOT) National Transportation Atlas Database (NTAD). The HPMS highway data is maintained by states and compiled by DOT.

The HPMS data are collected at the state level, and the traffic counting program is designed to cover all interstate, principal arterial, other National Highway System and HPMS sample sections on a 3-year maximum cycle where at least one-third of roads are counted each year. More details on the HPMS are available at http://www.fhwa.dot.gov/policyinformation/hpms.cfm.

Data Version

The 2019 version of EJSCREEN uses 2017 HPMS data (U.S. Department of Transportation, 2019)[37]. The full HPMS dataset contains 6,809,077 records. Of these we selected 3,784,158 road segments that represent "major road segments. This was defined as interstates, expressways, principal arterials and minor arterials in urban areas. These have total length of approximately 593,270 miles of the 4,177,073 total miles in HPMS. While this is only 14% of total road miles of public road in the US, the roads not included (e.g., local streets) tend to have much lower levels of traffic, so the roads included appear to account for almost two thirds of all US traffic (vehicle-miles travelled).[38] For the 2019 version of EJSCREEN, a total of 11,078,297 Census 2010 blocks were analyzed to find all road segments within 500 meters of each block's internal point, or the nearest single segment if none were found within 500 meters.

Discussion

The traffic proximity indicator is based on average annual daily traffic (AADT) divided by distance in meters. For example, a single highway with 16,000 AADT at 400 meters distance would result in a score of 16,000/400=40, which is close to the median person's block group traffic proximity indicator value in the US. About 5% of the population has traffic proximity indicator values more than ten times as high as the median, because traffic volumes vary widely across roads and communities. The most traveled highway section in the United States, the I-405 in the Los Angeles-Long Beach-Santa Ana area, had 374,000 vehicles of AADT in 2008.[39] About forty other highway sections in the US exceeded 250,000 AADT, but about half were in just one state – California – and the rest were spread over just a dozen states.

The proximity score is based on the traffic within a search radius of 500 meters (or further if none is found in that radius). It is important to understand that this distance was selected to be large enough to capture the great majority of road segments (with traffic data) that could have a significant impact on the local residents, balanced against the need to limit the scope due to computational constraints. The closest traffic is given more weight, and the distant traffic given less weight, through inverse distance weighting. For example, traffic 500 meter away is given only one tenth as much weight as traffic 50 meters away.

Based on the available evidence, a distance of roughly 100–300 meters, or perhaps up to 500 meters, from traffic should be considered as most important. This

[37] http://www.fhwa.dot.gov/policyinformation/hpms.cfm

[38] The included 340,000 miles is comparable to the road miles covered by all interstate, freeway/expressway, and principal arterials, plus 50% of the minor arterial miles in the US, which together carry 64% of VMT. Collector and local roads are the balance of public roads. See http://www.fhwa.dot.gov/policyinformation/statistics/2014/hm220.cfm and http://www.fhwa.dot.gov/policyinformation/statistics/2014/vm202.cfm

[39] Office of Highway Policy Information, US DOT, 2008 Highway performance monitoring system (HPMS) July 27, 2010.

distance focuses on the types of exposures typically studied and shown to be associated with health impacts in near-roadway epidemiology. Epidemiologic studies of the impacts of proximity to traffic often utilize distances of 50–1,500 meters to define a cutoff between less and more exposed locations (Health Effects Institute, 2010). For example, a major study of coronary heart disease prevalence used distances greater than 200 meters as the reference group and found adjusted odds ratios of 1.08 for residences within 100–200 meters, 1.71 for 50–100 meters and 1.55 within 50 meters of a major road. Only 15% of participants lived within 200 meters of a major road, and only 3% within 50 meters in this study of heart disease (Hoffman et al., 2009).

Additionally, a distance cutoff of 500 meters captures exposures of concern for most definitions of mobile source impact. In a review of numerous prior studies of proximity to roads, in combination with a modeling case study, Zhou & Levy (2007) suggested that a distance of 500 meters should capture exposures of concern, although impacts may be largely limited to just 100 meters from roads for ultrafine particles and $PM_{2.5}$ mass from mobile sources alone.

A critical review of literature on traffic-related air pollution in 2010 "identified an exposure zone within a range of up to 300 to 500 meters from a highway or a major road as the area most highly affected by traffic emissions… and estimated that 30% to 45% of people living in large North American cities live within such zones" (Health Effects Institute, 2010, p. 7-5). A 2009 analysis of $PM_{2.5}$ levels in Southern California found that traffic within 300 meters of a monitor was the most informative predictor of monitored $PM_{2.5}$ levels, out of a wide range of factors considered such as various distances from roads, population density and the presence of industry (Krewski et al., 2009).

On the other hand, some studies have shown a dramatic drop in at least ultrafine levels within the first 100 meters downwind from a freeway, and an even sharper (essentially immediate) drop in the upwind direction (Zhu, Hinds, Kim, & Sioutas, 2002). This pattern has been seen in more recent measurements—levels on California highways (measured using monitors on vehicles) were compared to levels near those roads (roughly 50–300 meters away), and black carbon levels in particular were as much as 10 times higher on the road than near the road, for 1-hour averages (Fujita, Campbell, Zielinska, Arnott, & Chow, 2011). The same study found much higher levels (generally 2–5 times higher) on the road than near the road, for $PM_{2.5}$ mass, CO, NO, NO_x, VOCs, benzene, toluene, ethylene, xylene, formaldehyde and acetaldehyde. This reinforces the idea that exposures very close to a busy highway are most important, and that levels drop rapidly within tens of meters, falling to much lower levels within the first 50–300 meters (Spengler et al., 2011).

Many studies have analyzed roadway proximity categorically, including only major roads, but roads vary in the amount of traffic they carry, so AADT provides a better starting point for considering impacts than simply whether a road is a major road. Several land use regression studies and other research (Hoek et al., 2011; Hystad et al., 2011) have suggested that inverse distance-weighted traffic volume is a reasonably good proxy for ambient air concentrations of NO_x, $PM_{2.5}$ mass ($\mu g/m^3$), black carbon or ultrafine PM (as particle number concentration) nearby (50–500 meters). Levels are clearly higher downwind of the road, and higher where wind speeds are lower, but in the absence of detailed location-specific data, traffic volume and distance are useful indicators (Hoek et al., 2011).

Wastewater Discharge Indicator (Stream Proximity and Toxicity-Weighted Concentration)

The wastewater discharge indicator takes into account pollutant loadings from the Discharge Monitoring Report (DMR) Loading Tool (which include NPDES DMR discharges and TRI releases) for toxic chemicals reported to the Toxics Release Inventory. The data were input into the RSEI model to incorporate chemical toxicity and fate and transport in order to estimate concentrations of pollutants in downstream water bodies (i.e., stream reaches) and derive a toxicity-weighted concentration. The new indicator takes into account proximity from the stream reaches to Census blocks as described in the EJSCREEN indicator methodology section in Appendix C. Such an indicator allows EJSCREEN to provide users with a better approximation of the potential impact of a water discharge in several ways:

- Account for dilution in the discharge stream - a large discharge has a higher impact in a stream with a low flow volume than a stream with a high flow volume.
- Give greater weight to releases of highly toxic chemicals.
- Give greater weight to communities downstream of a discharge than to communities the same distance upstream or in the general area.

Technical details about the methodology used to determine proximity to major and minor direct dischargers to water are provided in Appendix C.

The water environmental indicator in previous versions of EJSCREEN was solely based on the number of major NPDES facilities within a specific radius and took into account the distance the facilities are from the centroid of a census block group. The indicator has no relationship to whether there is a discharge of pollutants, nor what waterbody those are discharged into, or whether there is potential for exposure to the discharged pollutants.

Indicator	The toxicity-weighted concentration in stream reach segments within 500 meters of a block centroid, divided by distance in meters, presented as the population-weighted average of blocks in each block group. Adjustments are made so that the minimum distance used is reasonable when very small.
Rationale for Inclusion	Water pollutants can have human health or adverse ecological effects, depending on concentration in the water, exposure to the water, toxicity of the particular chemical and other factors. There are approximately 6,700 major direct dischargers in the United States. These facilities discharge around 50 billion pounds of pollutants each year directly into the nation’s streams and rivers (including conventional pollutants such as nitrogen and phosphorus) (U.S. EPA, 2012a).

People may be exposed to the discharged pollutants either directly or through indirect pathways. People swimming in the downstream waters or engaging in water-based recreation may be directly exposed dermally, orally or through inhalation of volatized substances. If the released substances reach a downstream drinking water intake, consumers of the finished waters may consume whatever portion of the substances is not removed by the drinking water utility. Some portion of the discharged materials may enter the groundwater of neighboring areas and reach people through drinking water derived from wells that draw upon that aquifer.

Some EPA EJ screening tools have included measures of proximity to all facilities regulated by EPA, which includes major direct water dischargers.

More Information

More information is available at the water website (https://www.epa.gov/environmental-topics/water-topics), the NPDES webpage (https://www.epa.gov/npdes), and the RSEI webpage (https://www.epa.gov/rsei).

Relevant Studies

Some examples of studies of disparities in proximity to water dischargers include Fitos and Chakraborty (2010); Brender et al., 2011; VanDerslice, 2011; and a recent study (Deganian & Thompson, 2012) that tallied the number of facilities of different types within 10 km^2 squares and compared these counts to the demographics of the squares.

Data Source

RSEI, NPDES, DMR, and TRI loadings by chemical; RSEI model results mapped to NHD Version 2.0 stream reach segments. The NHD used by RSEI does not include stream reach segments for Alaska, so the provided RSEI model results include data for the Continental US, Hawaii, and Puerto Rico, but not for Alaska.

Data Version

The 2019 version of EJSCREEN uses locational information retrieved from the RSEI Version 2.3.7 with 2017 releases in May 2019.

Discussion

The indicator was developed using data including loadings from NPDES major and minor dischargers from the DMR Loading Tool, TRI releases, location of the discharges, decay rates for each chemical, and reach location information into which the chemical is discharged. All the necessary hydrography is available in the RSEI modeling environment. After input tables were created, running the RSEI model provided calculated in-stream concentrations for each chemical release, which were then weighted by chemical-specific toxicity weights that were already built into RSEI. The toxicity-weighted stream concentrations were then used as an input to the EJ Indicator.

The Wastewater Discharge Indicator was constructed using the following data and steps, calculating toxicity-weighted receiving waterbody concentrations:

- DMR Loading Tool loadings data, combining both DMR and TRI facility results, for TRI chemicals only as an input.
- Initial discharge reach (if reach is not available, RSEI has a program to pick the nearest reach based on the latitude/longitude of a facility or discharge point).
- Hydrography (stream flow, connectivity, etc.) are available in the RSEI modeling context.
- Individual chemical loadings.
- Water decay coefficient for each chemical (can be estimated by HYDROWIN, an EPA estimation program based on chemical structure). If no decay coefficient is available, RSEI only models dilution.
- RSEI toxicity weights, applied to the calculated in-stream concentrations.

Proximity to NPL Sites

Congress enacted the Comprehensive Environmental Response, Compensation and Liability Act (CERCLA), commonly known as Superfund, in 1980. This law was established to provide broad federal authority to respond to uncontrolled abandoned hazardous waste sites. Under CERCLA, EPA's response can involve remedial (long-term) cleanup actions or short-term removal actions.

EPA places sites on the National Priorities List (NPL) (a key subset of all "Superfund" sites) based on a defined set of criteria and a public comment process. Inclusion of a site on the NPL does not impose a financial obligation on EPA, nor does it assign liability to any party. The NPL serves primarily informational purposes, identifying sites that appear to warrant remedial actions, thereby conveying to policymakers and the public the size and nature of the nation's cleanup challenges.

Sites can be placed on the NPL in one of three ways[40]:

1. The site receives a score of 28.5 or higher in EPA's Hazard Ranking System (HRS);
2. States or territories designate a top-priority site; or
3. A site meets these requirements:
    a. The Agency for Toxic Substances and Disease Registry (ATSDR) of the U.S. Public Health Service has issued a health advisory that recommends removing people from the site;
    b. EPA determines the site poses a significant threat to public health; and
    c. EPA anticipates it will be more cost-effective to use its remedial authority (available only at NPL sites) than to use its emergency removal authority to respond to the site.

Technical details about the methodology used to determine proximity to NPL sites are provided in Appendix C.

Indicator	The count of sites proposed and listed on the National Priorities List (NPL), each represented by a point on the map (latitude/ longitude coordinate), within 5 km of the average resident in a block group, divided by distance, calculated as the population-weighted average of blocks in each block group. Adjustments are made if there are no NPL sites within 5 km, and so that the minimum distance used is reasonable when very small.
Rationale for Inclusion	Soon after the passage of CERCLA and the Superfund Amendments and Reauthorization Act, questions started to be raised about the locations, listing decisions and pace of cleanup at NPL sites in low-income and minority communities (Hird, 1993; Probst, 1990; United Church of Christ, 1987), and such concerns have continued to this day (Anderton, Oakes, & Egan, 1997; Baden, Noonan, & Turaga, 2007; O'Neil, 2007). The study by Deganian & Thompson (2012) included NPL sites in the tally of pollution points in each of the 10 km^2 squares in the study area, for comparison with demographic variables. Earlier studies related the presence of NPL sites to population characteristics for different definitions of the host areas—

40 https://www.epa.gov/superfund

counties (Hird, 1993), Census places or minor civil divisions where places are not defined (Zimmerman, 1993), and Census tracts (Anderton et al., 1997).

The contaminants in NPL sites may reach humans in a number of ways. Volatile contaminants may enter the atmosphere and reach individuals via the inhalation route. Particularly in dry climates or seasons, contaminants on the surface of some sites can become airborne and reach people directly through inhalation or indirectly after being deposited on surfaces that people may contact. Contaminants can also enter the food chain if the wind disperses them onto land used for agriculture. Some contaminants may migrate into groundwater. People may be exposed via drinking water derived from the aquifer, through vapor intrusion into their residences or through other routes.

Some EPA EJ screening tools have included measures of proximity to all facilities or other sites regulated by EPA, which include NPL sites.

More Information

More information is available at the Superfund website (http://www.epa.gov/superfund).

Relevant Studies

Some examples of studies focused on disparities in proximity to NPL sites include Brender et al. (2011) and those reviewed in Liu (2001).

Data Source

A single point location (latitude/ longitude coordinates) for each proposed and listed NPL site was obtained from EPA's CERCLIS database. The database does not provide details on the boundaries of each site, so this point data had to serve as a way to represent site locations. For residents close to very large sites, the available data may not provide an accurate representation of proximity to relevant portions of the site. These points are approximations of the locations of sites, and are not necessarily at the "center" of a given site. In a few cases a site's coordinates were located in a major body of water according to the database, so EPA manually specified new, plausible, nearby coordinates for use in EJSCREEN.

The count excludes deleted sites and sites in U.S. territories. Sites located in Guam are not included in the 2019 version of EJSCREEN.

Data Version

The 2019 version of EJSCREEN uses locational information retrieved from the SEMS database in July 2019. A total of 1,387 proposed and listed NPL sites were included in the EJSCREEN indicator. During processing of the SEMS extract, it was found that 5 sites did not have latitudes/longitudes, these were corrected using FRS-provided coordinates for the 2 sites, 1 coordinate set from the previous year's dataset, and 2 sites were geocoded. Note that the point data used to show site locations in the "Map Supplementary Layers" menu is a different database than the database used to calculate the EJSCREEN NPL proximity indicators and indexes. The Superfund

"Map Supplementary Layer" database includes deleted NPL sites, and NPL sites in U.S. territories, excluded from the EJSCREEN NPL proximity indicator database.

