

Data Sources for an Environmental Quality Index: Availability, Quality, and Utility

Danelle T. Lobdell, PhD, Jyotsna S. Jagai, PhD, Kristen Rappazzo, MPH, and Lynne C. Messer, PhD

Environmental public health research is conducted at multiple geospatial levels. State and national indicators are typically used by agencies to assess country-wide trends. For instance, the US Environmental Protection Agency (USEPA) and the Centers for Disease Control and Prevention (CDC) report the results of state and national tracking efforts regularly, in reports like the USEPA's Report on the Environment (ROE),¹ America's Children and the Environment,² and CDC's Environmental Public Health Tracking.³ Environmental public health research also occurs at smaller geographic units. County-level research exploring the impacts of air,⁴⁻⁸ water,^{9,10} land,¹¹⁻¹³ and built environment^{14,15} exposures on morbidity and mortality outcomes is well-represented in the literature. Research conducted at more "local" units of aggregation is also critical to understanding the effects of environmental insults on public health. What constitutes "local" varies within the research, with such distinctions as city,^{16,17} census tract,¹⁸⁻²⁰ or even rural or farm.²¹⁻²⁷ It is only through evidence, collected at multiple nested geospatial levels, that the full range of environmental public health effects can be observed.

Although 1 strength of environmental public health literature is its geographic breadth (ranging from national to local levels), 1 limitation is its restriction to single media—or even single contaminant—assessment, which fails to address the multiple environmental conditions to which people are simultaneously exposed. Environmental disamenities such as landfills or industrial plants are often located in neighborhoods with a high percentage of minority and/or poor residents,²⁸⁻³⁴ potentially contributing to adverse health outcomes and health disparities.^{33,35-38} High income neighborhoods are more likely to contain amenities conducive to promoting good health (e.g., parks, grocers).^{39,40} The differential distribution of disamenities results in the clustering of adverse exposures; this environmental

Objectives. An environmental quality index (EQI) for all counties in the United States is under development to explore the relationship between environmental insults and human health. The EQI is potentially useful for investigators researching health disparities to account for other concurrent environmental conditions. This article focused on the identification and assessment of data sources used in developing the EQI. Data source strengths, limitations, and utility were addressed.

Methods. Five domains were identified that contribute to environmental quality: air, water, land, built, and sociodemographic environments. An inventory of possible data sources was created. Data sources were evaluated for appropriate spatial and temporal coverage and data quality.

Results. The overall data inventory identified multiple data sources for each domain. From the inventory (187 sources, 617 records), the air, water, land, built environment, and sociodemographic domains retained 2, 9, 7, 4, and 2 data sources for inclusion in the EQI, respectively. However, differences in data quality, geographic coverage, and data availability existed between the domains.

Conclusions. The data sources identified for use in the EQI may be useful to researchers, advocates, and communities to explore specific environmental quality questions. (*Am J Public Health.* 2011;101:S277–S285. doi:10.2105/AJPH.2011.300184)

injustice has been well-noted within the literature.⁴¹⁻⁴⁴

Single exposure models remain the typical approach to environmental exposure assessment^{31,45} for a variety of reasons. Collecting exposure data on multiple media can be time-consuming and expensive, investigators tend to specialize in 1 environmental medium, and most research projects are insufficiently powered to accommodate multiple variables representing different environmental exposures in a single model. Estimating a more complete range of environmental exposures, as opposed to the single exposure model, would produce a more realistic exposure characterization. More complete environmental exposure estimation would also move the field toward addressing the Institute of Medicine's recommended increased efforts toward "...the collection and coordination of environmental health information and to better link it to specific population and communities of concern."⁴⁶

In an effort to learn more about how multiple environmental factors combine to

contribute to adverse health outcomes, and to better estimate the larger environmental and social context to which humans are continuously exposed, we are developing an environmental quality index (EQI) for all counties in the United States.⁴⁷ We identified 5 environmental domains: air, water, land, built environment, and sociodemographic. Within the 5 domains, we explored the availability and quality of data that could be used to develop the EQI. This article describes the quality and availability of the data used to construct the EQI.

METHODS

We initially identified 3 environmental domains: air, water, and land, based on the media chapters from the ROE.¹ We added 2 additional domains to account for the built and sociodemographic environments as part of overall environmental quality. An inventory of possible data sources representing each of the 5 domains were identified using web-based search engines (e.g., Google), site-specific search engines (e.g., federal and state data sites), literature-reported

data sources (e.g., PubMed, ScienceDirect, Toxnet), and personal communication from data owners. We sought data that were available at—or had the potential to be aggregated to—the United States county level. We restricted data to 2000 to 2005 to coincide with the available sociodemographic and health data used for initial testing of the EQI. For each data source identified, we collected the following information (when available): data title, source URL, data description, data ownership, data provider, data format, secondary data format, data geometry, geographic coverage, smallest geographic unit represented, data resolution, record start and end years, date data published, data refresh frequency, metadata availability, metadata link/location, method to obtain data, point of contact information, data constraints, and data limitations. Within each domain, a database containing information on each identified dataset was compiled.

Data sources were assessed for EQI inclusion based on temporal, spatial, and quality-related criteria. Temporal appropriateness required data to be maintained within the 2000 to 2005 period. Data sources were considered spatially appropriate if 2 criteria were met: (1) data were available at the county level; and (2) data were available for all 50 states. Temporal and spatial coverage of the data sources were balanced against data quality. Data quality, especially related to data source documentation, was judged by the data source managers (in data reports and internal documentation), project investigators, and by the larger field of environmental research, through use and critique of the various data sources. Additionally, some data sources were redundant (e.g., EPA air monitoring data and state reports of the same), resulting in the elimination of specific data sources.

