



## PROJECT REPORT

# Identifying Environmental Data, Barriers, and Opportunities for Health Services Research (HSR) and Patient- Centered Outcomes Research (PCOR): An Environmental Scan

---

Prepared for  
The Office of the Assistant Secretary for Planning and Evaluation (ASPE)  
at the U.S. Department of Health and Human Services

By  
NORC at the University of Chicago

September 2022

## Office of the Assistant Secretary for Planning and Evaluation

---

The Assistant Secretary for Planning and Evaluation (ASPE) advises the Secretary of the U.S. Department of Health and Human Services (HHS) on policy development in health, disability, human services, data, and science; and provides advice and analysis on economic policy. ASPE leads special initiatives; coordinates the Department's evaluation, research, and demonstration activities; and manages cross-Department planning activities such as strategic planning, legislative planning, and review of regulations. Integral to this role, ASPE conducts research and evaluation studies; develops policy analyses; and estimates the cost and benefits of policy alternatives under consideration by the Department or Congress.

## The Office of Science and Data Policy

---

The Office of Science and Data Policy is the departmental focal point for policy research, analysis, evaluation, and coordination of department-wide public health science policy and data policy activities and issues. The Office provides authoritative advice and analytical support to the ASPE and departmental leadership on public health science policy and data policy issues and initiatives, coordinates science and data policy issues of interagency scope within HHS and manages interagency initiatives in science policy and data policy. The Office works closely with staff from across the Department on strategic plan development and implementation efforts. The Office also carries out a program of policy research, analysis, evaluation, and data development in these issues.

This report was produced by NORC at the University of Chicago under Contract No. HHSP233201500048I, Task Order No. 75P00121F37021 for the Office of Science and Data Policy.

## Suggested Citation

---

Redman, S., Wagstaff, L., Flanagan, E., & Trevino, M. (2022) *Identifying Environmental Data, Barriers, and Opportunities for Health Services Research (HSR) and Patient-Centered Outcomes Research (PCOR): An Environmental Scan*. Report for the Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health & Human Services.

## CONTRIBUTING AUTHORS

SARAH REDMAN, PHD, MPAFF, NORC  
LAURA WAGSTAFF, MPH, NORC  
ELIZABETH FLANAGAN, MPH, NORC

## ACKNOWLEDGMENTS

GRETCHEN TORRES, MPP, NORC  
JENNIFER SATORIOUS, MSW, NORC  
TARCIA JOHNSON, MHA, LSSB, NORC

## PROJECT OFFICERS AND PROJECT LEADERSHIP

LATEEFAH HUGHES, DRPH, VP, AND PROJECT DIRECTOR, NORC  
SONAL PARASRAMPURIA, PHD, ASPE  
OLUWARANTIMI ADETUNJI, PHD, ASPE  
MARCOS TREVINO, MPH, ASPE



# Table of Contents

---

Introduction .....	5
I. Background .....	5
II. Purpose .....	6
Methodology.....	7
I. Data Source Identification .....	7
II. Data Quality Assessment .....	7
Results.....	8
III. Environmental Data Suitable for HSR and PCOR .....	8
IV. Illustrative HSR Questions.....	10
Discussion.....	13
V. Data-Driven Opportunities .....	13
VI. Barriers and Limitations.....	15
Conclusion.....	15
References .....	17
Appendix A. Federal Data Sources.....	21
Appendix B. Data Quality Assessment.....	28

# Introduction

---

## I. Background

People interact with the natural physical environment on a daily basis: the air they breathe, the land they walk upon, and the water they drink. Unfortunately, many diseases and chronic conditions are caused or worsened by environmental contaminants in the air, water, and soil<sup>1</sup> and socially vulnerable populations are disproportionately affected.<sup>2</sup> Air pollution alone is one of the leading causes of illness and premature death in the United States<sup>3</sup> and contributes to over \$800 billion in health care costs each year.<sup>4</sup> Extreme heat events have been associated with an increase in hospitalizations due to fluid deficiencies, renal failure, urinary tract infections, heat stroke, and mental health conditions, as well as an increase in mortality.<sup>5-7</sup>

As a key indicator of climate change and global warming, the National Oceanic and Atmospheric Association has documented an increase in the frequency, duration, and intensity of heat waves across major cities from the 1960s to 2020.<sup>8</sup> Climate change, marked by rising temperatures, changing precipitation, and extreme weather, affects nearly every aspect of our environment, greatly influencing human health.<sup>9</sup> The Intergovernmental Panel on Climate Change stated that “any increase in global warming is projected to affect human health, with primarily negative consequences.”<sup>10</sup>

While research in this area is ongoing, there are several adverse health outcomes that the U.S. Environmental Protection Agency has found to be most highly associated with detrimental environmental exposures including extreme heat, air pollution, and chemicals.<sup>11</sup>

- **Birth outcomes:** Pregnant people are biologically vulnerable to some environmental exposures because of physiologic and immunologic changes in pregnancy.<sup>12</sup> Heat and air pollution can contribute to maternal, fetal, and newborn risks, including elevated blood pressure and preeclampsia, preterm birth, premature rupture of membranes, infection, low birth weight, and stillbirth.<sup>13</sup> Lead exposure during pregnancy can contribute to adverse birth outcomes including preterm delivery and low birthweight.<sup>14</sup>
- **Cardiovascular disease:** Air pollution is a known risk factor for cardiovascular disease.<sup>15</sup> Studies suggest particulate matter exposure may increase the risk of hospital admissions or emergency department visits for cardiovascular related impacts.<sup>16</sup> Warmer environmental temperatures can also lead to increased risk of hospitalization for heart disease.
- **Respiratory disease:** Air pollution is a risk factor in several respiratory health outcomes, such as onset or exacerbation of asthma and chronic obstructive pulmonary disorder (COPD).<sup>17, 18</sup> Excess heat in the environment can also worsen asthma and COPD.<sup>19, 20</sup>
- **Cancer:** Some environmental exposures are known risk factors for certain types of cancers (e.g., radon and lung cancer, arsenic and several types of cancer) while others may be one of many contributing factors.<sup>21</sup>

- **Infectious disease:** Contaminated food and water are the major environmental sources of gastrointestinal illness.<sup>22</sup> In addition, illnesses such as Lyme disease, Rocky Mountain spotted fever, and West Nile virus, can be contracted from certain ticks and mosquitoes.<sup>9</sup>

High quality environmental health data are critical to understanding and addressing these adverse health outcomes. Many federally collected datasets include measures of health outcomes or environmental exposures, and on occasion, elements of both. These data can be used to answer a multitude of research questions by themselves and when linked to other datasets.