Discussion

Each Census block group in the United States was assigned a proximity score that was the population-weighted sum of block-level proximity scores. Appendix C provides more details on how proximity scores were calculated. First, each Census block was given a proximity score that was the sum of inverse distance-weighted count of sites anywhere within 5 km of the block's internal point. This score can be thought of as the number of NPL sites per kilometer of distance from the average person. It is also equal to the number of sites divided by the harmonic mean of their distances. This means one site 2 km away gives a score of 1/2, while three sites each 4 km away give a score of 3/4, and five sites all at 1 km away give a score of 5.[41] If there is no site within 5 km of a block centroid, 1/d is used, where d is the distance to the single nearest at any distance.

As with all proximity-based indicators, proximity alone may not represent any actual risk or even exposure.

[41] An adjustment was made so that any distance smaller than the block's "radius" was set equal to 90% of that radius, with radius defined as the square root of (area/pi). This adjustment accounted for the fact that the average location (residence) within a circle of radius R is 0.9R away from the average point (site) that is within the circle. For more detail, see Appendix C.

Proximity to TSDFs

The Resource Conservation and Recovery Act (RCRA), an amendment to the Solid Waste Disposal Act, was enacted in 1976 to address the growing volumes of municipal and industrial solid waste generated nationwide. RCRA was further amended in 1984 with the addition of the Hazardous and Solid Waste Amendments. RCRA Subtitle C establishes a federal program to manage hazardous wastes from "cradle to grave," or from generation to disposal, to ensure that hazardous waste is managed in a manner that protects human health and the environment. EPA has developed Subtitle C regulations governing hazardous waste generation, transportation, and the several hundred active treatment, storage or disposal facilities (TSDFs).[42]

Technical details about the methodology used to determine proximity to TSDF facilities are provided in Appendix C.

Indicator	The count of all commercial TSDF facilities within 5 km, divided by distance, presented as population-weighted averages of blocks in each block group. Adjustments are made if there are none within 5 km, and so that the minimum distance used is reasonable when very small.
Rationale for Inclusion	The substances at TSDF facilities may reach humans in a number of ways. Volatile substances may enter the atmosphere and reach residents via the inhalation route. Particularly in dry climates or seasons, substances on the surface of some sites may be entrained in the atmosphere and reach people directly through inhalation or indirectly after being deposited on surfaces that people may contact or on arable land. Some substances may migrate from the site into groundwater. People may be exposed via drinking water derived from the aquifer, through vapor intrusion into their residences or through other routes. Some EPA EJ screening tools have included measures of proximity to all facilities regulated by EPA, which includes TSDFs.
More Information	More information is available at the hazardous waste webpage (https://www.epa.gov/hw), the TSD webpage (http://www.epa.gov/hwpermitting), the waste information page (https://www.epa.gov/hw) and a RCRAInfo page (https://www.epa.gov/enviro/rcrainfo-overview).

[42] Basic information and TSDF counts are available here: https://echo.epa.gov/
More information is available from EPA's RCRA Orientation Manual, available at https://www.epa.gov/hwgenerators/resource-conservation-and-recovery-act-rcra-orientation-manual (U.S. EPA, 2012e).

Relevant Studies

Some examples of studies or reviews that have focused on disparities in proximity to TSDFs include Liu (2001) and Brender et al., (2011). Issues around environmental justice and TSDFs influenced the early origins of EJ work (General Accounting Office, 1983; United Church of Christ, 1987) and have been the topic of ongoing research (Been & Gupta, 1997; Boer, Pastor Jr., Sadd, & Synder, 1997; Mohai & Saha, 2007; Oakes, Anderton, & Anderson, 1996; Pastor Jr., Sadd, & Hipp, 2001; Saha & Mohai, 2005; United Church of Christ, 2007).

The study by Deganian & Thompson (2012) included Hazardous Waste Inventory sites, RCRA hazardous waste storage sites and active solid waste landfills sites in the tally of pollution points in each of the 10 km^2 squares in the study area, for comparison with demographic variables.

Earlier studies related the presence of TSDFs to population characteristics for different definitions of the host areas—Census-designated areas (General Accounting Office, 1983), postal ZIP codes (United Church of Christ, 1987), and Census tracts (Anderton, Anderson, Oakes, & Fraser, 1994).

Data Source

Latitudes/Longitudes for all active commercial TSDF sites were obtained from the RCRAInfo database.

Data Version

The 2019 version of EJSCREEN uses locational information retrieved from the RCRAInfo database in July 2019 (U.S. EPA, 2011e). This was augmented by an extract of RCRA Industrial Generators from 2017. A total of 17,306 TSDF facilities were included in this version of EJSCREEN. Note that the point data used to show facility locations in the “Map Supplementary Layers” menu is updated more often than the database with calculated EJSCREEN indicators and indexes, so in some small number of cases a facility may be in one data source but not the other.

Discussion

Each block group in the United States was assigned a proximity score that was the population-weighted sum of block-level proximity scores. Appendix C provides more details on the calculation of proximity scores. First, each block was given a proximity score that was the sum of inverse distance-weighted count of TSDFs anywhere within 5 km of the block’s centroid. This score can be thought of as the number of facilities per kilometer of distance from the average person. It is also equal to the number of facilities divided by the harmonic mean of their distances. This means one facility exactly 2 km away gives a score of 1/2, while three facilities exactly 4 km away give a score of 3/4, and five facilities all at 1 km away give a score of 5.[43] If there is no facility within 5 km of a block centroid, 1/d is used, where d is the distance to the single nearest at any distance.

As with all proximity-based indicators, proximity alone may not represent any actual risk or even exposure.

[43] An adjustment was made so that any distance smaller than the block's "radius" was set equal to 90% of that radius, with radius defined as the square root of (area/pi). This adjustment accounted for the fact that the average location (residence) within a circle of radius R is 0.9R away from the average point (facility) that is within the circle. For more detail, see Appendix C.

Proximity to RMP Sites

Accidental releases of toxic substances and incidents involving fires and explosions can result from the production, use, or transport of industrial materials. Evacuations, injuries and deaths have resulted in some cases. Concern about the risks of chemical accidents led Congress to pass the Emergency Planning and Community Right-to-Know Act of 1986 (EPCRA), and amendments to the Clean Air Act (CAA) (section 112(r)), which together created reporting and planning obligations for a variety of facility types, the EPA, and state and local planning and response organizations.

The facilities discussed here as "RMP facilities" are those facilities required by the CAA to file risk management plans. The regulations under CAA section 112(r) establishes a List of Regulated Substances—72 substances listed because of their high acute toxicity and 60 because of their flammable or explosive potential—along with threshold quantities (TQs) for each. The listed substances are those that pose the greatest risk of harm from accidental releases. If a facility maintains a quantity of any such chemical above those TQs, it must file an RMP with EPA.

It should be noted that some concerns related to proximity to facilities are already accounted for in NATA indicators for ambient air pollutants (e.g., cancer risk and hazard indexes), but NATA is based on one year of reported annual releases, which would not account for accidental releases unless they occurred that year.

Technical details about the methodology used to determine proximity to RMP facilities are provided in Appendix C.

Indicator	The count of RMP facilities within 5 km, divided by distance, presented as population-weighted averages of blocks in each block group. Adjustments are made if there are none within 5 km, and so that the minimum distance used is reasonable when very small.
Rationale for Inclusion	RMP facilities are diverse in their size, structure, activities and the makeup of the regulated substances. As with many types of industrial facilities, there may be routine releases to the air and water of the residuals after pollution control devices remove what is generally a large fraction of the waste stream. Thus, people may be exposed to some substances directly through inhalation or indirectly through water routes or via ingestion of food. But the primary concerns with RMP facilities are the accidental release of substances and fires or explosions. The sudden release of relatively large quantities of acutely toxic substances can cause serious health effects including death after inhalation or dermal exposure. These effects may be prompt or may occur or persist for some time after exposure. Fires may affect neighboring areas and the associated smoke may expose people to toxic combustion products. Explosions may cause material damage and injuries to people in neighboring areas. Local

residents, as well as workers and emergency responders, may suffer severe adverse effects.

Some EPA EJ screening tools have included measures of proximity to all facilities regulated by EPA, which include RMP facilities.

More Information

More information is available at the RMP program webpage (https://www.epa.gov/rmp) and the RMP Info database stored in Envirofacts Facility Registry Service (FRS) (https://www.epa.gov/frs/frs-query).

Relevant Studies

The EJ literature contains numerous studies that have examined proximity to various types of sites, including some relevant to the possibility or frequency of chemical accidents. Since the 1980s, many studies have examined the frequency and consequences of accidental releases of acutely toxic chemicals or events resulting in fires or explosions (Binder, 1989, and many other studies). After the RMP program was established, researchers examined the characteristics of the RMP reporting facilities and their reported accident histories for insights into causes, consequences, prevention and emergency response (Kleindorfer et al., 2003).

Fewer studies have focused specifically on the relationship of RMP facilities to the demographics of the surrounding populations. Disparity in acute exposures to hazardous substances was addressed by Chakraborty (2001). The study by Deganian & Thompson (2012) included two categories of facilities—facilities having air pollution permits and facilities reporting to the Toxics Release Inventory program—which include some overlap with RMP facilities, but are not limited to RMP facilities. M. R. Elliott, Wang, Lowe, & Kleindorfer (2004) examined the characteristics of RMP-reporting facilities and their reported 5-year accident history versus the demographic characteristics of the counties in which they are located. The demographic characteristics examined included total population, race, education and income. The study found an association between the presence of larger and more chemical-intensive facilities and counties with larger African-American populations, and in counties with high levels of income inequality but higher median incomes. Further, the study found a greater risk of accidents for facilities in heavily African-American counties.

Data Source

Latitudes/Longitudes for RMP facilities were obtained from EPA's RMP database.

Data Version

The 2019 version of EJSCREEN uses locational information retrieved from the RMP database in June 2019. A total of 20,368 RMP facilities were included in the proximity indicators and related EJ indexes in this version of EJSCREEN. Note that the point data used to show facility locations in the "Map Supplementary Layers" menu is updated more often than the database with calculated EJSCREEN indicators and indexes, so in some small number of cases a facility may be in one data source but not the other.

Discussion

Each block group in the United States was assigned a proximity score that was the population-weighted sum of block-level proximity scores. Appendix C provides more details on the calculation of proximity scores. First, each block was given a proximity score that was the sum of inverse distance-weighted count of facilities anywhere within 5 km of the block's internal point. This score can be thought of as the number of RMP facilities per kilometer of distance from the average person. It is also equal to the number of facilities divided by the harmonic mean of their distances. This means one facility exactly 2 km away gives a score of 1/2, while three facilities exactly 4 km away give a score of 3/4, and five facilities all at 1 km away give a score of 5.[44] If there is no facility within 5 km of a block centroid, 1/d is used, where d is the distance to the single nearest at any distance.

As with all proximity-based indicators, proximity alone may not represent any actual risk or even exposure.

[44] An adjustment was made so that any distance smaller than the block's "radius" was set equal to 90% of that radius, with radius defined as the square root of (area/pi). This adjustment accounted for the fact that the average location (residence) within a circle of radius R is 0.9R away from the average point (facility) that is within the circle. For more detail, see Appendix C.

4 BIBLIOGRAPHY

Anderton, D. L., Anderson, A. B., Oakes, J. M., & Fraser, M. R. (1994). Environmental equity: The demographics of dumping. *Demography, 31*(2), 229-248.

Anderton, D. L., Oakes, J. M., & Egan, K. L. (1997). Environmental equity in Superfund. Demographics of the discovery and prioritization of abandoned toxic sites. *Eval Rev, 21*(1), 3-26.

Baden, B. M., Noonan, D. S., & Turaga, R. M. R. (2007). Scales of justice: Is there a geographic bias in environmental equity analysis? *J Environ Planning Manage, 50*(2), 163-185.

Baumann, L. M., Robinson, C. L., Combe, J. M., Gomez, A., Romero, K., Gilman, R. H., . . . Checkley, W. (2011). Effects of distance from a heavily transited avenue on asthma and atopy in a periurban shantytown in Lima, Peru. *J Allergy Clin Immunol, 127*(4), 875-882.

Been, V., & Gupta, F. (1997). Coming to the Nuisance or Going to the Barrios-A Longitudinal Analysis of Environmental Justice Claims. *Ecol Law Q, 24*, 1-56.

Bell, M. L., & Ebisu, K. (2012). Environmental Inequality in Exposures to Airborne Particulate Matter Components in the United States. *Environ Health Perspect, 120*(12), 1699-1704.

Bell, M. L., Zanobetti, A., & Dominici, F. (2014). Who is more affected by ozone pollution? A systematic review and meta-analysis. *Am J Epidemiol*, *180*(1), 15-28. https://academic.oup.com/aje/article/180/1/15/2739254

Berrocal, V. J., Gelfand, A. E., & Holland, D. M. (2010a). A bivariate space-time downscaler under space and time misalignment. *Ann Appl Stat, 4*(4), 1942-1975.

Berrocal, V. J., Gelfand, A. E., & Holland, D. M. (2010b). A Spatio-Temporal Downscaler for Output From Numerical Models. *J Agric Biol Environ Stat, 15*(2), 176-197.

Berrocal, V. J., Gelfand, A. E., & Holland, D. M. (2011). Space-Time Data fusion Under Error in Computer Model Output: An Application to Modeling Air Quality. *Biometrics*, First published online: 29 Dec 2011.

Binder, S. (1989). Deaths, injuries, and evacuations from acute hazardous materials releases. *Am J Public Health, 79*(8), 1042-1044.

Boer, J. T., Pastor Jr., M., Sadd, L. L., & Synder, L. D. (1997). Is There Environmental Racism? The Demographics of Hazardous Waste in Los Angeles County. *Soc Sci Quart, 78*(4), 793-810.

Bravo, M. A., Fuentes, M., Zhang, Y., Burr, M. J., & Bell, M. L. (2012). Comparison of exposure estimation methods for air pollutants: Ambient monitoring data and regional air quality simulation. *Environ Res, 116*(1-10).

Brender, J. D., Maantay, J. A., & Chakraborty, J. (2011). Residential Proximity to Environmental Hazards and Adverse Health Outcomes. *Am J Public Health, 101*(S1), S19-S26.

Briguglio, L. (1997). Alternative Economic Vulnerability Indices for Developing Countries. Report prepared for the Expert Group on Vulnerability Index. UN (DESA), 17-19, December.

Brochu, P. J., Yanosky, J. D., Paciorek, C. J., Schwartz, J., Chen, J. T., Herrick, R. F., & Suh, H. H. (2011). Particulate air pollution and socioeconomic position in rural and urban areas of the Northeastern United States. *Am J Public Health, 101*(S1), S224-S230.

Byun, D., & Schere, K. (2006). Review of the governing equations, computational algorithms, and other components of the models-3 Community Multiscale Air Quality (CMAQ) modeling system. *Appl Mech Rev, 59*, 51-77.

Cauthen, N. K., & Fass, S. (2008). Measuring Poverty in the United States: National Center for Children in Poverty. Mailman School of Public Health.

Centers for Disease Control and Prevention. (2010). Behavioral Risk Factor Surveillance System Survey Data. Atlanta: U.S. Department of Health and Human Services. http://www.cdc.gov/brfss/

Centers for Disease Control and Prevention. (2011a). *CDC Health Disparities and Inequalities Report — United States, 2011. MMWR, January 14, 2011; Vol. 60 (Suppl)*. Retrieved from http://www.cdc.gov/mmwr/pdf/other/su6001.pdf.