In this article we present the final data sources selected for use in the EQI construction.⁴⁷

RESULTS

For each of the 5 domains, we described selected data sources and identified their respective strengths and limitations (Table 1). The sources discussed were those that will be included in the EQI. A more comprehensive version of Table 1 is available as

a supplement to the online version of this article at <http://www.ajph.org>.

Air Domain

Three data categories were considered: monitoring data, emissions data, and modeled estimates representing concentrations of either criteria air pollutants or hazardous air pollutants (HAPs). Twelve data sources were identified and 7 were considered for inclusion; ultimately 2 were identified as the most complete for use in the air domain of the EQI.

The Air Quality System (AQS)⁴⁸ is a repository for criteria ambient air pollution data collected by federal, state, local, and tribal agencies from thousands of monitors for the EPA's ambient air monitoring program across the United States. Monitored pollutants include all criteria air pollutants, particulate matter species, and approximately 60 ozone precursors. Major strengths of the AQS are that data are measured, rather than modeled, and these measurements are synchronized across the country. Monitors in the network and the reported data are audited regularly for accuracy and precision. However, most of the ambient air monitors are located in or near urban areas, leaving many United States counties without reported data. In addition, the AQS provides sparse and limited data collection for HAPs.

The National-Scale Air Toxics Assessment (NATA) database⁴⁹ uses data from the National Emissions Inventory to construct air dispersion models for estimating ambient concentrations of HAPs at the county and census tract level. Emissions data are constructed every 3 years, beginning in 1996, and are used to provide annual estimates. The NATA databases contain estimated ambient concentrations for 177 to 180 of the 187 HAPs, and uses validated models that take meteorology and chemical dispersion into account. The methodology for estimating concentrations may change between assessments, but these modifications are well-documented and justified. As a result of minor modifications, however, although the ambient concentrations may be comparable over time, some difference between estimates is attributable to methodological modifications. The temporal resolution of the assessments is adequate for the intended EQI, but because of the 3-year release schedule, there are gaps in temporal coverage.

Water Domain

Five broad data categories within the water domain were identified: modeled, monitoring, reported, survey/study, and miscellaneous data. Eighty data sources were identified, and 9 were selected. Two of the data sources are repositories maintained for compliance with federal regulations, and both were categorized as “miscellaneous” because they included monitored, reported, and survey/study data. The National Water Information System⁵⁰ is a repository maintained by US Geological Survey, which includes monitoring data from streams. The STORET (STORage and RETrieval) Database⁵¹ is a repository of water quality monitoring data maintained by EPA and collected by water resource management groups across the country. Both repositories include several measures of water quality; however, few have the spatial and temporal coverage required for the EQI. Therefore, selected data from both repositories will be combined to create a complete dataset with the appropriate spatial and temporal coverage for specific parameters.

The Watershed Assessment, Tracking & Environmental Result Program⁵² database represents the surface water assessment programs under the Clean Water Act. Data are maintained at the state level and reported to the federal system. Although all states report county level data, there is little consistency. These data are geocoded to a specific stream length in the National Hydrography Dataset via the Reach Attribute Database. The geocoded Watershed Assessment, Tracking & Environmental Result Program data can be used to calculate human exposure related variables, such as percentage of stream length impaired for recreational use.

The National Contaminant Occurrence Database⁵³ is a surveillance database maintained to satisfy the requirements of the Safe Drinking Water Act and includes information on contaminants in public water supplies. This survey is conducted every 6 years, and data are provided by public water supplies. The Safe Drinking Water Information System⁵⁴ contains information from public water systems and violations of EPA's drinking water regulations. The number and type of violations reported by each water supply can be calculated using this database. The Estimates of Water Use in the United States,⁵⁵

TABLE 1—Sources of data for air, water, land, built environment, and sociodemographic domains for use in the EQI

Source of Data	Description	Strengths	Limitations
Air domain			
Air Quality System ⁴⁸	Repository of ambient air quality data, including both criteria and hazardous air pollutants	Measured values, criteria air pollutants network is substantial, measurement occurs regularly and is synchronized, data are audited for accuracy and precision	HAPs network is sparse, some counties have no monitors, necessitating interpolation of concentrations for unmonitored locations
National-Scale Air Toxics Assessment ⁴⁹	Estimates of HAPs concentrations using emissions information from the National Emissions Inventory and meteorological data input into the Assessment System for Population Exposure Nationwide model	Validated models, coverage for all US counties, majority of HAPs included	Data are available at 3-y intervals; may underestimate concentrations; uses simplifying assumptions when information is missing or of poor quality; changes in methodology may result in different estimates between years
Water domain			
National Water Information Systems ⁵⁰	Large data repository of monitoring stream quality data maintained by USGS, also includes data from several National Water Quality Assessment Program studies	Includes data on many measured parameters	Coverage for most parameters is limited both spatially and temporally; therefore may only be able to use a few parameters for EQI
STORET (STORage and RETrieval) ⁵¹	Repository maintained by EPA of water quality monitoring data collected by water resource management groups across the country	Includes data on many measured parameters	Coverage for most parameters is limited both spatially and temporally; therefore may only be able to use a few parameters for EQI
WATERS Program Database/Reach Address Database ⁵²	Collection of EPA water assessments programs, including impairment, water quality standards, pollutant discharge permits, and beach violations	Only database maintaining information on EPA Clean Water Act regulations	Data maintained and provided by states and therefore difficult to compare across states and not consistently reported
National Contaminant Occurrence Database ⁵³	Samples both regulated and unregulated contaminants in public water supplies, maintained by EPA to satisfy statutory requirements for Safe Drinking Water Act	Provides measures for several chemicals and pathogens that are not measured elsewhere	Data provided by public water supplies; therefore need to use spatial aggregation to get county level estimates
Safe Drinking Water Information System ⁵⁴	Large database of measured drinking water quality data maintained by EPA	Only database with information on drinking water quality	Coverage for many parameters is limited both spatially and temporally; therefore may only be able to use a few parameters for EQI
Estimates of Water Use in US ⁵⁵	County level estimates of water withdrawals for domestic, agricultural, and industrial use calculated by USGS	County level data	Estimated based on various data sources
Drought Monitor Data ⁵⁶	Geographic Information System raster files reporting weekly modeled drought conditions. A collaboration, including National Atmospheric and Oceanic Administration, US Agricultural Service, and academic partners	Weekly coverage for the entire country	Modeled data Raster data; therefore will require spatial aggregation
National Atmospheric Deposition Program ⁵⁷	Measures deposition of various pollutants, such as calcium, sodium, potassium, and sulfate, from rainfall	Weekly coverage for the entire country	Data not at the county level and will require spatial aggregation
Nutrient Loss Database for Agricultural Fields in US ⁵⁸	Database maintained by the Agricultural Research Service of nitrogen and phosphorous loss on agricultural fields	Only database that considers agricultural impacts on water quality	Limited to areas with heavy agriculture