## II. Purpose

As part of the Patient-Centered Outcomes Research Trust Fund (PCORTF) established in 2010, the Assistant Secretary for Planning and Evaluation (ASPE) was tasked with setting a strategic framework and vision for PCOR data infrastructure. Key pillars of this framework include the linkage of clinical and other data for research, including environmental data, and the use of publicly funded data systems for research.<sup>23</sup>

To support the PCORTF strategic framework, the authors conducted this scan to identify environmental data suitable for health services research (HSR), which examines the many factors that impact healthcare access, quality and cost,<sup>24</sup> and patient-centered outcomes research (PCOR), which studies the effectiveness of prevention and treatment options to help patients and caregivers make informed decisions about their care.<sup>23</sup> Additional research was done to uncover examples of and opportunities for data linkage with other U.S. Department of Health and Human Services (HHS) data. The authors identified the following research questions to guide the data source search and review effort:

- Which federal agencies collect environmental data for HSR?
- What are potential HSR questions involving environmental data?
- What is the data quality for federal environmental data appropriate for PCOR?
- What is the status of federal environmental data access, sharing, and linkage activities to other HHS data?

# Methodology

---

## I. Data Source Identification

To identify data sources that could address the research questions, the authors searched federal agency websites for mention of health or environment-related datasets. Federal agencies included the Agency for Healthcare Research and Quality (AHRQ), the Centers for Disease Control and Prevention (CDC), the Centers for Medicare and Medicaid Services (CMS), and the National Institutes of Health (NIH), among others. Additionally, the authors conducted a review of the gray and peer-reviewed literature through Google and Google Scholar to identify resources that discussed the use of federal environmental health data. Search terms included a combination of “federal,” “environment/environmental,” and/or “health/disease/illness.” Inclusion criteria consisted of the following: 1) federal data source, 2) data collected in 2012 or after, 3) included measures either directly associated with environmental health factors or with the potential to link with environmental data to develop relevant insights.

Resulting data sources are listed in [Appendix A](#), which were reviewed with ASPE prior to finalizing. The authors then further reviewed these data sources to identify specific health and environmental measures, direct environment and health connections, and geographic and demographic indicators included within each of the data sources.

## II. Data Quality Assessment

The authors applied the Framework for Data Quality<sup>25</sup> to assess the identified data sources. The framework consists of three domains—utility, objectivity, and integrity—and 11 dimensions within these domains. The Framework for Data Quality includes the following definitions for each domain:

- Utility: “the usefulness of the information to the intended users.”
- Objectivity: “whether the information is accurate, reliable, and unbiased.”
- Integrity: “the maintenance of rigorous scientific standards and protection of information from manipulation or influence as well as unauthorized access or revisions.”

Since this environmental scan only included published federal data sources, and federal agencies are required to issue guidelines on upholding data quality,<sup>26</sup> the data quality assessment focuses on the utility domain and its dimensions: relevance, accessibility, timeliness and punctuality, and granularity. Specifically, the assessment rated the utility of federal environmental datasets for health services research and patient-centered outcomes research.

The authors assigned a value of ‘high,’ ‘moderate,’ or ‘low’ for each utility dimension for every data source identified. To earn a ‘high’ value in a dimension, the data source needed to meet the following operational definitions:

- **Relevance:** the source must include both health and environmental indicators within the dataset. Alternatively, it may include either health or environmental indicators and also allow for linkage with other datasets. This dimension aims to address the following research questions of this environmental scan:
  - What are potential HSR questions involving environmental data?
  - What environmental data are suitable for PCOR?
- **Accessibility:** the source must be in a format that can be easily used by researchers (e.g., with a clean, raw dataset available for download in common formats such as .csv or .sass and/or able to interact with the data on the web browser). This dimension addresses the following research question:
  - What is the status of federal environmental data access, sharing, and linkage activities to other HHS data appropriate for PCOR?
- **Timeliness and punctuality<sup>a</sup>:** the source must include data collected in 2012 or later.
- **Granularity:** the source must include estimates for populations of interest to ASPE (e.g., racial and ethnic minorities, rural populations, and the elderly), specifically at the county level to allow linkage with other data sources.

A full description of the quality assessment and definitions for each value can be found in [Appendix B](#).

## Results

---

The data source review resulted in the identification of a total of twenty federal data sources that focused on either environmental exposures, health outcomes, or both. This section describes results of the quality assessment conducted on these resources as well as the identified HSR research questions and data linkages that can be made with these datasets.

### III. Environmental Data Suitable for HSR and PCOR

Based on our data quality analysis, we found several data sources to be of the highest quality and suitability for HSR and PCOR. These data sources are produced and maintained by CDC, NIH, AHRQ, and the EPA. As described above, the primary component of the data quality review for assessing whether data were suitable for HSR and PCOR was identifying whether the dataset included both health and environmental indicators within the dataset or allowed for linkages with other datasets to provide insight on the connections between the environment and health. We also considered patient-centered outcomes that are available in these datasets and can be assessed for their connection with the

---

<sup>a</sup>Timeliness and punctuality were considered together, given timeliness involves when the data collection occurred, and punctuality describes when the data were released. This data quality review looked only at data that had already been released.

environment. These included health outcomes (e.g., cancer, respiratory disease, etc.) and hospital utilization.

## Centers for Disease Control and Prevention (CDC)

Several relevant data sources can be pulled from CDC networks:

- **The National Health and Nutrition Examination Survey (NHANES).**<sup>b</sup> This dataset includes a variety of up-to-date health and environment measures that can be used by themselves or together with other datasets to examine how the environment may impact health services and outcomes. Questionnaire items include air quality experiences and health condition status (including respiratory, heart, and chronic disease); biological specimens are also collected to determine levels of various contaminants in blood or urine.<sup>27</sup>
- **Adult Blood Lead Epidemiology and Surveillance (ABLES).**<sup>c</sup> Although limited in its topical area, this dataset includes state-level data for blood lead levels in adults. These data could be used in combination with other datasets to assess environmental impacts on health among certain populations and healthcare utilization for lead related health outcomes, among others.<sup>28</sup>
- **National Environmental Public Health Tracking Network.**<sup>d</sup> This dataset includes many health and environmental measures from national, state, and city sources often available at the county-level, making a clear connection between how the environment influences health. Health measures include asthma, cancer, heart disease, heat-related illness, while environmental measures include air quality, biomonitoring, drinking water, pesticide exposure, and more.<sup>29</sup>

## National Institutes of Health (NIH)

The **Surveillance, Epidemiology, and End Results (SEER) Program**<sup>e</sup> is a key NIH-produced data source. SEER collects cancer data from hospital and healthcare registries, which are available at the state level. SEER has been linked with other federal databases including Medicare data to provide detailed information about Medicare beneficiaries with cancer. It also uses Geographic Information System (GIS) methods to explore contextual influences on cancer risk, such as environmental exposures. GIS connects data to maps to help visualize trends through geographic patterns. The data tools available through this program's website allow for the connection of health and environmental measures that can inform HSR.<sup>30</sup>

---

<sup>b</sup> <https://www.cdc.gov/nchs/nhanes/index.htm>

<sup>c</sup> <https://www.cdc.gov/niosh/topics/ables/default.html>

<sup>d</sup> <https://ephtracking.cdc.gov/>

<sup>e</sup> <https://seer.cancer.gov/>

## Environmental Protection Agency (EPA)

**EJScreen: Environmental Justice Screening and Mapping Tool.** <sup>f</sup> EJScreen is a dataset that provides census-block level data on environmental and demographic indicators, as well as a few health indicators, such as life expectancy, heart disease, and asthma. Given its high granularity, this dataset can be used in combination with health and healthcare datasets to answer HSR research questions.<sup>31</sup>

The full results of this data quality review, including ‘high,’ ‘moderate,’ and ‘low’ ratings for each dimension of utility for all data sources, can be found in [Appendix B](#).