Centers for Disease Control and Prevention. (2011b). *Deaths: Leading Causes for 2007. National Vital Statistics Reports. Vol. 59, No. 8*. Retrieved from http://www.cdc.gov/nchs/data/nvsr/nvsr59/nvsr59_08.pdf.

Centers for Disease Control and Prevention. (2011c). Epi Info™. Retrieved April 12, 2012, from http://wwwn.cdc.gov/epiinfo/.

Centers for Disease Control and Prevention. (2011d). *MMWR, September 9, 2011; Vol. 60, No. 35*. Retrieved from http://www.cdc.gov/mmwr/pdf/wk/mm6035.pdf.

Centers for Disease Control and Prevention. (2012). *CDC Response to Advisory Committee on Childhood Lead Poisoning Prevention Recommendations in "Low Level Lead Exposure Harms Children: A Renewed Call of Primary Prevention"*. Retrieved from http://www.cdc.gov/nceh/lead/acclpp/cdc_response_lead_exposure_recs.pdf.

Chakraborty, J. (2001). Acute Exposure to Extremely Hazardous Substances: An Analysis of Environmental Equity. *Risk Analysis*, 21(5), 883-894.

Chakraborty, J. (2006). Evaluating the Environmental Justice Impacts of Transportation Improvement Projects in the US. *Transp Res D Transp Environ*, 11(5), 315-323.

Chakraborty, J., Maantay, J. A., and Brender, J. D. (2011). Disproportionate Proximity to Environmental Health Hazards: Methods, Models, and Measurement. *Am J Public Health*, 101 (S1), S27-S36.

Clark, L. P., Millet, D. B., & Marshall, J.D. (2014). National Patterns in Environmental Injustice and Inequality: Outdoor NO2 Air Pollution in the United States. *PLoS ONE 9*(4): e94431. http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0094431

Cohen, R. A., & Martinez, M. E. (2011). Health Insurance Coverage: Early Release of Estimates From the National Health Interview Survey, January–March 2011: National Center for Health Statistics.

deFur, P. L., Evans, G. W., Cohen Hubal, E. A., Kyle, A. D., Morello-Frosch, R. A., & Williams, D. R. (2007). Vulnerability as a Function of Individual and Group Resources in Cumulative Risk Assessment. *Environ Health Perspect, 115*(5), 817-824.

Deganian, D., & Thompson, J. (2012). The Patterns of Pollution: A Report on Demographics and Pollution in Metro Atlanta: GreenLaw.

Elliott, M. R., Wang, Y., Lowe, R. A., & Kleindorfer, P. R. (2004). Environmental justice: Frequency and severity of US chemical industry accidents and the socioeconomic status of surrounding communities. *J Epidemiol Community Health, 58*, 24-30.

Environmental Health Sciences. Environmental Health News. Online periodical. Retrieved August 24, 2012, from www.environmentalhealthnews.org.

Fann, N., Lamson, A. D., Anenberg, S. C., Wesson, K., Risley, D., & Hubbell, B. J. (2012). Estimating the national public health burden associated with exposure to ambient PM2.5 and ozone. *Risk Analysis, 32*(1), 81-95.

Fann, N., & Risley, D. (2011). The public health context for PM2.5 and ozone air quality trends. *Air Qual Atmos Health*, [Epub ahead of print January 5, 2011].

Fann, N., Roman, H. A., Fulcher, C. M., Gentile, M. A., Hubbell, B. J., Wesson, K., & Levy, J. I. (2011). Maximizing health benefits and minimizing inequality: Incorporating local-scale data in the design and evaluation of air quality policies. *Risk Analysis, 31*(6), 908-922.

Fedstats. (2007). FedStats. Retrieved April 12, 2012, from http://fedstats.sites.usa.gov/.

Finkel, A. M., & Golding, D. (Eds.). (1994). *Worst Things First? The Debate Over Risk-Based National Environmental Priorities*. Washington DC: Resources for the Future.

Fitos, E., and Chakraborty, J. (2010). Race, Class, and Wastewater Pollution. In J. Chakraborty and M. M. Bosman (Eds.) *Spatial and Environmental Injustice in an American Metropolis: A Study of Tampa Bay, Florida.* Cambria Press: Amherst, NY; 59-82.

Fujita, E. M., Campbell, D. E., Zielinska, B., Arnott, W. P., & Chow, J. C. (2011). Concentrations of air toxics in motor vehicle-dominated environments, Research Report 156. Boston: Health Effects Institute.

Gaitens, J. M., Dixon, S. L., Jacobs, D. E., Nagaraja, J., Wilson, J. W., & Ashley, P. J. (2009). Exposure of U.S. children to residential dust lead, 1999-2004: I. Housing and demographic factors. *Environ Health Perspect, 117*(3), 461-467.

Galea, S., Tracy, M., & Hoggatt, K. J. (2011). Estimated Deaths Attributable to Social Factors in the United States. *Am J Public Health, 101*(8), 1456-1465.

General Accounting Office. (1983). *Siting of Hazardous Waste Landfills and their Correlation with Racial and Economic Status of Surrounding Communities*. Washington, DC: Retrieved from http://www.gao.gov/products/121648.

Grineski, S. (2007). Incorporating Health Outcomes into Environmental Justice Research: The Case of Children's Asthma and Air Pollution in Phoenix, Arizona. *Environmental Hazards*, 7, 360–371.

Grineski, S., Bolin, B., and Boone, C. (2007). Criteria Pollution and Marginal Populations: Environmental Inequity in metropolitan Phoenix, Arizona, USA. *Soc Sci Q*, 88, 535–554.

Health Effects Institute. (2010). Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects (HEI Special Report 17).

Hird, J. A. (1993). Environmental policy and equity: The case of superfund. *J Policy Anal Manag, 12*(2), 323-343.

Hoek, G., Beelen, R., Kos, G., Dijkema, M., Zee, S. C. v. d., Fischer, P. H., & Brunekreef, B. (2011). Land Use Regression Model for Ultrafine Particles in Amsterdam. *Environ Sci Technol, 45*(2), 622-628.

Hoffman, B., Moebus, S., Dragano, N., Möhlenkamp, S., Memmesheimer, M., Erbel, R., & Jöckel, K. H. (2009). Residential traffic exposure and coronary heart disease: results from the Heinz Nixdorf Recall Study. *Biomarkers, 14*(S1), 74-78.

Hystad, P., Setton, E., Cervantes, A., Poplawski, K., Deschenes, S., Brauer, M., . . . Demers, P. (2011). Creating national air pollution models for population exposure assessment in Canada. *Environ Health Perspect, 119*(8), 1123-1129.

Jacobs, D. E., Clickner, R. P., Zhou, J. Y., Viet, S. M., Marker, D. A., Rogers, J. W., . . . Friedman, W. (2002). The Prevalence of Lead-Based Paint Hazards in U.S. Housing. *Environ Health Perspect, 110*(10), a599-a606.

Jones, R. L., Homa, D. M., Meyer, P. A., Brody, D. J., Caldwell, K. L., Pirkle, J. L., & Brown, M. J. (2009). Trends in blood lead levels and blood lead testing among US children aged 1 to 5 years, 1988-2004. *Pediatrics, 123*(3), e376-e385.

Kleindorfer, P. R., Belke, J. C., Elliott, M. R., Lee, K., Lowe, R. A., & Feldman, H. I. (2003). Accident epidemiology and the U.S. chemical industry: Accident history and worst-case data from RMP*Info. *Risk Analysis, 23*(5), 865-881.

Krewski, D., Jerrett, M., Burnett, R. T., Ma, R., Hughes, E., Shi, Y., . . . Thun, M. J. (2009). Extended Follow-Up and Spatial Analysis of the American Cancer Society Study Linking Particulate Air Pollution and Mortality, Research Report 140. Boston: Health Effects Institute.

Levin, R., Brown, M. J., Kashtock, M. E., Jacobs, D. E., Whelan, E. A., Rodman, J., . . . Sinks, T. (2008). Lead Exposures in U.S. Children, 2008: Implications for Prevention. *Environ Health Perspect, 116*(10), 1285-1293.

Levy, J. I., Wilson, A. M., & Zwack, L. M. (2007). Quantifying the Efficiency and Equity Implications of Power Plant Air Pollution Control Strategies in the United States. *Environ Health Perspect, 115*(5), 743-750.

Liu, F. (2001). *Environmental Justice Analysis: Theories, Methods, and Practice*. Boca Raton: CRC Press. https://www.crcpress.com/product/isbn/9781566704038

Lobdell, D. T., Jagai, J. S., Rappazzo, K., & Messer, L. C. (2011). Data sources for an environmental quality index: Availability, quality, and utility. *Am J Public Health, 101*(S1), S277-S285.

Maantay, J., Chakraborty, J., & Brender, J. (2010). Proximity to Environmental Hazards: Environmental Justice and Adverse Health Outcomes, Final Revised Draft. Prepared for the U.S. Environmental Protection Agency “Strengthening Environmental Justice Research and Decision Making: A Symposium on the Science of Disproportionate Environmental Health Impacts”. Final publication available at http://dx.doi.org/10.2105/AJPH.2011.300183

Maroko, A. R. (2012). Using air dispersion modeling and proximity analysis to assess chronic exposure to fine particulate matter and environmental justice in New York City. *Appl Geogr, 34*, 533-547.

McMillan, N. J., Holland, D. M., Morara, M., & Feng, J. (2010). Combining numerical model output and particulate data using Bayesian space–time modeling. *Environmetrics, 21*(1), 48-65.

Messer, L. C., Jagai, J. S., Rappazzo, K. M., Lobdell, D. T. (2014). Construction of an environmental quality index for public health research. Environ Health. May 22;13(1):39. http://dx.doi.org/10.1186%2F1476-069X-13-39

Miranda, M. L., Edwards, S. E., Keating, M. H., & Paul, C. J. (2011). Making the Environmental Justice Grade: The Relative Burden of Air Pollution Exposure in the United States. *Int J Environ Res Public Health, 8*(6).

Mohai, P., & Saha, R. (2007). Racial Inequality in the Distribution of Hazardous Waste: A National-Level Reassessment. *Social Problems, 54*(3), 343-270.

Morello-Frosch, R., & Jesdale, B. M. (2006). Separate and Unequal: Residential Segregation and Estimated Cancer Risks Associated with Ambient Air Toxics in U.S. Metropolitan Areas. *Environ Health Perspect, 114*(3), 386-393.

National Research Council of the National Academies Committee on Improving Risk Analysis Approaches Used by the U.S. EPA. (2009). *Science and Decisions: Advancing Risk Assessment*. Washington, DC: The National Academies Press. http://www.nap.edu/catalog/12209/science-and-decisions-advancing-risk-assessment

National Environmental Justice Advisory Council (NEJAC). (2010). *Recommendations for Nationally Consistent Environmental Justice Screening Approaches*. http://www.epa.gov/environmentaljustice/resources/publications/nejac/ej-screening-approaches-rpt-2010.pdf via https://www.epa.gov/sites/production/files/2015-02/documents/fact-sheet-nejac_0.pdf

O'Neil, S. G. (2007). Superfund: Evaluating the Impact of Executive Order 12898. *Environ Health Perspect, 115*(7), 1087-1093.

Oakes, J. M., Anderton, D. L., & Anderson, A. B. (1996). A Longitudinal Analysis of Environmental Equity in Communities with Hazardous Waste Facilities. *Soc Sci Res, 25*, 125-148.

The Organisation for Economic Co-operation and Development (OECD). (2008). Handbook on Constructing Composite Indicators. http://www.oecd.org/els/soc/handbookonconstructingcompositeindicatorsmethodologyanduserguide.htm

Pastor Jr., M., Morello-Frosch, R., & Sadd, J. (2010). Final Report: Contract Number 04-308. Air Pollution and Environmental Justice: Integrating Indicators of Cumulative Impact and Socio-Economic Vulnerability into Regulatory Decision-Making.

Pastor Jr., M., Sadd, J., & Hipp, J. (2001). Which Came First? Toxic Facilities, Minority Move-In, and Environmental Justice. *J Urban Affairs, 23*(1), 1-21.

Post, E. S., Belova, A., & Huang, J. (2011). Distributional Benefit Analysis of a National Air Quality Rule. *Int J Environ Res Public Health, 8*(6), 1872-1892.

Probst, K. N. (1990). *Hazardous Waste and the Rural Poor: A Preliminary Assessment*: Clean Sites, Inc.

Rosenbaum, A., Hartley, S., & Holder, C. (2011). Analysis of diesel particulate matter health risk disparities in selected US harbor areas. *Am J Public Health, 101*(S1), S217-S223.

Rowangould, G.M. (2013). A census of US near-roadway population: Public health and environmental justice considerations. *Transp Res D Transp Environ,* 25, 59-67.

Sadd, J. L., Pastor, M., Morello-Frosch, R., Scoggins, J., & Jesdale, B. (2011). Playing it safe: Assessing cumulative impact and social vulnerability through an environmental justice screening method in the South Coast Air Basin, California. *Int J Environ Res Public Health, 8*(5), 1441-1459.

Saha, R., & Mohai, P. (2005). Historical Context and Hazardous Waste Facility Siting: Understanding Temporal Patterns in Michigan. *Social Problems, 52*(4), 618-648.

Sørensen, M., Hvidberg, M., Andersen, Z. J., Nordsborg, R. B., Lillelund, K. G., Jakobsen, J., . . . Raaschou-Nielsen, O. (2011). Road traffic noise and stroke: A prospective cohort study. *Eur Heart J, 32*(6), 737-744.

Spengler, J., Lwebuga-Mukasa, J., Vallarino, J., Melly, S., Chillrud, S., Baker, J., & Minegishi, T. (2011). Air Toxics Exposure from Vehicle Emissions at a U.S. Border Crossing: Buffalo Peace Bridge Study, Research Report 158. Boston: Health Effects Institute.

Su, J. G., Morello-Frosch, R., Jesdale, B. M., Kyle, A. D., Shamasunder, B., & Jerrett, M. (2009). An index for assessing demographic inequalities in cumulative environmental hazards with application to Los Angeles, California. *Environ Sci Technol, 43*(20), 7626-7634.

Tian, N., Xue, J., & Barzyk, T.M. (2013). Evaluating socioeconomic and racial differences in traffic-related metrics in the United States using a GIS approach. *J Exposure Science Envt Epidem,* 23, 215-222.

U.S. Census Bureau. (2016). American Community Survey, 5-year Estimates 2011-2015. Data retrieved in 2016, from http://www2.census.gov/acs2015_5yr/summaryfile/ and https://www.census.gov/programs-surveys/acs/data/summary-file.html

U.S. Department of Transportation. (2011). National Transportation Atlas Database. Highway Performance Monitoring System. Data retrieved April 12, 2012, from http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transportation_atlas_database/2011/index.html.

U.S. EPA. (1987). *Unfinished Business: A Comparative Assessment of Environmental Problems*. Washington, DC: Retrieved from http://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=2000BZOP.txt.

U.S. EPA. (1992). *Environmental Equity: Reducing the Risk for All Communities*. Washington, DC: Retrieved from http://nepis.epa.gov/Exe/ZyNET.exe/40000JLA.TXT.

U.S. EPA. (2006a). *Air Quality Criteria for Lead (2006) Final Report*. EPA/600/R-05/144aF-bF. Washington, DC. http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=158823

U.S. EPA. (2006b). *Air Quality Criteria for Ozone and Related Photochemical Oxidants*. Washington, DC: Retrieved from http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=149923.