Continued

TABLE 1—Continued

	Land domain		
County Pesticide Use Estimates ⁵⁹	Information for 220 pesticides on the average amount (pounds) applied to 87 agricultural crops and the acres of crops treated for US counties	Combined state pesticide use coefficients from National Center for Food and Agricultural Policy and county harvested crop acres from the Census of Agriculture	May not be updated to include relevant time periods
2002 Census of Agriculture Full Report ⁶⁰	Summary of agricultural activity, including number of farms by size and type, inventory and values for crops and livestock, operator characteristics	Can be used to approximate land and water-related agricultural outputs (e.g., potential pesticide burden per acre, potential exposure to cattle, dust)	Not direct measures of pesticides or probable exposures
Dun and Bradstreet Agriculture Data ⁶¹	Agricultural data licensed for the entire US, including information about crops and livestock	Additional data source for exposure and sensitivity analyses	Proprietary data, not publically available
Web Feature Service for National Priority List (NPL) Sites ⁶²	Maintained by EPA and provides locations of and information on sites that have been placed on the NPL throughout the US	Indicators for major facilities (e.g., superfund sites, Large Quantity Generators [LQG], Toxics Release Inventory [TRI]) are available	Contains much more information than just the facilities, type, and location
Superfund NPL Sites ⁶³	Maintained by EPA and provides locations of and information on sites that have been placed on the Superfund NPL throughout the US	The Superfund NPL lists the national priorities among the known or threatened hazardous substance releases throughout the US and its territories	These data, including locations, have been gathered from multiple sources; users of the data (or any portion thereof) may wish to verify its accuracy before use
Resource Conservation and Recovery Act (RCRA) Treatment, Storage and Disposal (TSD) Facilities and Corrective Action Facilities ⁶⁴	The RCRA TSD and Corrective Action Facility data in this file are part of the EPA Geodata shapefile available through the EPA Geospatial Data Access Project and Envirofacts database	Other data about facilities, sites, or places subject to environmental regulation or of environmental interest are also included	These data, including locations, have been gathered from multiple sources; users of the data (or any portion thereof) may wish to verify its accuracy before use
RCRA LQG ⁶⁵	The RCRA LQG data in this file are part of the EPA Geodata shapefile available through the EPA Geospatial Data Access Project and Envirofacts database	5 different types of hazardous waste generation tracked and recorded; includes addresses and types of hazardous material generated; data about areas of environmental interest are also included	These data, including locations, have been gathered from multiple sources. Users of the data (or any portion thereof) may wish to verify its accuracy before use
TRI sites ⁶⁶	TRI data in this file are part of the EPA Geodata shapefile available through the EPA Geospatial Data Access Project and Envirofacts database	The TRI is a publicly available database that contains information on more than 650 toxic chemicals that are being used, manufactured, treated, transported, or released into the environment, among others	These data, including locations, have been gathered from multiple sources; users of the data (or any portion thereof) may wish to verify its accuracy before use
Assessment, Cleanup, and Redevelopment Exchange (ACRES) Brownfield sites ⁶⁷	ACRES Brownfields data in this file is part of the EPA Geodata shapefile available through the EPA Geospatial Data Access Project and Envirofacts database.	ACRES is an online database for Brownfields Grantees to electronically submit data directly to EPA only brownfields property interests maintained in the ACRES are provided in the file	These data, including locations, have been gathered from multiple sources; users of the data (or any portion thereof) may wish to verify its accuracy before use
Section Seven Tracking System (SSTS) pesticide producing site locations ⁶⁸	SSTS data in this file are part of the EPA Geodata shapefile available through the EPA Geospatial Data Access Project and Envirofacts database	SSTS tracks the registration of all pesticide-producing establishments and the types and amounts of pesticides, active ingredients, and related devices that are produced, sold, or distributed, annually	These data, including locations, have been gathered from multiple sources; users of the data (or any portion thereof) may wish to verify its accuracy before use
National Geochemical Survey ⁶⁹	Geochemical data (e.g., arsenic, selenium, mercury, lead, zinc, magnesium, manganese, iron) for US based on stream sediment samples	Provides county-level means and standard deviations for each element; sampled data interpolated over nonsampled space results in variance measures	Includes data from several surveys, therefore sampling locations and number of samples available vary by location