## IV. Illustrative HSR Questions

Based on the available data, several illustrative HSR questions were developed that focus on chronic conditions and exposures that can be answered in part using the high-quality datasets identified through this review. Some questions may require a combination of other data sources and literature to fully address. The list of questions included in this section represents a sampling of the types of questions that may be answered with these datasets but is not meant to be exhaustive. Table 1 below organizes these questions by HSR focus (access, quality, and cost) and crosswalks questions with health outcomes and environmental factors found in the identified datasets or hubs. It also highlights additional datasets for potential linking.

---

<sup>f</sup> <https://www.epa.gov/ejscreen>

**Table 1. Illustrative HSR Questions and Related Outcomes**

<i>HSR Focus</i>	<i>HSR Question</i>	<i>Health Outcome(s) Include</i>	<i>Environmental Factor(s) Include</i>	<i>Potentially Applicable Datasets</i>
Access	How does the association between chronic environmental exposures and health outcomes differ by various factors that affect healthcare access among the U.S. population (i.e., geographic area, race/ethnicity, income level, etc.)?	<ul style="list-style-type: none"> <li>- Cancer</li> <li>- Heart disease</li> <li>- Respiratory disease</li> </ul>	<ul style="list-style-type: none"> <li>- Chemical exposure</li> <li>- Air quality</li> <li>- Heat exposure</li> </ul>	<ul style="list-style-type: none"> <li>- NIH NIEHS Disaster Research Response (DR2) Resources</li> <li>- National Environmental Public Health Tracking Network*</li> <li>- Surveillance, Epidemiology, and End Results (SEER) Program*</li> <li>- Behavioral Risk Factor Surveillance System (BRFSS)</li> </ul>
	To what extent is there a connection between access to healthcare and areas with higher concentrations of environmental exposures (e.g., lead in water, proximity to superfund sites, high levels of PM 2.5)?	<ul style="list-style-type: none"> <li>- Medically underserved areas</li> <li>- Hospital utilization data</li> </ul>	<ul style="list-style-type: none"> <li>- Geographic data from large pollutant sources</li> </ul>	<ul style="list-style-type: none"> <li>- EPA EJScreen*</li> <li>- National Health and Nutrition Examination Survey (NHANES)*</li> <li>- Surveillance, Epidemiology, and End Results (SEER) Program*</li> <li>- EPA Superfund Site list<sup>6</sup></li> <li>- Factory emissions data<sup>6</sup></li> </ul>
	To what extent are special focus populations (i.e., racial and ethnic minorities, rural populations) accessing healthcare in routine vs. ambulatory settings for conditions resulting from environmental exposures?	<ul style="list-style-type: none"> <li>- Cancer</li> <li>- Heart disease</li> <li>- Respiratory disease</li> <li>- Hospital utilization data</li> </ul>	<ul style="list-style-type: none"> <li>- Chemical exposure</li> <li>- Air quality</li> </ul>	<ul style="list-style-type: none"> <li>- National Health and Nutrition Examination Survey (NHANES)*</li> <li>- CMS claims data<sup>6</sup></li> </ul>

<b>HSR Focus</b>	<b>HSR Question</b>	<b>Health Outcome(s) Include</b>	<b>Environmental Factor(s) Include</b>	<b>Potentially Applicable Datasets</b>
<i>Quality</i>	What local social and/or environmental factors should health care providers be aware of to provide high-quality, comprehensive care?	<ul style="list-style-type: none"> <li>- Cognition impacts</li> <li>- Mood disorders</li> <li>- Reproductive health effects</li> <li>- Various Noncommunicable Diseases</li> </ul>	<ul style="list-style-type: none"> <li>- Chemical exposure</li> </ul>	<ul style="list-style-type: none"> <li>- Adult Blood Lead Epidemiology and Surveillance (ABLES)*</li> <li>- Work-Related Lung Disease Surveillance System (eWoRLD)</li> <li>- AHRQ Social Determinants of Health Database*</li> <li>- CDC/ATSDR Social Vulnerability Index</li> <li>- U.S. Census Community Resilience Estimates<sup>5</sup></li> </ul>
	What types of biomonitoring (blood, urine) could healthcare providers implement to identify environmental factors affecting health to enable more targeted, quality healthcare services?	<ul style="list-style-type: none"> <li>- Cardiovascular disease</li> <li>- Cancer</li> </ul>	<ul style="list-style-type: none"> <li>- Volatile organic compounds</li> <li>- Personal care products (Bisphenol A (BPA), methyl paraben)</li> <li>- Pesticide metabolites</li> </ul>	<ul style="list-style-type: none"> <li>- National Environmental Public Health Tracking Network*</li> <li>- Adult Blood Lead Epidemiology and Surveillance (ABLES)*</li> <li>- National Health and Nutrition Examination Survey (NHANES)*</li> </ul>
<i>Cost</i>	How are healthcare costs affected by environmental factors?	<ul style="list-style-type: none"> <li>- Asthma</li> <li>- Hospital utilization data</li> </ul>	<ul style="list-style-type: none"> <li>- Pollution/air quality</li> <li>- Biologic allergens</li> </ul>	<ul style="list-style-type: none"> <li>- National Health and Nutrition Examination Survey (NHANES)*</li> <li>- National Environmental Public Health Tracking Network*</li> <li>- CMS claims data<sup>5</sup></li> <li>- Healthcare Cost and Utilization Project (HCUP)<sup>5</sup></li> <li>- Medical Expenditure Panel Survey (MEPS)<sup>5</sup></li> </ul>
	How do healthcare costs differ between geographic areas with high and low amounts of environmental exposures? Do these trends differ by social and economic factors?	<ul style="list-style-type: none"> <li>- Hospital utilization data</li> <li>- Prescription drug spending data</li> </ul>	<ul style="list-style-type: none"> <li>- Water lead levels</li> <li>- Particulate matter</li> <li>- Heat exposure</li> </ul>	<ul style="list-style-type: none"> <li>- National Environmental Public Health Tracking Network*</li> <li>- CMS claims data<sup>5</sup></li> <li>- Healthcare Cost and Utilization Project (HCUP)<sup>5</sup></li> <li>- Medical Expenditure Panel Survey (MEPS)<sup>5</sup></li> <li>- CMS Medicare Part D Prescription Drug Event (PDE) data<sup>5</sup></li> </ul>

\*Indicates resources that authors identified as the highest quality and relevance for PCOR and HSR

<sup>5</sup>Resource not included in accompanying data source list (Appendix A)

## Discussion

---

Although investigation into the connection between environmental exposures and human health is not novel, interest in and availability of high-quality data sources to both measure impacts and identify potential solutions is increasing. Several data source opportunities and challenges noted in the literature are explored below.