U.S. EPA. (2009a). *A Framework for Categorizing the Relative Vulnerability of Threatened and Endangered Species to Climate Change (External Review Draft)*. EPA/600/R-09/011. Washington, DC. http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=203743

U.S. EPA. (2009b). *Integrated Science Assessment for Particulate Matter (Final Report)*. Washington, DC: Retrieved from http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=216546.

U.S. EPA. (2009c). National-Scale Air Toxics Assessment for 2002 - Fact Sheet. Retrieved April 12, 2012, from https://www.epa.gov/national-air-toxics-assessment/nata-overview.

U.S. EPA. (2009d). Original list of hazardous air pollutants. Retrieved April 12, 2012, from http://www.epa.gov/ttnatw01/187polls.html.

U.S. EPA. (2015). *Guidance on Considering Environmental Justice During the Development of an Action*. Retrieved from https://yosemite.epa.gov/sab/sabproduct.nsf/4D46D5F60CC33393852578650059C7CF/$File/considering-ej-in-rulemaking-guide-07-2010.pdf.

U.S. EPA. (2011a). *Exposure Factors Handbook: 2011 Edition*. Washington, DC: Retrieved from http://www.epa.gov/ncea/efh/pdfs/efh-complete.pdf.

U.S. EPA. (2011c). *Integrated Science Assessment for Lead (1st External Review Draft 2011 & Final Report 2013)*. Washington, DC. EPA/600/R-10/075F. http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=255721

U.S. EPA. (2011e). RCRAInfo database search in Envirofacts. Data retrieved April 12, 2012, from http://www.epa.gov/enviro/facts/rcrainfo/search.html.

U.S. EPA. (2012a). Facilities and Enforcement Activities related to the Clean Water Act National Pollutant Discharge Elimination System (NPDES) Program. Data retrieved April 12, 2012, from https://www3.epa.gov/enviro/facts/rcrainfo/search.html .

U.S. EPA. (2012b). Industrial Regulations. Retrieved April 12, 2012, from http://water.epa.gov/scitech/wastetech/guide/industry.cfm.

U.S. EPA. (2012c). Integrated Risk Information System (IRIS). Retrieved April 12, 2012, from http://www.epa.gov/IRIS/.

U.S. EPA. (2012d). *Our Nation's Air: Status and Trends Through 2010*. Research Triangle Park: Retrieved from http://www.epa.gov/airtrends/2011/report/fullreport.pdf.

U.S. EPA. (2012e). RCRA Orientation Manual 2011: Resource Conservation and Recovery Act. Retrieved August 24, 2012, from https://www.epa.gov/hwgenerators/resource-conservation-and-recovery-act-rcra-orientation-manual.

U.S. EPA. (2012f). Search Superfund Site Information. Data retrieved May, 2012, from http://cumulis.epa.gov/supercpad/cursites/srchsites.cfm.

U.S. EPA. (2012g). Summary Nonattainment Area Population Exposure Report. Retrieved August 24, 2012, from http://nepis.epa.gov/Exe/ZyNET.exe/P100GD35.TXT?ZyActionD.

U.S. EPA. (2015a). Fused Air Quality Surfaces Using Downscaling. Data retrieved from (website address pending).

U.S. EPA. (2015b). National-Scale Air Toxics Assessments. Data retrieved from http://www.epa.gov/ttn/atw/natamain/index.html.

United Church of Christ. (1987). Toxic Waste 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. http://www.ucc.org/environmental-ministries_toxic-waste-20

United Church of Christ. (2007). Toxic Wastes and Race at Twenty: 1987-2007. Cleveland: Justice and Witness Ministries. http://www.ucc.org/environmental-ministries_toxic-waste-20

van Donkelaar, A., Martin, R. V., Brauer, M., Kahn, R., Levy, R., Verduzco, C., & Villeneuve, P. J. (2010). Global Estimates of Ambient Fine Particulate Matter Concentrations from Satellite-Based Aerosol Optical Depth: Development and Application. *Environ Health Perspect, 118*(6), 847-855. http://dx.doi.org/10.1289/ehp.0901623

VanDerslice, J. (2011). Drinking water infrastructure and environmental disparities: evidence and methodological considerations. *Am J Public Health, 101*(S1), S109-S114.

Villagrán de Léon, J. C. (2006). Vulnerability: A Conceptual and Methodological Review (Vol. 4/2006): SOURCE (Studies Of the University: Research, Counsel, Education), United Nations University - Institute for Environmental and Human Security.

Yang, T.-C., & Matthews, S. A. (2010). The Role of Social and Built Environments in Predicting Self-rated Stress: A Multilevel Analysis in Philadelphia. *Health & Place, 16*(5), 803-810.

Zhou, Y., & Levy, J. I. (2007). Factors influencing the spatial extent of mobile source air pollution impacts: A meta-analysis. *BMC Public Health, 7*(89).

Zhu, Y., Hinds, W. C., Kim, S., & Sioutas, C. (2002). Concentration and size distribution of ultrafine particles near a major highway. *J Air Waste Manag Assoc, 52*(9), 1032-1042.

Zimmerman, R. (1993). Social Equity and Environmental Risk. *Risk Analysis, 13*(6), 649-666.

APPENDIX A. DEVELOPMENT OF EJSCREEN

Review of Available Data and Other Tools

Preliminary planning for EJSCREEN began in late 2010. The first steps involved a review of existing or planned EJ screening methods from EPA program and regional offices. EJSCREEN draws upon a great deal of prior research, analysis and public involvement in the development of very closely related screening efforts. Early steps also included a review of current EJ research on methods and data for EJ analysis. Information was gathered from the following sources, among others:

- Stakeholder and expert presentations at EPA's March 2010 conference on environmental justice.
- EPA's 2010 expert workshop on economics and environmental justice.
- EPA's ORD's C-FERST research program, including a review of data sources for EJ analysis.
- EPA's ORD's Environmental Quality Index (EQI) compilation and review of data sources for environmental indicators.
- A review of several national reviews of analytic methods including the use of inequality metrics (e.g. as presented in the expert workshop).
- Review of EPA guidance documents and related documents on environmental justice[45]
- Review of the NEJAC report of May 2010 on EJ screening (NEJAC, 2010).
- Review of prior tools including EJSEAT, EJVIEW, various EPA Regional tools, and some state EJ screening tools such as CalEnviroScreen.

Selecting an Approach to EJ Screening

A number of important considerations must be balanced when selecting an approach to an EJ screening tool:

- Useful to end-users and other stakeholders.
- Reflects EPA policies and EJ policy goals.
- Reflects sound science.
- Is feasible to develop and maintain, update and upgrade.

Data coverage and quality considerations are also discussed in the chapters describing the environmental indicators.

EJSCREEN was developed through an EPA workgroup with participation from a very wide range of program offices and Regional offices, and in consultation with management and scientists representing the various offices, building upon the public input and scientific information developed in the course of prior screening efforts such as EJSEAT and Regional experience with EJ screening. Quality control and peer review of EJSCREEN are described in Appendices F and G.

[[45] https://www.epa.gov/sites/production/files/2014-08/documents/enviro_justice_309review.pdf

[intentionally blank]

APPENDIX B. TECHNICAL DETAILS ON PERCENTILES, ROUNDING, BUFFERING, AND DEMOGRAPHIC DATA

How Percentiles are Calculated and Displayed

Percentiles, such as "80th percentile," are displayed in EJSCREEN as rounded down to the closest percentile that is lower than the exact value. This is called showing the "floor" of the exact percentile (rather than rounding off the percentile to the nearest 1 percentile). For example, if the exact percentile is equal to or greater than 79 but less than 80, it is displayed as "79th percentile." If the exact percentile is equal to or greater than 80 but less than 81, it is displayed as "80th percentile." The reason for this is to ensure that EJSCREEN only displays "80th percentile" if the exact percentile truly is as high as 80.[46]

Ties in indicator values are fairly common, especially for the lead paint indicator and percent linguistic isolation, where large shares are tied with values of zero. Ties are assigned a percentile that can be thought of as the upper edge of the range of tied values. For example, if 3 percent of the US population were tied at the maximum indicator value, they would all be shown as being at the 100th percentile, and the next lower value would be assigned the 97th percentile. If 4 percent were all tied for the lowest value, they would all be shown as being at the 4th percentile, and nobody would be shown as being at the 0-3 percentiles. The percent linguistic isolation was zero in block groups comprising about 45% of the US population in the 2011-2015 ACS data, so all those places were shown as tied for the 45th percentile, and none were reported as being at any lower percentiles. However, it is worth noting that a group of tied values is usually not shared by more than 1% of the population, so once the percentiles are converted to integers 0-100, the tied raw values do not cause jumps in percentiles except in a few cases. A jump would be a case where there are no block groups assigned a percentile between zero and 45, because around 45% of the population is tied with a value of zero.

Percentiles are assigned to calculated values (such as in buffer reports) by use of national, region-specific, and state-specific lookup tables that show the raw value cutoff value that corresponds to each integer percentile 0-100. To ensure that exact matches are found when looking for the 100th percentile, for example, the cutoff values in the lookup tables are all stored with exactly 6 decimal places, and a raw value is rounded to exactly 6 decimal places before it is looked up in those tables. If a value matches the cutoff, it is assigned that percentile. If it falls between two cutoffs, it is assigned the lower of the two percentiles, to provide displayed results that are consistent with the way percentiles are displayed using the "floor" function described above. The lookup tables are stored in a geodatabase used for EJCSREEN.

[46] This also ensures that map colors correspond to displayed percentiles. For example, if the exact percentile is 79.99, the map will show the place as gray, meaning it is still below the 80th percentile, and the percentile will be shown as "79th percentile." If the map is yellow, it indicates the exact percentile is at least 80, and the displayed percentile will also be at least 80. Without using the "floor" of the exact percentile, the map colors and displayed percentile would sometimes disagree, and a user would not be sure if "80th percentile" actually meant the exact percentile was actually at least 80. Using the "floor" instead of rounding ensures clarity about whether a place actually reaches a given percentile.

In output tables, percentages are rounded to the nearest percent but percentiles are displayed using a "floor" function as described above. Occasionally, this can lead to some potentially confusing situations in some tables. For example, a place may be shown as 100% minority but only at the 98th percentile. This is because the place may actually be 99.6% minority, which is displayed as 100% minority. But if 1.8% of the US population lives where there is an even higher percent minority (e.g., 100%), this place is only at the 98.2 percentile, which is displayed as 98th percentile.

The percentiles and lookup tables were calculated using the statistical software called R, using code written by EPA, based on wtd.quantile() and wtd.Ecdf() functions in the Hmisc package (http://cran.r-project.org/web/packages/Hmisc/index.html). The scripting language R is documented here: http://cran.r-project.org

How Percentages and Raw Values are Rounded and Displayed

ESJSCREEN displays raw indicator numbers and percentiles in a standard report (on-screen or PDF format), the "Explore reports" window, a tabular view (on-screen or downloaded text file), bar graphs, popup windows on maps, and in the downloadable raw data files. Several standard rules have been applied to keep these formats consistent, clear, and at an appropriate level of detail.

The raw data stored in the database used in EJSCREEN are stored as the "exact" values calculated from estimated counts obtained from the Bureau of Census, or from the development of environmental indicators, at the highest degree of precision used by the software calculating the indicators and by the GIS database. This ensures that all internal calculations use the best estimate of a given number rather than relying on a rounded off approximation, and is standard best practice in working with such data. All calculations use the "exact" (unrounded) numbers from Census or stored in the GIS database. This includes converting raw Census data into a demographic indicator, calculating an EJ index, or estimating the values for a buffer report.

When displaying data, such as in reports, popup windows on maps, or the tabular view, EJSCREEN presents formatted numbers that follow certain conventions:

- Raw environmental indicator values are displayed using specified numbers of significant figures (also known as significant digits). This is a standard way of communicating precision appropriately. Precision of these estimates depends largely on sample size in Census survey data, and the ability of measurements and models to estimate environmental conditions. Two significant figures are shown for all environmental indicators other than PM2.5, ozone, and diesel PM, which are shown with three significant figures. For example, a cancer risk calculated to be 144.44 per million would be displayed as 140 (i.e., using 2 significant figures). A PM2.5 concentration of 14.44 would be shown as 14.4 (i.e., using 3 significant figures). Proximity scores of at least 0.185 but less than 0.195 would be displayed as 0.19 (which shows 2 significant figures). This means a proximity score displayed as 0.010 came from an exact value of at least 0.0095 but less than 0.015, for example. Note that these significant figure rules have been applied in all cases, and if in some case the number is missing a trailing zero that should appear, it is simply a limitation of print formatting. For example, if a proximity score is shown as 0.1, best practice would be to display it as 0.10 to make explicit the use of 2 significant figures, a printout may only display it as 0.1 instead of 0.10, but it was still rounded using the 2 significant figures rule.

- Demographic percentages, such as "34% low-income," are displayed as rounded to the nearest 1%. For example, any values equal to or greater than 79.5% but less than 80.5% are displayed as "80%."

It is also important to keep in mind that all of the numbers are estimates, so small differences in raw values or percentiles should not be regarded as certain and meaningful, given the uncertainty in the environmental and demographic estimates.

Calculations for Buffer Reports

EJSCREEN allows a user to define a buffer, such as the circle that includes everything within 1 mile of a specific point. Non-circular, user-defined shapes also can be defined to represent buffers of any shape. A report summarizes the demographics of residents within this buffer, as well as the environmental indicators and EJ index values within the buffer.

The summary within a buffer is designed to represent the average resident within the buffer, and also provides an estimate of the total population residing in the buffer. For example, the traffic proximity indicator for a buffer is the population-weighted average of all the traffic indicator values in the buffer. Similarly, the percent minority would be a weighted average, which is the same as the overall percent minority for all residents in the buffer.

Some block groups will be partly inside and partly outside a buffer, and any buffer analysis must estimate how much of each block group's population is inside the buffer. Areal apportionment of block groups is one standard method, but it assumes that population is evenly spread throughout a block group, which may be far from the actual distribution of residents. Areal apportionment of blocks would be even more accurate but extremely computationally intensive.

To provide the most accurate counts that are currently feasible for a screening tool, EJSCREEN uses an approach based on Census block internal points. EJSCREEN estimates the fraction of the Census block group population that is inside the buffer by using *block*-level population counts from Census 2010. These blocks provide data about where residents are at a higher resolution than block groups. Each block has an internal point defined by the Census Bureau, and the entire block population is counted as inside or outside the buffer depending on whether the block internal point is inside or outside. This assumption typically introduces relatively little error because blocks are so small relative to a typical buffer, so a small fraction of the total buffer population is in blocks that span an edge of the buffer. Also, any blocks along the edge of a buffer whose populations are close to 0 or 100% inside the buffer will be well represented by this assumption.

As long as users draw buffers much larger than a local block group, this method should represent the average person inside the buffer reasonably well.

The calculation of a value for the buffer is essentially the population-weighted average of the indicator values in the blocks included in the buffer, where each block uses the indicator values of the block group containing it. A block group is weighted based on the fraction of the ACS block group population that is considered in the buffer. That fraction is estimated as the Census 2010 block population divided by the Census 2010 block group population. The formula below is used to estimate the population average of a raw indicator value in a buffer. This formula is simply a population-weighted average – it sums the population-weighted raw values, and then divides that sum by the total population in the buffer.

$$Value(A) = \sum_{\forall Blk, Blk \cap A} \frac{\frac{BlockPop10}{BGPop10} * BGACSPop * BG\_RawValue}{\sum_{\forall Blk, Blk \cap A} \frac{BlkPop10}{BGPop10} * BGACSPop}$$

"BlockPop10" refers to the Census 2010 block level population total (used here because the ACS does not provide block resolution), and "BG" indicates block group. "BGACSPop" is the block group estimated population count from the ACS, which is often different than the Census 2010 total for all blocks in the block group, because the ACS data used here is a composite estimate based on survey samples spanning five years, while the Census is a full count at one point in time.