Continued

TABLE 1—Continued

Map of Radon Zones ⁷⁰	Identifies areas of the US with the potential for elevated indoor radon levels, maintained by EPA	Each US county is assigned to 1 of 3 radon zones based on radon potential	Data are not actual measurements of radon and only 3 levels of radon potential reduces possible county level variability
Dun and Bradstreet North American Industry Classification System codes ⁶¹	Description of physical activity environment (e.g., recreation facilities, parks, physical fitness-related businesses) food environment (fast food restaurants, grocery, convenience stores) education environment (schools, daycares, universities) per county	Built environment domain Detailed, thorough data; geocoding to county level sufficiently large geographic unit to reduce geocoding errors; ongoing updates.	Proprietary data, not publically available
Topologically Integrated Geographic Encoding and Referencing ⁷¹	Road type and length per county	National coverage	Different road types may not be equivalent across US geography; confer different exposure risks
Fatality Annual Reporting System ⁷²	Annual pedestrian-related fatality per 100,000 population maintained by National Highway Safety Commission	County level reports and annual updates	Pedestrian fatalities result from diverse types of events and are not well-captured in database
Rural-Urban Continuum Codes ⁷³	A 9-part classification scheme to distinguish metropolitan, nonmetropolitan, and rural areas using data from the 2000 US Census	Allows researchers to break counties into finer residential groups (than metro–nonmetro dichotomy), for analysis of trends that may be related to degree of rurality and metro proximity	Updated every 10 y; definition changes make 2003 codes not comparable with previous years
US Census ⁷⁴	County level population and housing characteristics, including density, race, spatial distribution, education, socioeconomic, home and neighborhood features, and land use	Sociodemographic domain Uniformly collected and constructed across US and can be used for construction of a variety of different variables	Data are available at 3 y intervals, may underestimate concentrations; uses simplifying assumptions when information is missing or of poor quality
Uniform Crime Reports ⁷⁵	County level reports of violent crime	General estimate of public safety exposure	Reporting may differ across geography
Home Mortgage Disclosure Act Data ⁷⁶	Data collected from lending institutions from home mortgage loan applications, including the application outcome, property location, and applicant-related information	Annually updated, uniformly collected and useful for identifying disparities in home loan access and subprime lending	Not all institutions required to file; therefore coverage may be incomplete; areas with low home-ownership rates will have the least amount of information

Note: EPA = Environmental Protection Agency; EQI = Environmental Quality Index; HAPs = hazardous air pollutants; LQG = Large Quantity Generators; TRI = Toxics Release Inventory; US = United States; USGS = United States Geological Survey.

which is modeled by the US Geological Survey, provides county level estimates of water withdrawals for domestic, irrigation, livestock, and industrial use, an indication of water stress in a county.

Two data sources provide information on meteorological impacts on water quality. The Drought Monitor Data⁵⁶ are modeled weekly drought conditions. The National Atmospheric Deposition Program⁵⁷ provides weekly measures and national coverage of the deposition of various pollutants from rainfall using monitors around the country.

The Nutrient Loss Database for Agricultural Fields⁵⁸ will be used to estimate impacts of agriculture on water quality. It is the only database that estimates direct impacts of agriculture on water quality and provides data on nitrogen and phosphorous loss on agricultural fields. These data exist only for United States areas with high agricultural production. Although not nationally representative, these data will be included to account for water quality impacts in areas with heavy agriculture.

Land Domain

Land domain data sources were grouped into 4 categories: agriculture, industrial facilities, geology and mining, and land cover. Eighty sources were identified and 12 were retained; 3 from agriculture, 7 from facilities, and 2 from geology/mining. Because of the lack of previous associations with human health, none were retained from land cover.

The 3 agricultural data sources considered for inclusion in the EQI are County Pesticide Use Estimates⁵⁹ for 220 compounds, 2002 Census of Agriculture Full Report,⁶⁰ and Dun and Bradstreet Agriculture Data.⁶¹ Two of these databases will be used to estimate the pesticide burden in the ambient environment and 1 will be reserved for sensitivity analyses. A significant limitation of the County Pesticide Use Estimates involved the timing of data collection because these data were last collected in 1997, which was before our stated exposure window. These data also provide the best human exposure estimate for pesticides. The Census of Agriculture data provides mostly farm-related summary characteristics and does not offer direct pesticide measures or probable exposure information. As a strictly environmental indicator, the Census of Agriculture is useful, but its ability to link to

human health is somewhat limited. Because no single database provides complete coverage, these 2 data sources will be compared and merged to generate a pesticide level rank, which will be a more robust ambient environment measure. The Dun and Bradstreet agricultural data are similar to the Census of Agriculture data, with many of the same strengths and limitations. Therefore, we will use these data for sensitivity analyses.

The industrial facilities data source retained included the EPA's Web Feature Service for National Priority List Sites,⁶² the Superfund National Priorities List sites,⁶³ the Resource Conservation and Recovery Act Treatment, Storage, and Disposal and Corrective Action Facilities,⁶⁴ the Large Quantity Generators,⁶⁵ Toxic Release Inventory sites,⁶⁶ Assessment, Cleanup, and Redevelopment Exchange Brownfield sites⁶⁷ and the Section Seven Tracking System Pesticide Producing site locations.⁶⁸ All facilities-related data were retained for inclusion in the EQI with extensive information on each facility.

The 2 geology/mining data sources are the National Geochemical Survey (NGS)⁶⁹ and the Map of Radon Zones.⁷⁰ The NGS data provide the mean and standard deviations (SD) for multiple soil chemicals. However, these values are calculated from multiple surveys of soil samples over several years, and therefore, combine many varying sources of data. The radon map assigns a radon potential level to each county in the United States. As the data source provides radon potential, not actual measurement, these data are limited. The 3-level radon categorization masks important radon-level heterogeneity across the United States. Despite the limitations, both of these data sources provide land-related data not available elsewhere.

Built Environment Domain

Built environment data sources were grouped by topic: traffic-related, transit access, pedestrian safety, access to physical activity, food environment, school or educational environment, and household health measures. Twelve data sources were identified, and 4 were retained: 1 traffic-related, 1 for pedestrian safety, 1 for use in physical activity, food, and educational environments, and 1 for the urban/rural residence. Because of the noncomparable county level data quality, none of the transit

access or household health measures were retained.