### V. Data-Driven Opportunities

#### Prioritize Linkages to Healthcare Utilization and Cost Data

Linking environmental exposure data with healthcare utilization and cost data would provide deeper insights into the impacts of environmental exposures on health outcomes, especially among populations at higher risk for such exposures. CMS datasets—including Medicare fee-for-service, Medicare encounter data, Medicare prescription drug event data, and the Transformed Medicaid Statistical Information System, could be explored and prioritized for any efforts to develop a data linkage protocol. These data cover many of the populations at most risk of environmental exposures, given more than a third of the U.S. population is covered by Medicare, Medicaid, and the Children’s Health Insurance Program. AHRQ data sources also offer the potential to explore utilization and cost information for health services research, including hospital care data from the Healthcare Cost and Utilization Project (HCUP) and patient, provider, and employer data from the Medical Expenditure Panel Survey (MEPS).

#### Investigate Opportunities to Advance the Study of Priority Topics and Populations

Recent legislative actions signal an increased national interest in exploring the impact certain environmental exposures have on health outcomes, suggesting an opportunity for advancing environmental data collection, linkage, and analysis. For example, several provisions outlined in the 2022 Inflation Reduction Act (IRA) provide grants for community-level entities to address health disparities resulting from pollution and climate change through activities such as monitoring air quality, reducing greenhouse gas emissions, and mitigating health risks from extreme heat.<sup>32</sup> The IRA also sets aside substantial funding to support and expand data collection efforts to better understand the disproportionate impact of environmental exposures on certain communities.<sup>33</sup> Collectively, the IRA presents an opportunity to expand the collection and use of environmental health data, and by extension improve the utility of the publicly available datasets aggregated by data repository resources such as EJScreen, the Social Determinants of Health Database, and the Tracking Network.

In 2022, the White House released the Blueprint for Addressing the Maternal Health Crisis outlining goals and action steps to improve maternal health. One of the five goals is to “advance data collection, standardization, harmonization, transparency, and research,” stating that “data and research are foundational” to achieving improved maternal health.<sup>34</sup> According to the Census Bureau, national fertility data are “not well-known and underutilized.”<sup>35</sup> However, pregnant and postpartum people, as

well as newborns, are prominent populations for environmental health services research for a variety of reasons. Currently, Medicaid finances more than 40% of all births in the United States,<sup>36</sup> suggesting that a large proportion of pre- and post-natal health care utilization, cost, and biological sample data are maintained within a consolidated dataset. Additionally, increasing evidence suggests that fetuses are susceptible to environmental exposures, which can lead to low birth weight, preterm birth, and being small for gestational age. The Pregnancy Risk Assessment Monitoring System (PRAMS) was not identified as a priority dataset under the scope of this project given environmental health questions are optional and therefore only captured in select states.<sup>37</sup> However, should CDC decide to move these types of questions to the core questionnaire, vital information about maternal attitudes and experiences related to environmental exposures could be captured across the country. Promisingly, PRAMS has begun to link birth certificate and clinical outcomes data to further PCOR opportunities; a publicly accessible data linkage protocol and data set are forthcoming.<sup>38</sup>

Researchers at the state-level have taken children’s environmental health research a step further by merging not only physical environment and health outcome measures from disparate datasets (e.g., the EPA’s Superfund National Priorities List<sup>39</sup> and an Annie E. Casey Foundation database with county-level percentages of low birthweight births<sup>40</sup>), but also incorporating social determinants of health – such as insurance status or caregivers’ income – to develop a holistic view of vulnerability assessments.<sup>41</sup> Health services research on maternal and child health is increasingly important as recent evidence suggests environmental exposures can impact not only immediate offspring, but subsequent generations as well.<sup>42</sup>

### **Expansion of Clinical Data Collection and Utilization**

Near universal adoption of electronic health records (EHRs) across U.S. healthcare systems presents an opportunity for health services researchers to access a wealth of information related to both care utilization and clinical outcomes. However, health services researchers can face barriers when attempting to access EHR data, and data exchange between health systems remains insufficient; current CDC efforts aim to address these challenges to allow for more PCOR-focused research opportunities.<sup>43</sup> Researchers who have navigated the system were able to leverage EHR data to find that exposure to air particulate matter increased risk of gestational diabetes<sup>44</sup> and chronic kidney disease,<sup>45</sup> and that pediatric tobacco smoke exposure screening and intervention was increased through use of a clinical decision support system.<sup>46</sup> Reducing barriers to access and use of EHR data—including restructuring data collected to support clinical practice so it can be used for research; identifying both clinical and information technology personnel to facilitate EHR data use; and appropriate alignment of biostatistical methods for complex administrative data—could help answer future HSR research questions.<sup>47</sup>

Although data linkages allow inferences to be made among certain populations, additional data on biomarker-identified health outcomes would provide the most direct evidence of environmental health impacts, such as the chemical biomarkers collected each year through NHANES. This effort could be

undertaken through national surveys and/or clinical care. The availability of this data would allow for more targeted HSR and PCOR research questions and improved care for priority populations.

## VI. Barriers and Limitations

Although this scan identified several opportunities for improving environmental data for HSR and PCOR, certain data barriers – ranging from unavailability to incompatibility – pose significant challenges to the advancement of research in this field.

### Limited Environmental Health Datasets

Data on environmental exposures and human health are plentiful but often siloed and therefore are not adequate for PCOR or for answering illustrative HSR questions independently. Datasets that provide measures at the nexus of the two concepts are more difficult to obtain. This review did not find datasets that assess patient-centered outcomes (e.g., specific out-of-pocket costs, child and eldercare costs, time away from work, and others) and their connection with environmental exposures and chronic environmental health conditions. Often, datasets that focus specifically on environmental health target occupational health hazards (i.e., Adult Blood Lead Epidemiology and Surveillance, Work-Related Lung Disease Surveillance System) and have limited applicability to broader populations. Investigators often must create their own connections between datasets that may have disparate reporting measures, timeframes, coding, etc.<sup>48</sup> Similarly, only federally available data sources were examined for this environmental scan. Data at the state, academic, and private levels may help provide some insight into HSR research questions at a more localized level. Having more resources available that capture both human health and environmental exposure data would allow for more powerful analyses and improve the characterization of exposure-response relationships.