Demographic Data and Geographic Coverage

In the first decade of this century, the Census Bureau made a fundamental shift in how detailed demographic data are collected. Rather than collecting basic data from everyone, plus more detailed data from a one-in-six sample of households once a decade in the decennial census, a mixed approach has been adopted. The basic data, required for Congressional redistricting under the U.S. Constitution, are still collected every ten years in what is intended to be a 100% census. But that basic information, plus virtually all the more detailed demographic data, are also collected throughout each year in a stratified random sample of more than 200,000 households each month. This is the American Community Survey (ACS). Some of this information is then aggregated and displayed in yearly summaries, others in 3-year summaries, and others in 5-year summaries. Only the five-year summary files provide block group resolution. The result is a timelier, evolving picture of U.S. demographics. For instance, the ACS 2013 to 2017 average data were released in December 2018, while most demographics data users were still working with the April 2010 decennial census snapshot.

Extensive documentation of the ACS is available. For a general overview, see https://www.census.gov/programs-surveys/acs/ and for complete documentation see https://www.census.gov/content/dam/Census/newsroom/press-kits/2015/20150212_workshop_acso.pdf. For information on the 5-year summary file, which is what EJSCREEN uses, see https://www.census.gov/programs-surveys/acs/ . For information on using the data see https://www.census.gov/programs-surveys/acs/guidance/handbooks.html).

Race and ethnicity, the two items that determine minority status in our approach, are available from the 100% enumeration from the decennial census or the ACS, while all the other measures are only contained in ACS estimates. For the purposes of EJSCREEN, EPA did not believe that the increased precision of the minority measures that might be gained from combining those once per decade data with the other data items from ACS were worth the problems and ambiguities entailed in such a hybrid approach.

All of EJSCREEN's demographic data come from the latest annual update of the five-year average ACS estimates, with some lag time from publication by Census to inclusion in EJSCREEN. The Census Bureau does not recommend making direct comparisons between data from a five-year summary and a prior, overlapping five-year period, such as comparing 2007-2011 to 2011-2015.[47] This means attempting to look for trends in terms of year-to-year changes is not recommended – changes in ACS estimates at block group resolution can be reviewed every five years, but not more often. Yearly changes can be examined at county resolution using the 1-year ACS data.

Each of the nation's counties (or county equivalents, such as Municipios in Puerto Rico) is completely divided into Census tracts. Each tract is in turn divided into Census block groups. Census block groups generally have between 600 and 3,000 people, with an optimal population of 1,500; however, a few are much more populous and a small number have zero residents. A block group consists of one or more Census blocks. In urban areas, a block is typically a city block defined by streets. The Census Bureau collects data by household, but block groups are the smallest area for which the ACS presents estimates.

ACS 2013-2017 summary file tracts and block groups use Census 2017 boundaries and FIPS codes.[48]

Also, note that since NATA is collected using Census 2010 codes, details of how NATA data is converted to 2014 boundaries is discussed in Section 3.

There are 11,078,297 blocks in the Census 2010 data, not including Puerto Rico and the Island Areas,[49] but since then Census has made some changes in the FIPS codes that relate blocks to their parent block groups. Block to block group relationships are used by EJSCREEN in part of the buffer analysis (as described in Appendix C). This required making manual adjustments to reassign some Census 2010 blocks to their updated parent ACS block groups for purposes of calculating buffer estimates.

The EJSCREEN dataset based on the ACS 2013-2017 summary file has data for States/equivalents (includes DC) and Puerto Rico, counties/county equivalents, census tracts, and block groups that differ from the decenial census. The Census Bureau does not collect ACS data for the Virgin Islands or the other Island Territories.

The count of geographies (e.g., number of block groups) covered by the ACS is summarized by the Census Bureau for each update:

- ACS tallies, as counted in data downloaded from ACS FTP site, and as used in EJSCREEN (217,739 block groups, without Puerto Rico)

[47] http://www.census.gov/programs-surveys/acs/guidance.html
[48] https://www.census.gov/programs-surveys/acs/news/data-releases/2017/release.html
[49] https://www.census.gov/programs-surveys/geography.html

- ACS tallies, with Puerto Rico[50]: 220,333 with PR, 217,739 without Puerto Rico.
- Census 2010 tallies, by State/PR[51]: 220,334 with Puerto Rico, and 217,740 without Puerto Rico.

Table 4. Tallies of 2011-17 ACS Block Groups Used in 2019 Version of EJSCREEN

Geography	Number of Block Groups	In EJSCREEN?
Continental U.S., 48 States plus DC	216,330	Yes
Alaska & Hawaii	534 & 875	Yes
SUBTOTAL: Included in EJSCREEN	217,739	Yes
Puerto Rico	2,594	Yes
Virgin Islands and other Island Territories (American Samoa, Commonwealth of the Northern Mariana Islands, Guam)	408	No

Source: U.S. Census Bureau http://www.census.gov/geo/maps-data/data/tallies/tractblock.html

Demographic Variables and Formulas

This section provides details on the Census variables and formulas used to calculate demographic indicators for each block group. Short variable names used in EJSCREEN internal calculations are shown below, preceded by tabular data showing the ACS summary file table number, sequence number, variable number, and name of the table or variable. For example, total population, referred to in EJSCREEN calculations as "pop" and called "ACSTOTPOP" in the geodatabase, is taken from the Census variable B01001.001, in ACS 5-year summary file Table B01001. Field names as used in the EJSCREEN geodatabase differ from the names shown below, and a table was used to map between alternative fieldnames.

These details are based on the ACS 2013-2017 summary files.

ACS Summary File documentation is here: http://www2.census.gov/programs-surveys/acs/summary_file/2017/data/

[50] https://www.census.gov/programs-surveys/geography.html
[51] https://www.census.gov/programs-surveys/geography.html

ACS Summary File Variables

Tables used from ACS 2013-2017 are shown below. Block group data were obtained from the Census FTP site[52], and are not available from American Fact Finder at block group resolution.

Note that only selected variables from these tables are in the EJSCREEN geodatabase, for performance reasons. Many of the intermediate variables or detailed breakdowns are not in the geodatabase. Documentation supplied with the geodatabase download explains the variable names used in the geodatabase.

The URLs that can be used to view and download one ACS 2013-2017 table at a time, for the US total only, are as follows:

Table 5. ACS Tables Underlying EJSCREEN Demographic Data and Lead Paint Indicator

ACS Table ID	URL for US summary table via American Fact Finder	Table Title
B01001	https://factfinder.census.gov/bkmk/table/1.0/en/ACS/17_5YR/B01001/0100000US	SEX BY AGE
B03002	https://factfinder.census.gov/bkmk/table/1.0/en/ACS/17_5YR/B03002/0100000US	HISPANIC OR LATINO ORIGIN BY RACE
B15002	https://factfinder.census.gov/bkmk/table/1.0/en/ACS/17_5YR/B15002/0100000US	SEX BY EDUCATIONAL ATTAINMENT FOR THE POPULATION 25 YEARS AND OVER
C16002	https://factfinder.census.gov/bkmk/table/1.0/en/ACS/17_5YR/C16002/0100000US	HOUSEHOLD LANGUAGE BY HOUSEHOLDS IN WHICH NO ONE 14 AND OVER SPEAKS ENGLISH ONLY OR SPEAKS A LANGUAGE OTHER THAN ENGLISH AT HOME AND SPEAKS ENGLISH "VERY WELL"
C17002	https://factfinder.census.gov/bkmk/table/1.0/en/ACS/17_5YR/C17002/0100000US	RATIO OF INCOME TO POVERTY LEVEL IN THE PAST 12 MONTHS
B25034	https://factfinder.census.gov/bkmk/table/1.0/en/ACS/17_5YR/B25034/0100000US	YEAR STRUCTURE BUILT

URLs to view and download one ACS 2013-2017 table at a time, for the totals for the US, every state plus DC, and Puerto Rico are shown below.

[52] http://www.census.gov/programs-surveys/acs/technical-documentation/pums.html

B01001

https://factfinder.census.gov/bkmk/table/1.0/en/ACS/17_5YR/B01001/0100000US|0400000US01|0400000US02|0400000US04|0400000US05|0400000US06|0400000US08|0400000US09|0400000US10|0400000US11|0400000US12|0400000US13|0400000US15|0400000US16|0400000US17|0400000US18|0400000US19|0400000US20|0400000US21|0400000US22|0400000US23|0400000US24|0400000US25|0400000US26|0400000US27|0400000US28|0400000US29|0400000US30|0400000US31|0400000US32|0400000US33|0400000US34|0400000US35|0400000US36|0400000US37|0400000US38|0400000US39|0400000US40|0400000US41|0400000US42|0400000US44|0400000US45|0400000US46|0400000US47|0400000US48|0400000US49|0400000US50|0400000US51|0400000US53|0400000US54|0400000US55|0400000US56|0400000US72

B03002

https://factfinder.census.gov/bkmk/table/1.0/en/ACS/17_5YR/B03002/0100000US|0400000US01|0400000US02|0400000US04|0400000US05|0400000US06|0400000US08|0400000US09|0400000US10|0400000US11|0400000US12|0400000US13|0400000US15|0400000US16|0400000US17|0400000US18|0400000US19|0400000US20|0400000US21|0400000US22|0400000US23|0400000US24|0400000US25|0400000US26|0400000US27|0400000US28|0400000US29|0400000US30|0400000US31|0400000US32|0400000US33|0400000US34|0400000US35|0400000US36|0400000US37|0400000US38|0400000US39|0400000US40|0400000US41|0400000US42|0400000US44|0400000US45|0400000US46|0400000US47|0400000US48|0400000US49|0400000US50|0400000US51|0400000US53|0400000US54|0400000US55|0400000US56|0400000US72

B15002

https://factfinder.census.gov/bkmk/table/1.0/en/ACS/17_5YR/B15002/0100000US|0400000US01|0400000US02|0400000US04|0400000US05|0400000US06|0400000US08|0400000US09|0400000US10|0400000US11|0400000US12|0400000US13|0400000US15|0400000US16|0400000US17|0400000US18|0400000US19|0400000US20|0400000US21|0400000US22|0400000US23|0400000US24|0400000US25|0400000US26|0400000US27|0400000US28|0400000US29|0400000US30|0400000US31|0400000US32|0400000US33|0400000US34|0400000US35|0400000US36|0400000US37|0400000US38|0400000US39|0400000US40|0400000US41|0400000US42|0400000US44|0400000US45|0400000US46|0400000US47|0400000US48|0400000US49|0400000US50|0400000US51|0400000US53|0400000US54|0400000US55|0400000US56|0400000US72

C16002

https://factfinder.census.gov/bkmk/table/1.0/en/ACS/17_5YR/C16002/0100000US|0400000US01|0400000US02|0400000US04|0400000US05|0400000US06|0400000US08|0400000US09|0400000US10|0400000US11|0400000US12|0400000US13|0400000US15|0400000US16|0400000US17|0400000US18|0400000US19|0400000US20|0400000US21|0400000US22|0400000US23|0400000US24|0400000US25|0400000US26|0400000US27|0400000US28|0400000US29|0400000US30|0400000US31|0400000US32|0400000US33|0400000US34|0400000US35|0400000US36|0400000US37|0400000US38|0400000US39|0400000US40|0400000US41|0400000US42|0400000US44|0400000US45|0400000US46|0400000US47|0400000US48|0400000US49|0400000US50|0400000US51|0400000US53|0400000US54|0400000US55|0400000US56|0400000US72

C17002

https://factfinder.census.gov/bkmk/table/1.0/en/ACS/17_5YR/C17002/0100000US|0400000US01|0400000US02|0400000US04|0400000US05|0400000US06|0400000US08|0400000US09|0400000US10|0400000US11|0400000US12|0400000US13|0400000US15|0400000US16|0400000US17|0400000US18|0400000US19|0400000US20|0400000US21|0400000US22|0400000US23|0400000US24|0400000US25|0400000US26|0400000US27|0400000US28|0400000US29|0400000US30|0400000US31|0400000US32|0400000US33|0400000US34|0400000US35|0400000US36|0400000US37|0400000US38|0400000US39|0400000US40|0400000US41|0400000US42|0400000US44|0400000US45|0400000US46|0400000US47|0400000US48|0400000US49|0400000US50|0400000US51|0400000US53|0400000US54|0400000US55|0400000US56|0400000US72

B25034

https://factfinder.census.gov/bkmk/table/1.0/en/ACS/17_5YR/B25034/0100000US|0400000US01|0400000US02|0400000US04|0400000US05|0400000US06|0400000US08|0400000US09|0400000US10|0400000US11|0400000US12|0400000US13|0400000US15|0400000US16|0400000US17|0400000US18|0400000US19|0400000US20|0400000US21|0400000US22|0400000US23|0400000US24|0400000US25|0400000US26|0400000US27|0400000US28|0400000US29|0400000US30|0400000US31|0400000US32|0400000US33|0400000US34|0400000US35|0400000US36|0400000US37|0400000US38|0400000US39|0400000US40|0400000US41|0400000US42|0400000US44|0400000US45|0400000US46|0400000US47|0400000US48|0400000US49|0400000US50|0400000US51|0400000US53|0400000US54|0400000US55|0400000US56|0400000US72

TOTAL POPULATION COUNTS AND AGES

SEX BY AGE

Table.ID	Sequence.Number	Line.Number	Field
B01001	0002	NA	
B01001	0002	NA	Universe: Total population
B01001	0002	1	Total:
B01001	0002	2	Male:
B01001	0002	3	Under 5 years
B01001	0002	4	5 to 9 years
B01001	0002	5	10 to 14 years
B01001	0002	6	15 to 17 years
B01001	0002	7	18 and 19 years
B01001	0002	8	20 years
B01001	0002	9	21 years
B01001	0002	10	22 to 24 years
B01001	0002	11	25 to 29 years
B01001	0002	12	30 to 34 years
B01001	0002	13	35 to 39 years
B01001	0002	14	40 to 44 years
B01001	0002	15	45 to 49 years
B01001	0002	16	50 to 54 years
B01001	0002	17	55 to 59 years
B01001	0002	18	60 and 61 years
B01001	0002	19	62 to 64 years
B01001	0002	20	65 and 66 years
B01001	0002	21	67 to 69 years
B01001	0002	22	70 to 74 years
B01001	0002	23	75 to 79 years
B01001	0002	24	80 to 84 years
B01001	0002	25	85 years and over
B01001	0002	26	Female:
B01001	0002	27	Under 5 years
B01001	0002	28	5 to 9 years
B01001	0002	29	10 to 14 years
B01001	0002	30	15 to 17 years
B01001	0002	31	18 and 19 years
B01001	0002	32	20 years
B01001	0002	33	21 years
B01001	0002	34	22 to 24 years
B01001	0002	35	25 to 29 years
B01001	0002	36	30 to 34 years
B01001	0002	37	35 to 39 years
B01001	0002	38	40 to 44 years

```
B01001          0002        39           45 to 49 years
B01001          0002        40           50 to 54 years
B01001          0002        41           55 to 59 years
B01001          0002        42          60 and 61 years
B01001          0002        43           62 to 64 years
B01001          0002        44          65 and 66 years
B01001          0002        45           67 to 69 years
B01001          0002        46           70 to 74 years
B01001          0002        47           75 to 79 years
B01001          0002        48           80 to 84 years
B01001          0002        49        85 years and over
```

pop = B01001.001
ageunder5m = B01001.003
age5to9m = B01001.004
age10to14m = B01001.005
age15to17m = B01001.006
age65to66m = B01001.020
age6769m = B01001.021
age7074m = B01001.022
age7579m = B01001.023
age8084m = B01001.024
age85upm = B01001.025
ageunder5f = B01001.027
age5to9f = B01001.028
age10to14f = B01001.029
age15to17f = B01001.030
age65to66f = B01001.044
age6769f = B01001.045
age7074f = B01001.046
age7579f = B01001.047
age8084f = B01001.048
age85upf = B01001.049