For the traffic-related data source, Topologically Integrated Geographic Encoding and Referencing⁷¹ was retained. The files provide relatively uniform and nation-wide coverage. From these files, county-specific proportions for various road types will be characterized. Unfortunately, considerable heterogeneity may be lost; for instance, a tertiary road in Maryland may not be qualitatively equivalent to one located in Wyoming.

The fatality annual reporting system of the National Highway Safety Commission⁷² was retained as part of pedestrian safety because of its national coverage. The data are regularly updated and available from the website. One limitation of these data are that pedestrian fatalities result from diverse types of events (e.g., from crossing busy intersections or deserted highways), but this diversity is not well-captured.

Dun and Bradstreet NAICS (North American Industry Classification System) codes⁶¹ will be used as the data source to estimate 3 different topics: physical activity, food, and educational environments. Although these data have sometimes been criticized for inadequate spatial resolution (e.g., inaccurate geocoding to small units of aggregation like census tracts),⁷⁷ they should be sufficient as a construct for county level food, physical activity, and educational environments.

The US Census' rural-urban continuum code (RUCC)⁷³ was retained for urban/rural residence. RUCC was developed for use at the county level, which makes it appropriate for use in the EQI.⁷⁹ Further, it delineates nonmetropolitan counties by degree of urbanization and adjacency to a metropolitan area or areas. Like all decennial census data, it is only updated every 10 years, meaning those geographic areas where change occurs quickly (either growing or declining) will become outdated at a rate that is not comparable to the rest of the country.

Sociodemographic Domain

Few sociodemographic data sources are available. Only 3 data sources were identified and retained for sociodemographic data: the US Census Bureau,⁷⁴ the Uniform Crime Report⁷⁵ and the Home Mortgage Disclosure Act database.⁷⁶ Each of these data sources represent critical aspects of the human sociodemographic environment, are updated regularly, and are

available at the county level for the entire country.

DISCUSSION

The ability to estimate the full range of environmental exposures is critical for environmental justice, both to document and then address environmental health disparities. The process used to create the EQI, as well as the EQI tool itself, is useful for investigators researching health disparities. The process described is easily replicable at various units of geographic aggregation, and the EQI tool enables researchers interested in a given environmental exposure to control for the county level environmental conditions with only 1 (EQI) variable.

This article described the environmental quality data that are available at the county level across the United States. The majority of data sources identified are publically available, which community leaders, advocates, and residents can explore for environmental health questions. Data sources were found to represent each of the 5 *a priori* identified domains: air (12 data sources identified and 2 retained), water (80 sources identified and 9 retained), land (80 sources identified and 6 retained), built environment (12 sources identified and 4 retained), and sociodemographic (3 sources identified and retained). Each data source was used previously in published literature and was reasonably well documented. Despite finding a considerable number of data sources to represent the environmental domains, significant data gaps exist.

Environmental data sources are often plagued by inadequate spatial and temporal coverage. Most of the data sources presented in the “Results” section will require either spatial interpolation to achieve county level estimates or temporal interpolation to achieve annual estimates, or both. For example, even with extensive air monitoring networks, the measured spatial coverage of the United States is incomplete, particularly in rural areas. Some types of exposures are disproportionately located in urban areas (e.g., particulate matter), whereas others are found in rural areas (e.g., industrial livestock operations). The nonrandom distribution of environmental risk means that virtually all

interpolated data are inaccurate, impairing the ability to assess how pollutants differentially impact urban and rural areas. Ultimately, environmental justice efforts suffer from incomplete data.

Environmental data are also rarely collected with the temporal frequency required to adequately assess health relationships. Although data on some parameters are collected on a consistent and frequent basis, the majority are collected infrequently. Water data, for instance, are only sporadically collected in response to a particular query or based on regulatory statute. Within the sociodemographic domain, the complete United States census is collected decennially, which limits investigators’ capacity to explore temporal changes. Characteristics of places can change rapidly, but under current data collection schedules, this cannot be accounted for.

Many environmental parameters are compiled at a smaller unit of aggregation (e.g., for a municipality or city), and most are not maintained in a single source, such as a data repository. Although national repositories for some domains exist (e.g., water, air), often in response to federal regulations, no built environment repository exists (for transit, walkability/physical activity, street connectivity, presence of sidewalks, or pedestrian lighting measures). Localities with limited funds may not be motivated, or able, to collect these data.

From a human health perspective, probably the biggest limitation to existing environmental data sources is that data are collected with little thought given to potential health impacts. For instance, monitoring sites may collect relevant air pollutant data, but their location (e.g., air monitors located on top of buildings) is inappropriate for assessing the street level values to which humans are exposed. Pesticide data, from the land domain, usually reports pesticides sales in relation to crops and livestock, not application, handling, or disbursement. Even the US Census, which is widely used in health research, is primarily collected for tax and political districting purposes. Some of the data sources identified have not been used in human health research, and as such, are a limitation. Regularly collected, high quality data that considers probable human health impacts, would make the task of assessing differential exposures considerably easier.

Existing environmental data are not collected to address environmental justice concerns. Data monitoring and collection, especially in the air domain, are typically focused in highly populated urban areas; therefore, data are sparse for rural areas. The data are also often collected at a scale (e.g., county) that may mask local, heterogeneous environmental conditions, which may underestimate environmental injustices. These data gaps can be addressed by increasing monitoring and data collection in areas with known environmental injustices, such as near industrial livestock operations, landfills, or in low socioeconomic status urban neighborhoods.

The EQI is a national-level index that will allow understanding of how multiple environmental conditions affect United States counties. At its current county level scale, the EQI may not reveal environmental injustices seen at the local community level. However, it will highlight those counties experiencing an increased burden of environmental impacts. Further, we are contributing to environmental justice endeavors by describing the following: the process by which we obtained data, construction of the EQI, and the websites containing available data that can be used to construct indices at different levels of aggregation. Interested investigators are encouraged to consider constructing local EQIs, and to add relevant, local level data for focused comparisons. A review by Lioy et al⁷⁹ provided a useful way to incorporate different levels of data (national, state, local) for localized exposures that could be adapted for ambient environmental quality.