### Disparate Data Reporting Requirements

Environmental and health data are collected using very different methods. Environmental agencies are amassing vast amounts of data, largely publicly available; health data are often more difficult to access, stored in outdated databases, and subject to strict privacy standards.<sup>49</sup> Using separate environment and health data sources to identify environmental health outcomes requires linking location-level with individual-level data, which has limited feasibility depending on the availability for the data within each population and geographic level. Health services researchers also cite challenges linking health outcome and healthcare utilization data, given the lack of precision of diagnosis and procedure codes in federal claims data.<sup>50</sup>

## Conclusion

---

This report sought to identify federal sources of environmental data for HSR and PCOR, as well as associated barriers and opportunities for further research. While few individual data sources capture

both environmental exposures and health outcomes, there are substantial opportunities for researchers to leverage data linkage capabilities to elucidate environmental health trends.

The results of this report can inform decision-making for subsequent federal HSR and PCOR activities. This environmental scan can serve as a first step to understanding which federally available data sources that focus on the environment and health can be used for HSR and PCOR. However, more work is needed to understand how to use those environmental datasets to specifically answer PCOR questions. Further research could focus on the determination of strengths and weaknesses of current data infrastructures to identify where evidence is plentiful and could be used to support environmental health HSR and PCOR, and where additional data collection is required to provide impactful linkage resources.

Issues addressed in the barriers section of this report highlight the importance for not only increased efforts of primary data collection specific to environmental health measures, but a concerted effort to define appropriate protocols for data standardization, transformation, and linkage to help ensure data quality for connected datasets. Health services researchers may also consider convening subject matter experts to capture first-hand perspectives on the opportunities and challenges to using federal datasets to answer environmental health research questions.

## References

---

1. Centers for Disease Control and Prevention. (April 22, 2021). *About NCEH*. Retrieved April 29 from <https://www.cdc.gov/nceh/information/about.htm>
2. U.S. Environmental Protection Agency. (2021). *Climate Change and Social Vulnerability in the United States: A Focus on Six Impacts*. [https://www.epa.gov/system/files/documents/2021-09/climate-vulnerability\\_september-2021\\_508.pdf](https://www.epa.gov/system/files/documents/2021-09/climate-vulnerability_september-2021_508.pdf)
3. Lelieveld, J., Klingmüller, K., Pozzer, A., Burnett, R. T., Haines, A., & Ramanathan, V. (2019). Effects of fossil fuel and total anthropogenic emission removal on public health and climate. *Proceedings of the National Academy of Sciences*, 116(15), 7192-7197. <https://doi.org/10.1073/pnas.1819989116>
4. Limaye, V., & DeAlwis, D. (2021). *The Costs of Inaction: The Economic Burden of Fossil Fuels and Climate Change on Health in the United States*. <https://www.nrdc.org/sites/default/files/costs-inaction-burden-health-report.pdf>
5. Bobb, J. F., Obermeyer, Z., Wang, Y., & Dominici, F. (2014). Cause-Specific Risk of Hospital Admission Related to Extreme Heat in Older Adults. *JAMA*, 312(24), 2659-2659. <https://doi.org/10.1001/jama.2014.15715>
6. Sun, S., Weinberger, K. R., Nori-Sarma, A., Spangler, K. R., Sun, Y., Dominici, F., & Wellenius, G. A. (2021). Ambient heat and risks of emergency department visits among adults in the United States: time stratified case crossover study. *BMJ*, e065653-e065653. <https://doi.org/10.1136/bmj-2021-065653>
7. Khatana, S. A. M., Werner, R. M., & Groeneveld, P. W. (2022). Association of Extreme Heat With All-Cause Mortality in the Contiguous US, 2008-2017. *JAMA Network Open*, 5(5), e2212957-e2212957. <https://doi.org/10.1001/jamanetworkopen.2022.12957>
8. U.S. Environmental Protection Agency. (2022, August 1, 2022). *Climate Change Indicators: Heat Waves*. <https://www.epa.gov/climate-indicators/climate-change-indicators-heat-waves>
9. Ebi, K. L., Balbus, J. M., Luber, G., Bole, A., Crimmons, A., Glass, G., Saha, S., Shimamoto, M. M., Trtanj, J., & White-Newsome, J. L. (2018). Human Health. In D. R. Reidmiller, C. W. Avery, D. R. Easterling, K. E. Kunkel, K. L. M. Lewis, T. K. Maycock, & B. C. Stewart (Eds.), *Impacts, Risks, and Adaptation in the United States: Fourth National Climate Assessment* (Vol. II, pp. 539-571). U.S. Global Change Research Program. <https://doi.org/10.7930/NCA4.2018>
10. Intergovernmental Panel on Climate Change. (2022). Summary for Policymakers. In *Global Warming of 1.5°C* (pp. 1-24). Cambridge University Press. <https://doi.org/10.1017/9781009157940.001>
11. U.S. Environmental Protection Agency. (2021). *Report on the Environment*. Retrieved July 15, 2022 from <https://www.epa.gov/report-environment>
12. Laura Rudolph, C. H., Laura Buckley, Savannah North. (2018). *Climate Change, Health, and Equity: A Guide for Local Health Departments*. [https://www.apha.org/-/media/files/pdf/topics/climate/climate\\_health\\_equity.ashx](https://www.apha.org/-/media/files/pdf/topics/climate/climate_health_equity.ashx)
13. Roos, N., Kovats, S., Hajat, S., Filippi, V., Chersich, M., Luchters, S., Scorgie, F., Nakstad, B., Stephansson, O., Hess, J., Kadio, K., Kouanda, S., Lusambili, A., Marsham, J., Ngugi, A., & Wright, C. Y. (2021). Maternal and newborn health risks of climate change: A call for awareness and global action. *Acta Obstetrica et Gynecologica Scandinavica*, 100(4), 566-570. <https://doi.org/10.1111/aogs.14124>
14. Agency for Toxic Substances and Disease Registry. (2020). *Toxicological Profile for Lead*. <https://www.atsdr.cdc.gov/toxprofiles/tp13.pdf>