RACE/ETHNICITY

```
HISPANIC OR LATINO ORIGIN BY RACE

Table.ID Sequence.Number Line.Number             Field
B03002            0005    NA

B03002            0005    NA     Universe:  Total population
B03002            0005    1      Total:
B03002            0005    2      Not Hispanic or Latino:
B03002            0005    3      White alone
B03002            0005    4      Black or African American alone
B03002            0005    5      American Indian and Alaska Native alone
B03002            0005    6      Asian alone
B03002            0005    7      Native Hawaiian and Other Pacific Islander alone
B03002            0005    8      Some other race alone
B03002            0005    9      Two or more races:
B03002            0005    10     Two races including Some other race
B03002            0005    11     Two races excluding Some other race, and three or more
races
B03002            0005    12     Hispanic or Latino:
B03002            0005    13     White alone
B03002            0005    14     Black or African American alone
B03002            0005    15     American Indian and Alaska Native alone
B03002            0005    16     Asian alone
B03002            0005    17     Native Hawaiian and Other Pacific Islander alone
B03002            0005    18     Some other race alone
B03002            0005    19     Two or more races:
B03002            0005    20     Two races including Some other race
B03002            0005    21     Two races excluding Some other race, and three or more
races
```

pop3002 = B03002.001
nhwa = B03002.003

EDUCATIONAL ATTAINMENT FOR THOSE AGE 25+

```
SEX BY EDUCATIONAL ATTAINMENT FOR THE POPULATION 25 YEARS AND OVER

Table.ID Sequence.Number Line.Number   Field
B15002            0043    NA

B15002            0043    NA     Universe:  Population 25 years and over
B15002            0043    1      Total:
```

B15002	0043	2	Male:
B15002	0043	3	No schooling completed
B15002	0043	4	Nursery to 4th grade
B15002	0043	5	5th and 6th grade
B15002	0043	6	7th and 8th grade
B15002	0043	7	9th grade
B15002	0043	8	10th grade
B15002	0043	9	11th grade
B15002	0043	10	12th grade, no diploma
B15002	0043	11	High school graduate, GED, or alternative
B15002	0043	12	Some college, less than 1 year
B15002	0043	13	Some college, 1 or more years, no degree
B15002	0043	14	Associate's degree
B15002	0043	15	Bachelor's degree
B15002	0043	16	Master's degree
B15002	0043	17	Professional school degree
B15002	0043	18	Doctorate degree
B15002	0043	19	Female:
B15002	0043	20	No schooling completed
B15002	0043	21	Nursery to 4th grade
B15002	0043	22	5th and 6th grade
B15002	0043	23	7th and 8th grade
B15002	0043	24	9th grade
B15002	0043	25	10th grade
B15002	0043	26	11th grade
B15002	0043	27	12th grade, no diploma
B15002	0043	28	High school graduate, GED, or alternative
B15002	0043	29	Some college, less than 1 year
B15002	0043	30	Some college, 1 or more years, no degree
B15002	0043	31	Associate's degree
B15002	0043	32	Bachelor's degree
B15002	0043	33	Master's degree
B15002	0043	34	Professional school degree
B15002	0043	35	Doctorate degree

age25up = B15002.001
m0 = B15002.003 (males age 25+ with zero education)
m4 = B15002.004 (males age 25+ with >0 up to 4th grade)
m6 = B15002.005
m8 = B15002.006
m9 = B15002.007
m10 = B15002.008
m11 = B15002.009
m12 = B15002.010 (males age 25+ with high school diploma)

f0 = B15002.020
f4 = B15002.021
f6 = B15002.022
f8 = B15002.023
f9 = B15002.024
f10 = B15002.025
f11 = B15002.026
f12 = B15002.027

HOUSEHOLDS THAT ARE LINGUISTICALLY ISOLATED

```
"HOUSEHOLD LANGUAGE BY HOUSEHOLDS IN WHICH NO ONE 14 AND OVER SPEAKS ENGLISH ONLY OR
SPEAKS A LANGUAGE OTHER THAN ENGLISH AT HOME AND SPEAKS ENGLISH "VERY WELL"

Table.ID Seq.Number Line.Number Field
C16002       0044   NA

C16002       0044   NA     "Universe:  Households"
C16002       0044   1      "Total:"

C16002       0044   2      "English only"
C16002       0044   3      "Spanish:"
C16002       0044   4
No one 14 and over speaks English only or speaks English “very well”
C16002       0044   5
At least one person 14 and over speaks English only or speaks English “very well”

C16002       0044   6      Other Indo-European languages:
C16002       0044   7
No one 14 and over speaks English only or speaks English “very well”
C16002       0044   8
At least one person 14 and over speaks English only or speaks English “very well”

C16002       0044   9      Asian and Pacific Island languages:
C16002       0044   10
No one 14 and over speaks English only or speaks English “very well”
C16002       0044   11
At least one person 14 and over speaks English only or speaks English “very well”

C16002       0044   12     Other languages:
C16002       0044   13
No one 14 and over speaks English only or speaks English “very well”
C16002       0044   14
At least one person 14 and over speaks English only or speaks English “very well”
```

hhlds = C16002.001
lingisospanish = C16002.004
lingisoeuro = C16002.007
lingisoasian = C16002.010
lingisoother = C16002.013

INDIVIDUALS BY RATIO OF INCOME TO POVERTY THRESHOLD

```
RATIO OF INCOME TO POVERTY LEVEL IN THE PAST 12 MONTHS

Table.ID Sequence.Number Line.Number                Field
C17002            0049    NA

C17002            0049    NA Universe:  Population for whom poverty status is determined
C17002            0049    1                                                       Total:
C17002            0049    2                                                   Under .50
C17002            0049    3                                                  .50 to .99
C17002            0049    4                                                1.00 to 1.24
C17002            0049    5                                                1.25 to 1.49
C17002            0049    6                                                1.50 to 1.84
C17002            0049    7                                                1.85 to 1.99
C17002            0049    8                                               2.00 and over
```

povknownratio = C17002.001
pov50 = C17002.002 (below 0.50 times poverty threshold)
pov99 = C17002.003 (0.5 to 0.99 times poverty threshold)
pov124 = C17002.004
pov149 = C17002.005
pov184 = C17002.006
pov199 = C17002.007
pov2plus = C17002.008

AGE OF OCCUPIED HOUSING UNITS (CORRELATED WITH LEAD PAINT)

```
YEAR STRUCTURE BUILT

Table.ID Sequence.Number Line.Number              Field
B25034            0104           NA
```

```
B25034            0104          NA Universe:  Housing units

B25034       0104               1            Total:
B25034       0104               2            Built 2014 or later
B25034       0104               3            Built 2010 to 2013
B25034       0104               4            Built 2000 to 2009
B25034       0104               5            Built 1990 to 1999
B25034       0104               6            Built 1980 to 1989
B25034       0104               7            Built 1970 to 1979
B25034       0104               8            Built 1960 to 1969
B25034       0104               9            Built 1950 to 1959
B25034       0104               10           Built 1940 to 1949
B25034       0104               11           Built 1939 or earlier
```

builtunits = B25034.001
built1950to1959 = B25034.009
built1940to1949 = B25034.010
builtpre1940 = B25034.011

Calculated Demographic Data Fields

Based on the raw counts from the ACS described above, various demographic variables were calculated for use in EJSCREEN. Conditional formulas below are in R syntax, and generally indicate that a value of zero was used in cases where the denominator was zero, to avoid division by zero. For example, the formula "pctmin = ifelse(pop==0,0, as.numeric(mins) / pop)" indicates that percent minority was calculated as the ratio of number of minorities over total population of a block group, but was set to zero if the population was zero.

 # RACE/ETHNICITY COMBINED, CALCULATED VARIABLES

mins = pop - nhwa
pctmin = ifelse(pop==0,0, as.numeric(mins) / pop)

 # POVERTY, LOW-INCOME CALCULATED VARIABLES

# poverty ratios
num2pov = num1pov + pov124 + pov149 + pov184 + pov199
lowinc = num2pov
pct2pov = ifelse(povknownratio==0,0, num2pov/povknownratio)
pctlowinc = pct2pov
num2pov.alt = povknownratio - pov2plus
pct2pov.alt = ifelse(povknownratio==0,0, num2pov.alt/povknownratio)

 # EDUCATIONAL ATTAINMENT CALCULATED VARIABLES

```
lths = m0 + m4 + m6 + m8 + m9 + m10 + m11 + m12 +
                    f0 + f4 + f6 + f8 + f9 + f10 + f11 + f12
pctlths = ifelse(age25up==0,0, as.numeric(lths ) / age25up)
```

 # LINGUISTIC ISOLATION CALCULATED VARIABLES

```
lingiso = lingisospanish + lingisoeuro + lingisoasian + lingisoother
pctlingiso = ifelse( hhlds==0,0, lingiso / hhlds)
```

 # AGE GROUPS CALCULATED VARIABLES

```
under5 = ageunder5m + ageunder5f
pctunder5 = ifelse( pop==0,0, under5/pop)
over64 = age65to66m + age6769m + age7074m + age7579m + age8084m + age85upm +
                    age65to66f + age6769f + age7074f + age7579f + age8084f + age85upf
pctover64 = ifelse( pop==0,0, over64/pop)
```

 # HOUSING CALCULATED VARIABLES (LEAD PAINT INDICATOR)

```
pre1960 = builtpre1940 + built1940to1949 + built1950to1959
pctpre1960 = ifelse( builtunits==0,0, pre1960/builtunits)
```

Uncertainty and Limitations in the Demographic Data

Uncertainty in demographic data

As with every sample survey, sampling results in unavoidable approximations in every estimate that comes from the survey. The Census Bureau clearly labels every data item as an "estimate," and accompanies each with an estimate of its margin of error. Anyone using a screening tool should be aware of those demographic uncertainties, together with uncertainties in the environmental measures, in tables, graphical displays and descriptive materials.

Uncertainties are also discussed in section 1 (as general caveats), section 2 (with regard to buffer reports), and Appendix B (in discussions of buffering details and demographic data).

Users of EJSCREEN must keep in mind the substantial uncertainty in estimated demographic and environmental indicators used in screening tools such as EJSCREEN. Uncertainty is a critical consideration when using EJSCREEN because the tool relies on demographic and environmental estimates at block group resolution. As the Census Bureau makes clear in documentation of the American Community Survey (ACS), the margin of error for an estimate in a given block group is often very large relative to the estimate, so an estimate of percent low-income, for example, is often very uncertain for a single block group.

Combined with uncertainty in environmental data, this means EJ index values are often very uncertain at block group resolution. Therefore, modest differences in percentile scores between block groups or small buffers should not be interpreted as meaningful because of the uncertainties in demographic and environmental data at the block group level. We do not have a high degree of confidence when comparing or ranking places with only modest differences in estimated percentile. For this reason, it is critical that EJSCREEN results be interpreted carefully and that additional information be used to supplement or follow up on screening, where appropriate. Section 1 of this document discusses caveats and limitations further.

EPA cannot provide precise confidence intervals on EJ indexes or percentiles due to technical limitations in the data made public by the Census Bureau and the challenges of quantifying uncertainty for the environmental indicators. Technical documentation on methods and challenges in estimating uncertainty for calculated demographic indicators using the ACS is available from the Census Bureau[53] (with challenges described in related technical documents[54]). ESRI also provides useful discussions of margin of error.[55]

It is likely that block group errors in the various data fields reported by Census (e.g., count with income-poverty ratio below 0.5, count with ratio 0.5 to 1, etc.) are correlated. Relevant covariances, however, are not provided by the Census Bureau. This means simple methods of approximating margin of error for a calculated variable (e.g., percent low-income) may not be entirely adequate. In this case, it appears that a custom tabulation by the Census Bureau would be the most accurate way to generate reliable estimates of the margin of error for

---

[53] https://www.census.gov/programs-surveys/acs/data/race-aian.html
[54] https://www2.census.gov/programs-surveys/acs/tech_docs/accuracy/MultiyearACSAccuracyofData2015.pdf
[55] http://www.esri.com/software/american-community-survey/understanding-margin-error

variables such as the percent low-income or the demographic indicator, for use in creating confidence intervals around an EJ index or the percentile of that index. Future research may be able to produce reasonable approximations of confidence intervals around block group or buffer estimates. EJSCREEN users should keep in mind that using a buffer larger than the local block groups will produce more reliable estimates than a single block group can provide.

Using 2x poverty rate

The rationale for using twice the poverty threshold rather than just the poverty threshold includes the following considerations:

- The effects of income on baseline health and probably on other aspects of susceptibility are not limited to those below the poverty thresholds — those from 1x to 2x poverty also have worse health overall than those with higher incomes (Centers for Disease Control and Prevention, 2010), and asthma rates, for example, begin to increase as income falls below twice the poverty threshold (Centers for Disease Control and Prevention, 2011a).
- Many studies in various fields use 2x poverty, and many others use 1x poverty (e.g., see Su et al., 2009); the same is true for prior EPA screening tools. There is precedent for both. However, a rationale often mentioned is that today's poverty thresholds are too low to adequately capture the populations adversely affected by low income levels, especially in high-cost areas. Some analysts have concluded that the amount of income actually required for basic living costs without government support is far higher than the current Federal poverty thresholds (Cauthen & Fass, 2008).
- When using twice the poverty threshold, the number or percent low income happens to roughly equal number or percent minority in the United States. This makes it convenient and simple to use the average of the two without applying any other weights to them, and in this way each low-income person affects the susceptibility indicator about as much as each minority person.
- The Census Bureau has been developing experimental poverty measures that account for local costs of living, but these are not yet in widespread use.[56]

Interpretation of Demographic Indexes

The demographic indexes are meant to reflect some of the combined impacts of multiple demographic factors. The Census Bureau does not provide a tabulation of low-income residents by race/ethnicity at the block group level[57], so it is impossible to know what percentage of a block group is low-income minorities vs. low-income non-minorities, for example. EJSCREEN simply defines the demographic index as the average of the percentage of people who are low income and the percentage of people who are minorities. Therefore, this demographic index will be (equal to or) *smaller than* the percentage of *people who are in at least one of these groups*. In

[56] https://www.census.gov/topics/income-poverty/poverty/guidance/poverty-measures.html
[57] Table B17001 and related tables provide tract-level cross-tabulations of race-ethnicity and poverty, but not percent low-income as defined in EJSCREEN.

other words, it is typically smaller than the share of people who are in one or more of these groups – just low income, just minority, or both. The average will also be (equal to or) larger than the percentage of people who are simultaneously in all of these groups. It is larger than (or equal to) the share of *people who are simultaneously low income and minority*. The value of the demographic index is almost always *larger than* the number of people who are simultaneously minority and low-income, because usually some people are in only one of these demographic groups. Note that one person cannot be under five and over 64, so any one person can be in up to five of the six demographic groups used in EJSCREEN.