In this article, we described the process and lessons learned from reviewing national data for constructing the EQI. Although the EQI will be a tool for documenting environmental injustices, environmental conditions, particularly in communities with clustered environmental hazards, will likely be underrepresented by a county level EQI instrument. Future work will consider the EQI’s sensitivity to regional and aggregation level (county versus block group) construction.

This article presented only a fraction of the data sources identified as part of the data inventory. The full inventory is available at <http://www.epa.gov/nheerl/eqi> for others to explore.⁴⁷ This website will be updated and available as a resource to communities and other

researchers as each of the domains, constructed indices, and ultimately, the final EQI are developed. In this way, we hope that this project can contribute to increased effort in the “...coordination of environmental health information and to better link it to specific population and communities of concern” as advised by the 1999 Institute of Medicine report.⁴⁶ ■

About the Authors

Danelle T. Lobbell and Jyotsna S. Jagai are with the Office of Research and Development, US Environmental Protection Agency Research Triangle Park, NC. Kristen Rappazzo is with the Gillings School of Global Public Health, Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill. Lynne C. Messer is with the Center for Health Policy and Inequalities Research, Duke Global Health Institute, and the Children's Environmental Health Initiative, Nicholas School of the Environment, Duke University, Durham, NC.

Correspondence should be sent to Danelle T. Lobbell, US Environmental Protection Agency, MD 58A, Research Triangle Park, NC 27711 (e-mail: lobbell.danelle@epa.gov). Reprints can be ordered at <http://www.ajph.org> by clicking the “Reprints/Eprints” link.

This article was accepted February 15, 2011.

Contributors

D. T. Lobbell, J. S. Jagai, K. Rappazzo, and L. C. Messer contributed equally to the writing, reviewing, and editing of this article. D. T. Lobbell conceptualized the overall article. J. S. Jagai reviewed and organized the water domain. K. Rappazzo reviewed and organized the air domain. L. C. Messer reviewed and organized the land, built, and sociodemographic domains.

Acknowledgments

The Office of Research and Development (ORD), US Environmental Protection Agency (EPA), partially funded the research with CSC and L. C. Messer (Contracts WCF DP26H0001 and EP09D000003) with L. C. Messer and under EPA Cooperative Agreement with the University of North Carolina at Chapel Hill (CR83323601).

We are grateful to Barbara Rosenbaum, Suzanne Pierson, Mark Murphy, Tom Luben, Chris Heaney, Jane Gallagher, and Martha Keating for their input and comments on the article and David Hollandsworth for website development.

Note. The views expressed in this article are those of the authors and do not necessarily reflect the views or policies of the US Environmental Protection Agency.

Human Participant Protection

No protocol approval was required; no human data were acquired or used as part of this article.