15. An, Z., Jin, Y., Li, J., Li, W., & Wu, W. (2018). Impact of Particulate Air Pollution on Cardiovascular Health. *Current Allergy and Asthma Reports*, 18(3), 15-15. <https://doi.org/10.1007/s11882-018-0768-8>
16. Brook, R. D., Rajagopalan, S., Pope, C. A., Brook, J. R., Bhatnagar, A., Diez-Roux, A. V., Holguin, F., Hong, Y., Luepker, R. V., Mittleman, M. A., Peters, A., Siscovick, D., Smith, S. C., Whitsett, L., & Kaufman, J. D. (2010). Particulate Matter Air Pollution and Cardiovascular Disease. *Circulation*, 121(21), 2331-2378. <https://doi.org/10.1161/CIR.0b013e3181d8e3e1>
17. Cook, Q., Argenio, K., & Lovinsky-Desir, S. (2021). The impact of environmental injustice and social determinants of health on the role of air pollution in asthma and allergic disease in the United States. *Journal of Allergy and Clinical Immunology*, 148(5), 1089-1101.e1085. <https://doi.org/10.1016/j.jaci.2021.09.018>
18. Hansel, N. N., McCormack, M. C., & Kim, V. (2016). The Effects of Air Pollution and Temperature on COPD. *COPD: Journal of Chronic Obstructive Pulmonary Disease*, 13(3), 372-379. <https://doi.org/10.3109/15412555.2015.1089846>
19. Soneja, S., Jiang, C., Fisher, J., Upperman, C. R., Mitchell, C., & Sapkota, A. (2016). Exposure to extreme heat and precipitation events associated with increased risk of hospitalization for asthma in Maryland, U.S.A. *Environmental Health*, 15(1), 57-57. <https://doi.org/10.1186/s12940-016-0142-z>
20. Witt, C., Schubert, J. A., Jehn, M., Holzgreve, A., Liebers, U., Endlicher, W., & Scherer, D. (2015). The Effects of Climate Change on Patients With Chronic Lung Disease. *Deutsches Ärzteblatt international*. <https://doi.org/10.3238/arztebl.2015.0878>
21. National Cancer Institute. (n.d., June 17, 2022). *Cancer-Causing Substances in the Environment*. <https://www.cancer.gov/about-cancer/causes-prevention/risk/substances>
22. Trtanj, J., Jantarasami, L., Brunkard, J., Collier, T., Jacobs, J., Lipp, E., McLellan, S., Moore, S., Paerl, H., Ravenscroft, J., Sengco, M., & Thurston, J. (2016). Climate Impacts on Water-Related Illness. In U. S. G. C. R. Program (Ed.), *The Impacts of Climate Change on Human Health in the United States: A Scientific Assessment* (pp. 157-188). <http://dx.doi.org/10.7930/J03F4MH>
23. National Academies of Sciences, E., and Math,. (2022). Building Data Capacity for Patient-Centered Outcomes Research: Interim Report 3 - A Comprehensive Ecosystem for PCOR. In T. N. A. Press (Ed.). <https://doi.org/10.17226/26396>
24. Lohr, K. N., & Steinwachs, D. M. (2002). Health Services Research: An Evolving Definition of the Field [<https://doi.org/10.1111/1475-6773.01020>]. *Health Services Research*, 37(1), 15-17. <https://doi.org/https://doi.org/10.1111/1475-6773.01020>
25. Federal Committee on Statistical Methodology. (2020). *A Framework for Data Quality*. Federal Committee on Statistical Methodology.
26. Data.gov. (n.d.). *Data Quality and Documentation*. <https://data.gov/ocean/data-quality-and-documentation-subpage/#data-quality-standards>
27. Centers for Disease Control and Prevention National Center for Health Statistics. (n.d.). National Health and Nutrition Examination Survey Questionnaire. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention. <https://www.cdc.gov/nchs/nhanes/index.htm>
28. Centers for Disease Control and Prevention National Institute for Occupational Safety and Health. (n.d.). Adult Blood Lead Epidemiology and Surveillance (ABLES). Cincinnati, OH. <https://www.cdc.gov/niosh/topics/ables/default.html>
29. Centers for Disease Control and Prevention. (n.d.). National Environmental Public Health Tracking Network. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention. <https://ephtracking.cdc.gov/>

30. National Cancer Institute Division of Cancer Control and Population Sciences. (n.d.). Surveillance, Epidemiology, and End Results (SEER) Program. Bethesda, MD: U.S. Department of Health and Human Services, National Cancer Institute. <https://seer.cancer.gov/>
31. U.S. Environmental Protection Agency. (n.d.). EJScreen: Environmental Justice Screening and Mapping Tool. <https://www.epa.gov/ejscreen>
32. Harvard Law School Environmental & Energy Law Program. (2022, September 25). The Inflation Reduction Act’s Implications for Biden’s Climate and Environmental Justice Priorities. <https://eelp.law.harvard.edu/2022/08/ira-implications-for-climate-ej-priorities/>
33. Safford, H. (2022, August 16). Environmental Data in the Inflation Reduction Act. <https://fas.org/blogs/sciencepolicy/enviro-data-ira/>
34. The White House. (2022). *White House Blueprint for Addressing the Maternal Health Crisis*, . <https://www.whitehouse.gov/wp-content/uploads/2022/06/Maternal-Health-Blueprint.pdf>
35. Genadek, K., Sanders, J., & Stevenson, A. (2022). Measuring US fertility using administrative data from the Census Bureau. *Demographic Research*, 47, 37-58. <https://doi.org/10.4054/DemRes.2022.47.2>
36. Centers for Disease Control and Prevention National Center for Health Statistics. (2022). *National Vital Statistics System – Birth Data*. <https://www.cdc.gov/nchs/nvss/births.htm>
37. Korfmacher, K. S., Suter, B. J., Cai, X., Brownson, S. A., & Dozier, A. M. (2014). Environmental Risks and Children’s Health: What can PRAMS Tell Us? *Maternal and Child Health Journal*, 18(5), 1155-1168. <https://doi.org/10.1007/s10995-013-1345-3>
38. The Assistant Secretary for Planning and Evaluation. (2021). *Enhancing the Data Infrastructure for Women’s Health Research to Improve Women’s and Maternal Health Outcomes*. <https://aspe.hhs.gov/sites/default/files/2021-07/womens-maternal-health.pdf>
39. U.S. Environmental Protection Agency. (2022). *Superfund: National Priorities List*. <https://www.epa.gov/superfund/superfund-national-priorities-list-npl>
40. Annie E Casey Foundation. (n.d., 2022). *Kids Count Data Book*. <https://datacenter.kidscount.org>
41. Hubal, E. A. C., DeLuca, N. M., Mullikin, A., Slover, R., Little, J. C., & Reif, D. M. (2022). Demonstrating a systems approach for integrating disparate data streams to inform decisions on children’s environmental health. *BMC Public Health*, 22(1), 313-313. <https://doi.org/10.1186/s12889-022-12682-3>
42. Voorheis, J. (2021). *Exposure to Pollution Has Long-Term Effect on Multiple Generations*. <https://www.census.gov/library/stories/2021/02/air-pollution-diminishes-future-generations-economic-opportunities.html>
43. Assistant Secretary for Planning and Evaluation. (n.d.). *Making Electronic Health Record (EHR) Data More Available for Research and Public Health*. Retrieved August 15 from <https://aspe.hhs.gov/making-electronic-health-record-ehr-data-more-available-research-public-health>
44. Sun, Y., Li, X., Benmarhnia, T., Chen, J.-C., Avila, C., Sacks, D. A., Chiu, V., Slezak, J., Molitor, J., Getahun, D., & Wu, J. (2022). Exposure to air pollutant mixture and gestational diabetes mellitus in Southern California: Results from electronic health record data of a large pregnancy cohort. *Environment International*, 158, 106888-106888. <https://doi.org/10.1016/j.envint.2021.106888>
45. Ghazi, L., Drawz, P. E., & Berman, J. D. (2022). The association between fine particulate matter (PM2.5) and chronic kidney disease using electronic health record data in urban Minnesota. *Journal of Exposure Science & Environmental Epidemiology*, 32(4), 583-589. <https://doi.org/10.1038/s41370-021-00351-3>
46. Mahabee-Gittens, E. M., Dexheimer, J. W., Tabangin, M., Khoury, J. C., Merianos, A. L., Stone, L., Meyers, G. T., & Gordon, J. S. (2018). An Electronic Health Record–Based Strategy to Address