The demographic index is also bounded by these two percentages (percent low income and percent minority). For example, the actual percentage minority is larger than the value of the demographic index, if the percentage low-income is lower than the percentage minority.

APPENDIX C. TECHNICAL DETAILS ON PROXIMITY INDICATORS

Several of EJSCREEN's environmental indicators are direct or indirect estimates of potential exposure or health risks, such as the NATA cancer risk estimates and the ozone and $PM_{2.5}$ concentration estimates. There are other aspects of an individual's or a community's environmental concerns that are less readily quantified in terms of emissions, concentrations, or risk estimates.

People may be concerned about living near facilities that handle hazardous substances, and other potential sources of pollution, such as highways or abandoned waste sites. Concern over "locally undesirable land uses", or LULUs, is in some cases founded on the potential for routine or episodic releases of pollutants to the air, land or water, and the potential for such releases to cause human health or environmental adverse effects or other societal disamenities.

The purpose of the proximity measures in EJSCREEN is to systematically and consistently quantify different degrees of potential for these effects. We have developed a method to calculate a score that represents the relative magnitude of the proximity of the population within a block group to facilities, waste sites, or traffic surrounding it. A block group with more facilities closer to the block group's residential population will have a higher score than a block group where facilities are further away. We have applied this method to these facility or site types:

- National Priorities List (NPL) sites (a key subset of "Superfund," sites).
- Hazardous waste Treatment, Storage or Disposal Facilities (TSDFs), subject to regulations under the Resource Conservation and Recovery Act (RCRA).
- Risk Management Plan (RMP) facilities, which are facilities that maintain greater than certain quantities of extremely hazardous substances, and are required to take certain actions, including filing risk management plans, under section 112 (r) of the Clean Air Act.

We have developed a similar approach to represent proximity to and traffic volume on nearby highways and proximity to toxic concentrations on nearby water segments.

In the sections below, we will describe the general approach, in terms of facility proximity. We will then describe how it differs for traffic and wastewater proximity. Then we will discuss certain adjustments we have made, mostly to make the approach computationally efficient, and summarize the data sources and computational routine that we applied to implement this approach. We conclude with caveats and other observations.

Calculating Proximity to Facilities or NPL Sites

Each of the more than 217,000 block groups for the U.S. states and the District of Columbia is made up of between one block and several hundred blocks. Most block groups nationwide are smaller than approximately 0.5 square miles, an area that if circular would have a radius of about 640 meters. In block groups of this median size, the average residence generally would be about 430 to 720 meters (or less than half a mile) away from a given point within the block group, such as a facility, as explained at the end of Appendix C. About 20-25% of block groups covered an area smaller than a circle of radius 300-350 meters (almost one quarter of a mile), as of the 2005-2009 geographies. Also, a very small number of block groups are extremely large in area, in very rural locations.

All of a block group's blocks may have residential population estimated by the 5-year ACS, or only some, and some block groups have no residents at all. Blocks and block groups vary greatly in geographic area, and in population. The approach used here works first at the block level, based on measures of proximity to the facilities in or near the blocks. The block-level measures are then aggregated among all the blocks within a block group, weighted by the number of people in the different blocks.

Thus, while population is considered in aggregating the block scores, the measure does not increase or decrease for block groups with higher or lower populations. The measure is, rather, a characteristic of the residents of the block group, in the same way that cancer risk from NATA or ozone concentration are estimated measures of the conditions of those places.

Let
i represent a particular facility
j represent a block within a block group
k represent a block group

d_{ij} is the distance, in kilometers, from block j's centroid to the given location of facility i

pop_{jk} is the estimated population of block j within block group k

pop_k is the total estimated population of block group k

$f(d_{ij})$ is a function representing the proximity of facilty i to block j, a declining function of the distance, d_{ij}

$BlockScore_{jk}$ is the aggregation of the proximity influences of all facilities affecting block jk

$BlockGroupScore_k$ is the population-weighted aggregation of the block group's component blocks

We have chosen to define the proximity function as

$$f(d_{ij}) = 1 / d_{ij}$$

That is, a facility 1 kilometer from a block's population contributes twice the score as a facility 2 kilometers from the same block. We note that we have made a choice in using inverse distance for this function. Air dispersion modeling for pollutants following Gaussian plume assumptions would show a generally greater drop-off in concentration, roughly with the second power to 2.5 power of one over distance. But actual concentrations around individual plants follow often-complex patterns that depend on the particular mix of stack vs. fugitive emissions, characteristics of stack height, exit velocity and temperature, the presence of buildings or other land surface characteristics and meteorology. Some substances react readily with other substances in the atmosphere, or precipitate out readily. It is not uncommon for concentrations to rise for some distance from the emitting source, and then to fall from that peak concentration. The Gaussian plume model applies to gases, and emissions of particulates can drop off more quickly than gases.

Releases to land may follow extremely complex patterns of dispersion. Added to that are the very site-specific characteristics of potential human exposure via drinking water, vapor intrusion or contact with contaminated soils, etc. For water pollution, similar complexities exist, most notably that an effluent is carried away downstream of a running body of water, dilution can be complicated by the presence of other water entering stream segments, by volatilization, by biological and chemical interactions, and by deposition to sediments, and finally by the treatment and removal of a water pollutant sent to a publicly-owned treatment works.

We also note that researchers and others have taken varied approaches to representing the proximity of facilities to populations. The EJSM model of environmental justice concerns, developed for the state of California, scored facility proximity in concentric rings around a population centroid (Pastor Jr., Morello-Frosch, & Sadd, 2010; Sadd, Pastor, Morello-Frosch, Scoggins, & Jesdale, 2011). All facilities within 1 mile received a score of 3. All within the 1 to 3 mile band received a score of 2, and those between 3 and 5 miles received a score of 1. Anything beyond 5 miles received a score of zero. This step-wise scoring represents the judgment of the model developers, influenced by interactions with various stakeholders.

Finally, we note that EJSCREEN's measure of proximity is intended to represent more than simply real or potential human health adverse effects coming from exposure. Some parts of the environmental justice literature reflect semi-quantitative factors, such as increased psychological stress, fear and other reactions to the presence of LULUs. This is not the forum for sorting through those factors.

However, we have made a judgment call: For the purposes of this EJSCREEN tool, we represent a facility's measure of proximity by the inverse of its distance from the estimated location of the average person. A block's proximity score is the sum of the inverse distances of all the facilities of a particular type.

Note that for the minority of block groups in the United States with no residential population, we take a straight average of the block scores.

The units for these measures are facilities per kilometer. A block group could have a score of 1.0 if all residents were an average of one kilometer from a single facility, and all other facilities were so distant (> 5 km) as to make no contribution to the score. Another block group could have a score of 1.0 if there were five facilities that were all exactly five kilometers from the residents.

Calculating Proximity to Traffic

We have adopted essentially the same approach described above for representing proximity to highway segments – an inverse distance-weighted sum of highway segments surrounding each block, and a population-weighted sum of the individual blocks' contributions to the block group.

The highway segment database that we have used is described in section 3. These segments differ from a facility database in that they are lines on a geographic area, rather than points that represent the facilities. In our approach, we find the distance from the block centroids to the nearest part of each surrounding highway segment. The nearest point, d_{ij}, could be an end of the highway segment or some point in between the ends.

We also multiplied each d_{ij} by the annual average daily traffic estimate that is associated with each highway segment. This is meant to reflect the traffic intensity, and this differs from the facility approach, where we have taken each facility within each group as having equal importance. Also, for traffic proximity, the search radius is 500 meters and the score uses distance in meters, not kilometers.

Calculating Proximity to Toxic Weighted Wastewater Dischargers

The proximity to wastewater dischargers was calculated using the same methodology as the Proximity to Traffic—an inverse distance-weighted sum of water reach segments surrounding each block, and a population-weighted sum of the individual blocks' contributions to the block group. As described above in the Details on Environmental Indicators section, the source data is RSEI modeled output of toxics concentrations mapped to water segments. So essentially EPA replaced highways with National Hydrography Dataset (NHD) reach water segments, and average annual daily traffic counts with toxic concentrations.

The Water Discharger map layer (as is it currently referred to in EJSCREEN) was adjusted to display all NPDES facility latitude/longitude with pollutant loadings greater than zero. In addition, EPA provided more information in the popup box associated with each mapped facility, including a link to the detailed facility information in the DMR Pollutant Loading Tool, so that users can query the specific pollutants the facility may be discharging.

All relevant stream parameters were already included in the RSEI modeling context and was upgraded to NHD Version 2.0, which contained improved flow estimates and better connectivity.

Currently the DMR Loading Tool includes 672 reported chemicals from DMRs, of which 201 are also reported to TRI. The ideal set of loadings would be the union of the DMR and TRI data sets. For the overlap, that is, facility/chemical combinations reported to both DMR and TRI, DMR reported quantities were prioritized, as they are not subject to some of the constraints of TRI reporting and are usually more accurate, being based on actual monitoring data.

The concentrations of the modeled chemicals were weighted by their relevant toxicity weight. Only toxicity weights from RSEI were used. EPA excluded chemicals without RSEI toxicity weights or for which cannot be readily extrapolated for an existing chemical. RSEI toxicity weights are based solely on chronic human health effects, and therefore, may be more suitable for an application like EJSCREEN that is geared toward screening for human health concerns. For water releases, the RSEI toxicity weight is calculated as the reciprocal of the reference dose (RfD) for noncarcinogens or cancer potency factor/$1x10^{-6}$ for carcinogens. Crosswalks between chemical sets (DMR and TRI) allowed EPA to apply toxicity weights and decay rates of TRI chemicals to matching DMR chemicals, so that DMR data could be incorporated into the RSEI model.

The new indicator approach ran the standard RSEI modeling data processing procedures. EPA extracted DMR and TRI data from the DMR Pollutant Loading Tool and structured the data to closely mimic the standard RSEI input tables. EPA aimed to minimize the adjustments needed to accommodate DMR data in the RSEI modeling procedures.

The new water indicator calculates proximity as the distance between the centroid of the Census block to the midpoint of any stream reaches within a 500 m radius of the block centroid. If no reaches are found within 500m, the nearest neighbor within 3 km is used. The indicator uses the output from RSEI, to give more emphasis to stream reaches with higher toxicity-weighted pollutant concentrations.

$$f(d_{ij}, W) = \frac{W}{d_{ij}}$$

Where;

W is the weighting factor (described below) and $W > 0$;

d_{ij} is the distance, in km, from the Census block centroid, *j*, to the midpoint of the stream reach, *i*; and,

$f(d_{ij}, W)$ is the weighted proximity function.

The resulting block values were aggregated to Census block groups using a population-weighted method to produce block group level results. This is the same aggregation method used for the other EJSCREEN proximity indicators.

Calculating Proximity – Additional Details

We address two modifications to the general method described above. The first deals with instances where a facility or highway segment location is very close to the centroid of the block. The second is an accommodation to the computational intensity of the general method.

Extremely Small d_{ij} Values

Our intention is to represent the proximity of facilities or highway segments to the population within each block. All facilities and each part of all highway segments fall within one block. By chance, some portion of those points fall very close to the block centroids.

We do not know how the population is geographically distributed within any block, but we assume that people are more likely to be distributed across the blocks' expanses than to be concentrated at one point, such as the centroid. In fact, for rural, suburban and many non-high rise urban areas, people's residences are more likely to be closer to the blocks' peripheries (bounded by roads) than clustered at the centroids. Thus, when a facility location happens to be very close to the block centroid, it would result in an artificially high contribution to the block's score. This is not a hypothetical problem: We have observed d_{ij} values well below 100 meters, and some below 10 meters.

In looking for solutions to the problem, we conducted analyses and arrived at the approach we have adopted. Blocks vary widely in their total area and in their shapes. Both can be found in the Census Bureau's Tiger shape files. Dealing explicitly with the individual block shapes would be computationally very intensive because there are over 11 million blocks. Since we cannot easily find out how the residents are actually distributed in those areas, we made two simplifying assumptions:

- residents are evenly distributed across the surface area of each block, and
- each block can be represented by a circle whose radius is $[\text{Block area} / \text{Pi}]^{1/2}$.

We call this latter value the Block Area Equivalent Radius.

Our investigations indicate that for any d_{ij} less than the Block Area Equivalent Radius, 0.9 times that value is a reasonable representation of the average distance from the facility for all residents in the block. We call this the $d_{ij\ \text{corrected}}$.

Our computational scheme determines the d_{ij} values as described above, tests for the comparison with Block Area Equivalent Radius, and substitutes $d_{ij\ \text{corrected}}$ values. We found that we needed to make that correction for less than 1% of all facility / block combinations in an early testing data set that used 2005-2009 ACS data.

Accommodating to Computational Intensity – Combine a Distance Limit with a Nearest Facility Approach

Our task is to compute a proximity score for each of the facility or site types and highway segments for each of the more than 217,000 block groups, comprised of over 11 million blocks. The number of facilities nation-wide varies from hundreds of TSDFs to many thousands of RMP facilities. Computing all the combinations would require more computational time and resources than were available.

In addition, doing so would be wasteful and perhaps irrelevant. The one over distance function we have chosen to represent concerns about facilities and highways drops off greatly for most facilities beyond the nearest ones. The miniscule contribution of a facility 100 kilometers or more from a block is not only small, compared with those that may be within 5 to 10 kilometers, but has little common-sense meaning, in our view.

Consequently, we have followed the general approach described above only for facilities or sites within 5 kilometers of a block's centroid, and within 500 meters for highway segments. Depending on the facility or site type, we find that 30-40% of block groups have at least one facility (RMP or TSDF) within the 5 kilometers limit, and almost 10% have one or more NPL sites within 5 km, in the 2017 version of EJSCREEN.

Of course, every block and block group has one nearest facility, even though it may be beyond the 5 kilometers horizon, and some of those may be fairly close to that limit. We have also calculated the distance to the facility nearest to each of the blocks. For those blocks lacking anything within the 5 kilometers, we represented the facility proximity by one over the distance to that single nearest facility.

This added computational complexity to the approach, but at far less cost than computing the full matrix of millions of blocks times thousands of facilities and sites.

This hybrid approach results in every block (and thus every block group) having a nonzero, positive proximity score. All of the resulting block proximity scores are necessarily less than the score had we computed the full matrix, but we judge that this is a reasonable and practical compromise. Figure 1 shows histograms for proximity scores. Counting only the single nearest beyond 5 km has the effect of shifting scores under 0.2 to the left, to lower scores than if all were counted, but the graphs show no major discontinuities, suggesting this limitation (counting only the nearest one) has little impact overall.

Data and Computational Scheme

Using the Census 2010 block centroids, the distance to all facilities within 5 kilometers of all blocks (not just block groups) was determined, and distance to the nearest facility at any distance was determined if none were found within 5 km.

The d_{ij} values were compared to the Block Area Equivalent Radius and corrected values were used when necessary, before computing $1 / d_{ij}$. The $1 / d_{ij}$ values were summed for each block to compute the $BlockScore_{jk}$. These were then rolled up to the block group level, applying the population weighting described above, for the final $BlockGroupScore_k$.

Caveats and Observations

Several aspects of the proximity analysis approach have been mentioned above, but deserve summary here.