References

- US Environmental Protection Agency. *EPA's 2008 Report on the Environment*. Washington, DC; 2008.
- US Environmental Protection Agency. *America's Children and the Environment: Measures of Contaminants*. Washington, DC: Body Burdens, and Illnesses.; 2003.
- Centers for Disease Control and Prevention. National Environmental Public Health Tracking Program. Available at: <http://www.cdc.gov/nceh/tracking>. Accessed December 9, 2010.
- Basu R, Woodruff TJ, Parker JD, et al. Comparing exposure metrics in the relationship between PM_{2.5} and birth weight in California. *J Expo Anal Environ Epidemiol*. 2004;14(5):391–396.
- Sagiv SK, Mendola P, Loomis D, et al. A time-series analysis of air pollution and preterm birth in Pennsylvania, 1997–2001. *Environ Health Perspect*. 2005;113(5):602–606.
- Pope CA 3rd, Ezzati M, Dockery DW. Fine-particulate air pollution and life expectancy in the United States. *N Engl J Med*. 2009;360(4):376–386.
- Peng RD, Chang HH, Bell ML, et al. Coarse particulate matter air pollution and hospital admissions for cardiovascular and respiratory diseases among Medicare patients. *JAMA*. 2008;299(18):2172–2179.
- Bell ML, Ebisu K, Peng RD, et al. Hospital admissions and chemical composition of fine particle air pollution. *Am J Respir Crit Care Med*. 2009;179(12):1115–1120.
- Jagai JS, Griffiths JK, Kirshen PH, et al. Patterns of protozoan infections: spatiotemporal associations with cattle density. *EcoHealth*. 2010;7(1):33–46.
- Hitt NP, Hendryx M. Ecological integrity of streams related to human cancer mortality rates. *EcoHealth*. 2010;7(1):91–104.
- Elliott MR, Wang Y, Lowe RA, Kleindorfer PR. Environmental justice: frequency and severity of US chemical industry accidents and the socioeconomic status of surrounding communities. *J Epidemiol Community Health*. 2004;58(1):24–30.
- Griffith M, Tajik M, Wing S. Patterns of agricultural pesticide use in relation to socioeconomic characteristics of the population in the rural U.S. South. *Int J Health Serv*. 2007;37(2):259–277.
- Hendryx M, Fedorko E, Halverson J. Pollution sources and mortality rates across rural-urban areas in the United States. *J Rural Health*. 2010;26(4):383–391.
- Zahran S, Brody SD, Peacock WG, et al. Social vulnerability and the natural and built environment: a model of flood casualties in Texas. *Disasters*. 2008;32(4):537–560.
- Lopez-Zetina J, Lee H, Friis R. The link between obesity and the built environment. Evidence from an ecological analysis of obesity and vehicle miles of travel in California. *Health Place*. 2006;12(4):656–664.
- Cohen SA, Egorov AI, Jagai JS, et al. The SEEDs of two gastrointestinal diseases: socioeconomic, environmental, and demographic factors related to cryptosporidiosis and giardiasis in Massachusetts. *Environ Res*. 2008;108(2):185–191.
- Swan SH, Kruse RL, Liu F, et al. Semen quality in relation to biomarkers of pesticide exposure. *Environ Health Perspect*. 2003;111(12):1478–1484.
- Grady SC. Racial disparities in low birthweight and the contribution of residential segregation: a multilevel analysis. *Soc Sci Med*. 2006;63(12):3013–3029.
- Lupo PJ, Symanski E, Waller DK, et al. Maternal exposure to ambient levels of benzene and neural tube defects among offspring, Texas, 1999–2004. *Environ Health Perspect*. 2011;119(3):397–402.
- Kalkbrenner AE, Daniels JL, Chen JC, et al. Perinatal exposure to hazardous air pollutants and autism spectrum disorders at age 8. *Epidemiology*. 2010;21(5):631–641.
- Borchardt MA, Chyou PH, DeVries EO, Belongia EA. Septic system density and infectious diarrhea in a defined population of children. *Environ Health Perspect*. 2003;111(5):742–748.
- Burkholder J, Libra B, Weyer P, et al. Impacts of waste from concentrated animal feeding operations on water quality. *Environ Health Perspect*. 2007;115(2):308–312.
- Mirabelli MC, Wing S, Marshall SW, Wilcosky TC. Asthma symptoms among adolescents who attend public schools that are located near confined swine feeding operations. *Pediatrics*. 2006;118(1):e66–e75.
- Mirabelli MC, Wing S, Marshall SW, Wilcosky TC. Race, poverty, and potential exposure of middle-school students to air emissions from confined swine feeding operations. *Environ Health Perspect*. 2006;114(4):591–596.
- Wing S, Cole D, Grant G. Environmental injustice in North Carolina's hog industry. *Environ Health Perspect*. 2000;108(3):225–231.
- Wing S, Horton RA, Marshall SW, et al. Air pollution and odor in communities near industrial swine operations. *Environ Health Perspect*. 2008;116(10):1362–1368.
- Wing S, Horton RA, Muhammad N, et al. Integrating epidemiology, education, and organizing for environmental justice: community health effects of industrial hog operations. *Am J Public Health*. 2008;98(8):1390–1397.
- Payne-Sturges D, Gee GC, Crowder K, et al. Workshop summary: connecting social and environmental factors to measure and track environmental health disparities. *Environ Res*. 2006;102(2):146–153.
- Mohai P, Lantz PM, Morenoff J, et al. Racial and socioeconomic disparities in residential proximity to polluting industrial facilities: evidence from the Americans' Changing Lives Study. *Am J Public Health*. 2009;99(Suppl 3):S649–S656.
- Payne-Sturges D, Gee GC. National environmental health measures for minority and low-income populations: tracking social disparities in environmental health. *Environ Res*. 2006;102(2):154–171.
- Fan AM, Alexeeff G, Harris SB. Cumulative risks and cumulative impacts of environmental chemical exposures. *Int J Toxicol*. 2010;29(1):57.
- Martuzzi M, Mitis F, Forastiere F. Inequalities, injustices, environmental factors in waste management and health. *Eur J Public Health*. 2010;20(1):21–26.
- Johnson BL, Coulberson SL. Environmental epidemiologic issues and minority health. *Ann Epidemiol*. 1993;3(2):175–180.
- Norton JM, Wing S, Lipscomb HJ, et al. Race, wealth, and solid waste facilities in North Carolina. *Environ Health Perspect*. 2007;115(9):1344–1350.
- Miranda ML, Maxson P, Edwards S. Environmental contributions to disparities in pregnancy outcomes. *Epidemiol Rev*. 2009;31:67–83.
- Miranda ML, Kim D, Reiter J, et al. Environmental contributors to the achievement gap. *Neurotoxicology*. 2009;30(6):1019–1024.
- Rubin IL, Nodvin JT, Geller RJ, et al. Environmental health disparities: environmental and social impact of