- Child Tobacco Smoke Exposure. *American Journal of Preventive Medicine*, 54(1), 64-71.  
<https://doi.org/10.1016/j.amepre.2017.08.011>
47. Taksler, G. B., Dalton, J. E., Perzynski, A. T., Rothberg, M. B., Milinovich, A., Krieger, N. I., Dawson, N. V., Roach, M. J., Lewis, M. D., & Einstadter, D. (2021). Opportunities, Pitfalls, and Alternatives in Adapting Electronic Health Records for Health Services Research. *Medical Decision Making*, 41(2), 133-142. <https://doi.org/10.1177/0272989X20954403>
  48. NORC at the University of Chicago, A. H. (2020). *Challenges and Improvements for PCOR Data Infrastructure: Results from a Stakeholder Prioritization Activity*.  
[https://aspe.hhs.gov/sites/default/files/migrated\\_legacy\\_files/197426/PCOR-Data-Infrastructure.pdf](https://aspe.hhs.gov/sites/default/files/migrated_legacy_files/197426/PCOR-Data-Infrastructure.pdf)
  49. Levin, A. (n.d.). *Keeping Track, Promoting Health...* Centers for Disease Control and Prevention,.  
<https://www.cdc.gov/nceh/tracking/pdfs/healthtracks.pdf>
  50. National Cancer Institute Division of Cancer Control & Population Sciences. (n.d., August 2, 2022). *Measures that are Limited or not Available in the Data*.  
<https://healthcaredelivery.cancer.gov/seermedicare/considerations/measures.html>
-

# Appendix A. Federal Data Sources

Agency	Sub-Agency	Source	Link	Description	Health-Related Measures	Environment-Related Measures	Accessibility	Available Years
ACF	OPRE	Early Head Start Family and Child Experiences Study (Baby FACES 2009)	<a href="https://www.acf.hhs.gov/opre/project/early-head-start-family-and-child-experiences-study-baby-faces">https://www.acf.hhs.gov/opre/project/early-head-start-family-and-child-experiences-study-baby-faces</a>	Baby FACES continues a series of ongoing descriptive studies aimed at maintaining an up-to-date, extensive knowledge base to support Early Head Start policies and programs.	Primary caregiver health status (self-reported overall health, access to health insurance, smoking, substance abuse); Early Head Start children health status (premature birth, birth weight, overall health, health care access, immunization status)	Frequency of lead exposure discussed in home visit	The most currently approved documents are accessible online. Access to dataset unknown.	2007-2023
AHRQ	-	Social Determinants of Health (SDOH) Database	<a href="https://www.ahrq.gov/sdoh/data-analytics/sdoh-data.html">https://www.ahrq.gov/sdoh/data-analytics/sdoh-data.html</a>	These SDOH beta data files are curated from existing Federal datasets and other publicly available data sources. The purpose of the files is to make it easier to find a range of well documented, readily linkable SDOH variables across domains without having to access multiple source files, facilitating SDOH research and analysis.	Food access, HPSA (dentists, mental healthcare, primary care), number of clinics and providers, mortality due to cardiovascular disease, drug overdose, alcohol, premature death Emergency and other relief services for victims of domestic or international disasters per 1,000 people Community housing services (targeting low-income or elderly) per 1,000 people; fitness centers and recreational sports centers per 1,000 people; temporary shelters per 1,000 people	Percentage of days with good air quality Annual average ambient concentrations of PM2.5 micrograms per cubic meter Person-days with PM2.5 levels over national ambient air quality standard Percentage of days with PM2.5 levels over national ambient air quality standard Number of designated toxic sites	The AHRQ website has links to data files from 2011 to 2018 in XLSX format. There is an accompanying codebook with all variables in XLSX format. No data viz for the enviro/health measures (only poverty and internet access)	2011-2018
ASPE								
CDC	NCHS	National Health Interview Survey (NHIS)	<a href="https://www.cdc.gov/nchs/nhis/index.htm">https://www.cdc.gov/nchs/nhis/index.htm</a>	The main objective of the NHIS is to monitor the health of the United States population through the collection and analysis of data on a broad range of health topics.	Asthma, food security, cancer, diabetes, cardiovascular conditions, hypertension and cholesterol, immunosuppression, mental health, physical activity, among others	N/A	Through the interactive data query system, can view data tables for variables, grouped by key demographics (sex, race, age, education, MSA). Can also access CSV, ASCII data files, plus SAS, SPSS, and STATA input statements	2004-2020