- We recognize that our selection of the inverse of distance is a design choice that represents our judgment of a balance among competing factors.
- We recognize that one could potentially attempt to distinguish among facilities within each facility category by quantitative or qualitative measures of importance. These could include total pounds released or toxicity-weighted releases for NPDES facilities; the number of accidental releases and/or their apparent severity for RMP facilities; some classification of the likelihood of releases for NPL sites or TSDFs; and general indications of scale for all of them. We note that CalEnviroscreen has addressed this issue to some extent, and that the RSEI tool based on TRI data may be relevant to future work on this issue. At this point, we have chosen not to develop any such potential scaling adjustments.
- We recognize that all location data are subject to potential error. While we have high confidence in the block centroid locations, we know that the facility or site or roadway location data may contain larger or smaller errors, and that for large facilities or sites, one point may not be an entirely adequate representation of the location of its releases or of neighbors' perceptions.
- We recognize that the computational accommodation we describe above results in a hybrid of measures: For some block groups, all blocks have one or more facilities within 5 kilometers and the

score is the summation of all those potentially multiple facility/block combinations; for other block groups, none of the blocks have a facility within 5 kilometers and the score is the contribution of the single facility closest to each block; while for some block groups, we have a mix of those situations. We believe that this is a reasonable compromise

Note: The method described in the above section would be the case if homes and facilities were on average uniformly distributed within block groups that were roughly circular on average, because the average distance between two random points in a circle of radius R is 90% of R (Weisstein, Eric W. "Disk Line Picking." From *MathWorld*--A Wolfram Web Resource. http://mathworld.wolfram.com/DiskLinePicking.html). This means that if a population is randomly spread over a roughly circular block, a facility in the block typically would be 0.9R from the average person. Also, the average point in the circle is 0.67R from the center, and 1.13R from the edge of the circle. We can describe this relationship using an equation that is a portion of the formula for the distance between two random points in a circle of radius=1. The formula is

$$\int_0^1 \int_0^\pi \frac{1}{\pi} \sqrt{a + b^2 - \left(2\, b \sqrt{a}\,\right) \cos(t)}\ dt\, da$$

where b= the facility's distance from the center as a fraction of the radius, and the integral over *a* represents distances of residences from the center. We can solve this equation using http://WolframAlpha.com, for b=0, 0.5, or 1, representing points at the center, halfway to the edge, and at the edge of the circle. For example, we can use this equation for b=0.5 to find that the average person, if randomly located in a circle of radius R, is a distance of about 0.8 R from a facility that is halfway between the center and edge of the circle. For the distance between the average person and a randomly placed facility in the circle, we use b=sqrt(0.5) instead, and the following would be used as the input to WolframAlpha:
Integrate[(1/Pi) Sqrt[a + (Sqrt(0.5))^2 - 2 * (Sqrt(0.5)) * Sqrt[a] cos(t)], {a, 0, 1}, {t, 0, pi}] or
http://www.wolframalpha.com/input/?i=Integrate%5B%281%2FPi%29+Sqrt%5Ba+%2B+%28Sqrt%280.5%29%29%5E2+-+2+*+%28Sqrt%280.5%29%29+*+Sqrt%5Ba%5D+cos%28t%29%5D%2C+%7Ba%2C++0%2C++1%7D%2C++%7Bt%2C+0%2C+pi%7D%5D+

APPENDIX D. SUMMARY STATISTICS FOR INDICATORS

This appendix provides basic information and statistics on the environmental and demographic data used in EJSCREEN.

Table 6 shows summary statistics for the 11 environmental indicators, including the population mean and selected population percentiles. The mean score and high percentiles for each indicator provide useful perspective on the magnitude of these environmental indicators for the average or highly exposed individuals.

Table 5. Summary Statistics for Environmental Indicators

Environmental Indicator	Missing	Minimum	25%ile	50%ile (median)	*Pop. Mean*	75%ile	80%ile	90%ile	95%ile	99%ile
PM 2.5	1664	3.04	7.41	8.30	9.14	9.13	9.34	9.95	10.58	12.11
Ozone	1664	25.27	38.93	43.03	38.4	46.12	46.72	49.21	53.47	60.61
NATA DPM	542	4.8E-06	0.23	0.38	0.48	0.60	0.66	0.85	1.05	1.73
NATA cancer risk	542	6.13	25.72	31.16	31.92	36.85	38.21	41.72	44.53	49.55
NATA respiratory HI	542	0.06	0.34	0.43	0.44	0.52	0.55	0.62	0.68	0.77
% pre-1960 (lead paint)	0	0	.04	0.17	0.28	0.46	0.54	0.72	0.83	0.92
Proximity Traffic	0	-	50.15	203.95	753.67	654.58	849.77	1691.71	2862.75	6046.43
Proximity NPL	0	0	2.8E-02	6.0E-02	0.13	0.13	0.15	0.24	0.42	0.88
Proximity RMP	0	0	0.14	0.30	0.74	0.90	1.11	2.53	2.71	4.05
Proximity TSDF	0	0	0.14	0.42	3.95	1.64	2.17	4.31	7.45	45.46
Wastewater Discharge	534	-	-	2.5E-05	13.60	3.65E-03	8.91E-03	7.65E-02	0.51	6.56

Source: 2019 version of EJSCREEN, not including Puerto Rico. See body of report for sources and definitions of environmental indicators.

Notes: Population percentiles (and means) are shown, not block group percentiles (or means), so 80%ile means 80% of the population has a lower (or exactly tied) block group score. Values in table have been rounded to two significant digits, except for PM, ozone, and DPM, which use three significant digits. Numbers may differ slightly from those in EJSCREEN reports. Summary statistics for a given environmental factor exclude block groups where that environmental indicator was not available (missing).

Table 6. Summary Statistics for Demographics

Demographic variable	Missing	Minimum	25%ile	50%ile (median)	*Population mean*	75%ile	80%ile	90%ile	95%ile	99%ile
Demographic index	0	0	0.17	0.29	*0.36*	0.50	0.56	0.69	0.77	0.85
% low-income	0	0	0.18	0.29	*0.33*	0.46	0.50	0.62	0.70	0.81
% minority	0	-	0.11	0.29	*0.39*	0.61	0.70	0.88	0.95	0.99
% less than high school	0	-	0.0004	0.09	*0.13*	0.18	0.20	0.29	0.37	0.49
% linguistic isolation	0	-	-	0.01	*0.04*	0.06	0.07	0.14	0.21	0.34
% under 5	0	-	0.03	0.06	*0.06*	0.08	0.09	0.11	0.13	0.16
% over 64	0	0	0.08	0.13	*0.15*	0.19	0.20	0.25	0.29	0.39

Source: 2019 version of EJSCREEN, not including Puerto Rico. Calculated based on 2013-2017 5-year summary file, American Community Survey (ACS), from the US Census Bureau.

Note: Population percentiles (and means) are shown, not block group percentiles (or means), so 80%ile means 80% of the population has a lower (or exactly tied) block group value. Values in table have been rounded to an integer percentile 0-100. Numbers may differ slightly from those in EJSCREEN reports.

[INTENTIONALLY BLANK]

APPENDIX E. FORMULAS FOR DEMOGRAPHICS AND EJ INDEXES

The EJ indexes rely on demographic indexes combined with environmental indicators. The demographic and EJ indexes are calculated as follows:

Demographic Index

This is the average of percent minority and percent low income in the block group. Percent low income is defined in Appendix B, and is essentially all residents where household income is below twice the federally defined poverty threshold, as a percentage of all those for whom this poverty ratio could be determined (typically known for the vast majority of the block group's population).

> Demographic Index = (% minority + % low-income) / 2

EJ Index

The EJ Index measures how much a particular place contributes to overall nationwide differences in environmental indicator values between demographic groups. This EJ index is a combination of a block group environmental factor, the population of the block group, and the demographic composition of the block group. In this index, the demographic composition of the block group is the difference between the block group's composition and the national average, as measured by the demographic index.

> EJ Index =
>
> (Environmental Indicator)
> X (Demographic Index for Block Group – Demographic Index for US)
> X (Block Group Population)

APPENDIX F. QUALITY CONTROL / QUALITY ASSURANCE

EPA's quality control guidelines emphasize transparency and reproducibility as useful in ensuring the quality of data. EPA is providing a very high level of transparency in EJSCREEN by taking several steps described here.

The EJSCREEN Technical Documentation (this document) has extensive details on the precise sources and exact methods used, to ensure transparency. The transparency of the data inputs is also ensured through references to further technical documentation from the providers of those data inputs, such as the PM2.5 and ozone estimates, NATA, the Census demographic data, and the DOT traffic database. Metadata is linked from the web-based tool, providing quick access to further technical details.

Furthermore, the full raw database of EJSCREEN indicators and indexes, and supplementary material, will be available to expert users who wish to go beyond the web-based interface and conduct further analysis or research. EPA also hopes to make available the Python and R code used to develop all the indicators in EJSCREEN, including proximity scores, percentiles, and so on. Access will be provided through the data download section of the EJSCREEN website (http://www.epa.gov/ejscreen).

Extensive quality control/ quality assurance efforts were made in the development of EJSCREEN and a very brief summary is provided here.

The starting point for most of the environmental indicators was information provided by EPA Offices (i.e., latitude/ longitude data used to create proximity indicators, and the NATA results). Those sources of information had already been subject to QA procedures in the respective offices, and the information had already been released to the public. EPA's Office of Air and Radiation (OAR) provided PM2.5 and ozone estimates based on public monitoring data, CMAQ results, and a fusion model to combine them. The CMAQ and fusion model have previously been extensively documented in peer-reviewed journal articles (Byun & Schere, 2006; Berrocal, Gelfand, & Holland, 2010a, 2010b, 2011). The lead paint indicator was calculated from Bureau of Census data by an EPA contractor, and then independently replicated by EPA, through separate ACS downloads and calculations. The traffic indicator was calculated from publicly available DOT data, as explained in this report.

The calculations of environmental indicators from those inputs was conducted by an EPA contractor for the proximity scores and the lead paint score, using their established QA/QC procedures, so EPA did not attempt to replicate the proximity calculations. These calculations involved time-consuming proximity calculations, and simple calculation of the lead paint indicator. The NATA and PM and ozone indicators were simply taken directly from EPA and used in EJSCREEN.

The demographic indicators were calculated by an EPA contractor based on ACS data they obtained from the Census Bureau. EPA was able to independently replicate 100% of the resulting indicators by separately obtaining the raw ACS data.

The same is true for the EJ Indexes and all of the percentiles, map bins (for color-coded maps), and popup text fields used in EJSCREEN – the entire geodatabase was independently replicated by EPA using only the environmental indicators as a starting point, and applying alternative algorithms and code for development of percentiles and bins, as well as the rounding procedures defined for popup text.

EPA was also able to conduct some limited manual replication of buffer calculations, although truly independently replicating those GIS algorithms is challenging given the need to use data on millions of blocks and the challenge of identifying relevant blocks in a fashion that is independent of the geoprocessing tool used for buffer analysis in EJSCREEN. Spot checks were conducted on buffer reports to ensure raw data and percentile calculations, use of lookup tables, rounding, significant digits, and floored percentiles were all handled correctly.

The extensive QA/QC process did uncover numerous complex data challenges early in the process, and ultimately lead to a final database that could be fully independently replicated from environmental indicators calculated from public information, providing strong assurance of the integrity of the data processing and calculations.

APPENDIX G. PEER REVIEW

EJSCREEN was submitted for peer review in early 2014, through a letter review process conducted by a contractor with extensive experience in organizing peer review. Based on pre-defined criteria regarding level of expertise in relevant subject areas, four experts were identified.

The reviewers were provided with a draft of this technical documentation, describing EJSCREEN's development, purpose, and use of selected environmental indicators, demographics, and EJ indexes for screening and mapping.

Reviewers were also provided a live webinar presentation and demonstration, along with time for questions for EPA. One of the four was unable to attend the webinar but contacted EPA with questions that EPA responded to in a phone conference call.

The reviews were completed in March of 2014. Each of the four reviewers provided a detailed discussion of their technical comments, concerns, and recommendations.

All four agreed that the new environmental justice screening tool will be helpful to its users and is generally very well done. Each did point to some weaknesses in the tool, suggesting that correcting these shortcomings in the next version could strengthen the tool and help its users.

Of the more than 100 distinct comments from the expert peer reviewers, more than one third were positive statements about the quality of EJSCREEN and the documentation. A sampling of direct quotes includes the following:

- "I would like to commend the EPA … it does represent a major step forward and the EPA should be recognized for this achievement"
- "This documentation fairly represents the tremendously difficult task of creating this tool"
- "very impressed by the quality of the work"

All of the reviewers also agreed that the EJSCREEN documentation is generally well-written, clear, and easy to follow. Reviewers did ask for editorial changes, clarification, or further rationale in the documentation, and such comments represented about one fourth of all the comments received. They asked for clarification in some specific sections, such as more discussion of which indicators were chosen and why, and which were left out and why. Many of these comments have already been taken into consideration in this version of the Technical Document.

About one third of the comments were suggestions or requests for new data (in reports, maps, and data files), typically recommending new or improved environmental indicators (e.g., air quality, water quality, more facility types, etc.). Some comments made suggestions that would involve adding a new feature to the EJSCREEN tool, rather than improving the data layers or the documentation. These will be taken into consideration in discussions of possible future updates or upgrades to the tool.

A handful of comments raised policy considerations and inherently challenging issues in screening and mapping. These involved basic policy questions such as what is the best spatial resolution for these maps, and whether to combine all 11 EJ indexes (Reviewers were divided on this topic). The 2017 version of EJSCREEN continues to use block groups, but recommends an emphasis on buffers as less uncertain than a single block group estimate. It continues to use 11 separate indexes, but these issues can be a topic of continuing discussions and exploration in the future as the public and others work with the new tool.

On the whole, the reviewers' suggestions have already served to strengthen the tool and its documentation, and will continue to inform discussions. By elaborating and clarifying the options and choices made, EPA can help the users of EJSCREEN better understand its potential and its limitations. Improved data and methods should be considered as well in future versions of EJSCREEN. EPA looks forward to working across various offices, with stakeholders across the nation, and academics as well as the public, on implementation and future enhancement of EJSCREEN.

APPENDIX H. INITIAL FILTER APPROACH FOR SCREENING

What is the 80th percentile filter?

In past screening experience, EPA has found it helpful to establish a suggested Agency starting point for the purpose of identifying geographic areas that may warrant further consideration, analysis, or outreach. The use of an initial filter promotes consistency and provides a pragmatic first step for EPA programs and regions when interpreting screening results. For early applications of EJSCREEN, EPA identified the 80th percentile filter as that initial starting point. In other words, an area with any of the 11 EJ indexes at or above the 80th percentile nationally should be considered as a potential candidate for further review. Further review may include considering other factors and other sources of information such as health based information, local knowledge, proximity and exposure to environmental hazards, susceptible populations, unique exposure pathways, and other federal, regional, state, and local data. This filter is simply a starting point, and program offices and regions should perform additional analysis before making any decisions about potential environmental justice issues. As EPA gains further experience and insight into the performance of the tool and its applicability for different uses, program offices and regions may opt to designate starting points that are more inclusive or specifically tailored to meet programmatic needs more effectively.

The 80th percentile filter in EJSCREEN is not intended to designate an area as an “EJ community.” EJSCREEN provides screening level indicators, not a determination of the existence or absence of EJ concerns. Nor does the use of the 80th percentile filter suggest that all of the 11 environmental indicators are equal in terms of their impact on human health and the environment. Instead, the 80th percentile filter encourages programs to consider environmental indicators outside of their areas of concentration. The Agency may revise this approach in the future based on experience. This 80th percentile filter is for internal EPA use and is not intended to apply to States or other organizations.

[page intentionally blank]