- industrial pollution in a community—the model of Anniston, AL. *Pediatr Clin North Am*. 2007;54(2):375–398 ix.
38. Olden K, White SL. Health-related disparities: influence of environmental factors. *Med Clin North Am*. 2005;89(4):721–738.
39. Larson NI, Story MT, Nelson MC. Neighborhood environments: disparities in access to healthy foods in the U.S. *Am J Prev Med*. 2009;36(1):74–81.
40. Lovasi GS, Hutson MA, Guerra M, Neckerman KM. Built environments and obesity in disadvantaged populations. *Epidemiol Rev*. 2009;31:7–20.
41. Landrigan PJ, Rauh VA, Galvez MP. Environmental justice and the health of children. *Mt Sinai J Med*. 2010;77(2):178–187.
42. Rosen LD, Imus D. Environmental injustice: children's health disparities and the role of the environment. *Explore (NY)*. 2007;3(5):524–528.
43. Stokes SC, Hood DB, Zokovitch J, Close FT. Blueprint for communicating risk and preventing environmental injustice. *J Health Care Poor Underserved*. 2010;21(1):35–52.
44. Bullard RD, Wright BH. Environmental justice for all: community perspectives on health and research needs. *Toxicol Ind Health*. 1993;9(5):821–841.
45. DeFur PL, Evans GW, Cohen Hubal EA, et al. Vulnerability as a function of individual and group resources in cumulative risk assessment. *Environ Health Perspect*. 2007;115(5):817–824.
46. Institute of Medicine—Committee on Environmental Justice. *Toward Environmental Justice: Research, Education, and Health Policy Needs*. Washington, DC: The National Academies Press; 1999.
47. US Environmental Protection Agency. Environmental Quality Index. Available at: <http://www.epa.gov/nheerl/eqi/>. Accessed April 10, 2011.
48. US Environmental Protection Agency. The Ambient Air Monitoring Program. Available at: <http://www.epa.gov/air/oaqps/qa/monprog.html>. Accessed July 15, 2010.
49. US Environmental Protection Agency. National Air Toxics Assessments. Available at: <http://www.epa.gov/ttn/atw/natamain>. Accessed September 10, 2010.
50. US Geological Survey (USGS). National Water Information System (NWIS). Available at: <http://qwwebsiteservices.usgs.gov>. Accessed August 26, 2010.
51. US Environmental Protection Agency. STORET Data Warehouse. Available at: <http://www.epa.gov/storet/>. Accessed April 6, 2011.
52. US Environmental Protection Agency. Watershed Assessment, Tracking and Environmental Results (WATERS). Available at: <http://www.epa.gov/waters>. Accessed April 14, 2011.
53. US Environmental Protection Agency. National Contaminant Occurrence Database (NCO). Available at: <http://water.epa.gov/scitech/datait/databases/drink/nco/databases-index.cfm>. Accessed August 26, 2010.
54. US Environmental Protection Agency. Safe Drinking Water Information System. Available at: <http://water.epa.gov/scitech/datait/databases/drink/sdwsifed>. Accessed August 26, 2010.
55. US Geological Survey. Estimated Use of Water in the United States. Available at: <http://water.usgs.gov/watuse>. Accessed August 26, 2010.
56. National Drought Mitigation Center. Drought Monitor Data Downloads. Available at: http://www.drought.unl.edu/dm/dmshps_archive.htm. Accessed August 26, 2010.
57. National Atmospheric Deposition Program. National Atmospheric Deposition Program. Available at: <http://nadp.sws.uiuc.edu>. Accessed August 26, 2010.
58. Agricultural Research Service. Nutrient Loss Database for Agricultural Fields in the U.S. Available at: <http://www.ars.usda.gov/Research/docs.htm?docid=11079>. Accessed August 26, 2010.
59. United States Geologic Survey. 1997 County Pesticide Use Estimates for 220 compounds. Available at: http://water.usgs.gov/GIS/metadata/usgswrd/XML/pesticide_use97.xml. Accessed August 26, 2010.
60. United States Department of Agriculture. 2002 Census of Agriculture full report. Available at: <http://www.agcensus.usda.gov/Publications/2002/index.asp>. Accessed August 26, 2010.
61. Dun and Bradstreet. Dun and Bradstreet Products. Available at: http://www.dnb.com/us/dbproducts/product_overview/index.html. Accessed August 26, 2010.
62. US Environmental Protection Agency. EPA's web feature service for National Priority List (NPL) sites. Available at: <http://geodata.gov>. Accessed August 26, 2010.
63. US Environmental Protection Agency. Superfund National Priorities List (NPL) Sites. Available at: <http://www.epa.gov/superfund/sites/npl/index.htm>. Accessed April 6, 2011.
64. US Environmental Protection Agency. Resource Conservation and Recovery Act (RCRA) Treatment, Storage, and Disposal Facilities (TSD) and (RCRA) Corrective Action Facilities. Available at: <http://www.epa.gov/osw/hazard/tsd/index.htm>. Accessed August 26, 2010.
65. US Environmental Protection Agency. Resource Conservation and Recovery Act (RCRA) Large Quantity Generators (LQG). Available at: <http://www.epa.gov/osw/hazard/generation/lqg.htm>. Accessed August 26, 2010.
66. US Environmental Protection Agency. Toxics Release Inventory (TRI) Sites. Available at: <http://www.epa.gov/tri>. Accessed August 26, 2010.
67. US Environmental Protection Agency. Assessment, Cleanup, and Redevelopment Exchange (ACRES) Brownfield Sites. Available at: <http://www.epa.gov/brownfields>. Accessed August 26, 2010.
68. US Environmental Protection Agency (EPA). Section Seven Tracking System (SSTS) Pesticide Producing Site Locations. Available at: <http://www.epa.gov/compliance/data/systems/toxics/sstsys.html>. Accessed August 26, 2010.
69. US Geologic Services (USGS). National geochemical survey. Available at: <http://tin.er.usgs.gov/geochem/doc/averages/countydata.htm>. Accessed August 26, 2010.
70. US Environmental Protection Agency. Map of radon zones. Available at: <http://www.epa.gov/radon/zonemap.html>. Accessed August 26, 2010.
71. US Census Bureau. Topologically Integrated Geographic Encoding and Referencing. Available at: <http://www.census.gov/geo/www/tiger>. Accessed August 26, 2010.
72. National Highway Traffic Safety Administration (NHTSA) NCSaAN. Fatality Analysis Reporting System (FARS). Available at: <http://www.nhtsa.gov/people/nca/fars.html>. Accessed August 26, 2010.
73. Rural Health Research Center. Rural Urban Commuting Areas Definition. Available at: <http://depts.washington.edu/uwruca/ruca-download.php>. Accessed August 26, 2010.
74. US Census Bureau. Available at: <http://factfinder.census.gov>. Accessed April 6, 2011.
75. Federal Bureau of Investigation. Uniform Crime Reports. Available at: <http://www.fbi.gov/ucr/ucr.htm>. Accessed August 26, 2010.
76. The Urban Institute. A Guide to Home Mortgage Disclosure Act Data. Available at: http://www.urban.org/uploadedpdf/1001247_hdma.pdf. Accessed September 22, 2010.
77. Whitsel EA, Quibrera P, Smith RL, et al. Accuracy of commercial geocoding: assessment and implications. *Epidemiol Perspect Innov*. 2006;3:8.
78. Hall SA, Kaufman JS, Ricketts TC. Defining urban and rural areas in U.S. epidemiologic studies. *J Urban Health*. 2006;83(2):162–175.
79. Lioy PJ, Isukapalli SS, Trasande L, et al. Using national and local extant data to characterize environmental exposures in the National Children's Study: Queens County, New York. *Environ Health Perspect*. 2009;117(10):1494–1504.