Agency	Sub-Agency	Source	Link	Description	Health-Related Measures	Environment-Related Measures	Accessibility	Available Years
CDC	NCEH	National Environmental Public Health Tracking Network	<a href="https://ephtracking.cdc.gov/">https://ephtracking.cdc.gov/</a>	The Tracking Network brings together health data and environmental data from national, state, and city sources and provides supporting information to make the data easier to understand. The Tracking Network has data and information on environments and hazards, health effects, and population health.	Asthma, birth defects, cancer, carbon monoxide poisoning, childhood lead poisoning, COPD, developmental disabilities, heart disease, hormone disorders, reproductive and birth outcomes	Air quality, drought, environmental justice, heat & heat-related illness, pesticide exposures, precipitation & flooding, sunlight & UV, Radon testing	Data explorer allows users to select their content area, indicator, and measure of interest to create a data visualization; summary tables also available; raw data download available only for air pollution, solar radiation, and drought	Depends on variable
CDC	-	Behavioral Risk Factor Surveillance System (BRFSS)	<a href="https://www.cdc.gov/brfss">https://www.cdc.gov/brfss</a>	The BRFSS is a telephone survey conducted by all state health departments, the District of Columbia, Puerto Rico, the Virgin Islands, and Guam with assistance from CDC. The BRFSS is the largest continuously conducted telephone health survey in the world. States use BRFSS data to track critical health problems and to develop and evaluate public health programs. The BRFSS is the primary source of information on the health-related behaviors of adults in this country.	General health status, chronic health conditions (asthma, cancer, heart disease); childhood asthma prevalence is an optional module	N/A	Annual data files and codebook available; GIS Maps data files also available; online data visualization tools	1984-2022
CDC	-	Pregnancy Risk Assessment Monitoring System (PRAMS)	<a href="https://www.cdc.gov/prams/index.htm">https://www.cdc.gov/prams/index.htm</a>	Developed in 1987, PRAMS collects site-specific, population-based data on maternal attitudes and experiences before, during, and shortly after pregnancy.	Asthma (maternal)- asked in 21 states Zika	Varies - only six of the 40 states that use PRAMS include any environmental health question Water Quality (Lead, Arsenic)- asked in NH only Lead paint- asked in SC only Mosquitoes Lyme disease- NH only	Application is required to access data. Files sent within 6 weeks of application review. The PRAMS Analytic Research File contains a standard set of variables. There are five categories of the variables provided: birth certificate data, operational variables, weighting variables, questionnaire variables, analytic variables. Code book also available.	1988-2020 (8 phases of data collection/questionnaires)

Agency	Sub-Agency	Source	Link	Description	Health-Related Measures	Environment-Related Measures	Accessibility	Available Years
CDC	NCHS	National Health and Nutrition Examination Survey (NHANES)	<a href="https://www.cdc.gov/nchs/nhanes/index.htm">https://www.cdc.gov/nchs/nhanes/index.htm</a>	NHANES is designed to assess the health and nutritional status of adults and children in the United States. The survey is unique in that it combines interviews and physical examinations. The NHANES program includes a series of cross-sectional nationally representative health examination surveys. The health interview includes questions on demographic characteristics, health insurance status, health and diet behaviors, chronic and acute health conditions, mental health, and prescription medication use. Examination components may vary between survey cycles and generally include body measurements (weight, height, skin folds, body composition scans), blood pressure, and dental exams, and may also include vision, hearing, dermatology, fitness, balance and strength testing, respiratory testing, and taste and smell. Laboratory components may include hematology, organ and endocrine function (e.g., thyroid, kidney), environmental exposure, nutrition biomarkers, cardiovascular and metabolic health, and infectious disease.	Cardiovascular disease, blood pressure, cholesterol, diabetes, diet behavior and nutrition, immunization, kidney conditions, miscellaneous pain, oral health, physical activity and physical fitness, respiratory health and disease, sleep disorders, smoking, alcohol and drug use, mental health, pubertal maturation, prostate conditions, reproductive health	Pesticide exposure, carbon monoxide in blood, volatile toxicants (e.g., paint fumes, cooking with natural gas)	Data for each variable available for download in .XPT format. Codebooks available online. NHANES provides analytic guidance and online tutorials for how to use the data. Users can easily create dashboards on the website.	Currently available through 2020. 2020-2022 data expected shortly.

Agency	Sub-Agency	Source	Link	Description	Health-Related Measures	Environment-Related Measures	Accessibility	Available Years
CDC	NIOSH	National Occupational Mortality Surveillance (NOMS)	<a href="https://www.cdc.gov/niosh/topics/noms/default.html">https://www.cdc.gov/niosh/topics/noms/default.html</a>	NIOSH's NOMS program is a federal-state partnership that monitors changes in cause of death by usual occupation or industry in the United States over time. When repeated at intervals, this is known as occupational mortality surveillance and can identify new associations and opportunities for research and prevention activities. NOMS data can be used to conduct occupational mortality surveillance by calculating proportionate mortality ratios (PMRs). PMRs indicate whether the proportion of deaths due to a specific cause appears to be high or low for a particular occupation, compared to all other occupations, or a particular industry compared to all other industries.	Age-adjusted underlying cause of death proportionate mortality ratios (PMRs). Note: PMRs are calculated by comparing the proportion of deaths from a specified cause within a specified occupation or industry group with the proportion of deaths due to that cause among all decedents and age-adjusts after stratifying on race (White, Black)	Industry and occupation	Can query the data by selecting race, industry or occupation, and cause of death. The query returns the proportionate mortality ratio, deaths, and significance/confidence interval.	1985-1998; 1999; 2003-2004; 2007-2014
CDC	NIOSH	Work-Related Lung Disease Surveillance System (eWoRLD)	<a href="https://wwwn.cdc.gov/e-world/">https://wwwn.cdc.gov/e-world/</a>	eWoRLD provides up-to-date summary tables, figures, and maps on various occupationally-related respiratory diseases, including data on morbidity, mortality, and associated workplace exposures by geographic region, industry and occupation, time period, and demographic group. Diseases include, but are not limited to, the pneumoconioses, asthma and chronic obstructive pulmonary disease, respiratory tuberculosis, and certain cancers.	Pneumoconioses mortality; asthma morbidity; malignant mesothelioma mortality; hypersensitivity pneumonitis; smoking status by occupation and industry	Exposures most frequently reported by work-related asthma cases; environment-related occupations (mining, agriculture)	Downloadable CSV files for data summaries. Full datasets do not appear available for download	Many start in 1990 and go through early 2000s. The variables listed in Column G have data in 2012 or later. The most recent dataset is Percentage of examined underground miners with coal workers' pneumoconiosis from 2017. Many datasets cover 1990-1999.
CDC	NIOSH	Adult Blood Lead Epidemiology and Surveillance (ABLES)	<a href="https://www.cdc.gov/niosh/topics/ABLES/default.html">https://www.cdc.gov/niosh/topics/ABLES/default.html</a>	NIOSH's ABLES program examines U.S. trends in adult workplace lead exposure to identify where exposures occur. ABLES helps state health departments monitor workplace lead exposure trends within their state and find ways to prevent these lead exposures.	N/A	Blood Lead Levels (BLL)	Blood lead level data files available	1994-2016

Agency	Sub-Agency	Source	Link	Description	Health-Related Measures	Environment-Related Measures	Accessibility	Available Years
CDC	ATSDR	Social Vulnerability Index	<a href="https://www.atsdr.cdc.gov/placeandhealth/svi/index.html?msckid=ba829181d09511ec876a6dd9cf6e1a6a">https://www.atsdr.cdc.gov/placeandhealth/svi/index.html?msckid=ba829181d09511ec876a6dd9cf6e1a6a</a>	The CDC/ATSDR Social Vulnerability Index (CDC/ATSDR SVI) uses 15 U.S. census variables to help local officials identify communities that may need support before, during, or after disasters.	N/A	N/A	Downloadable CSV file	2000, 2010, 2014, 2016, 2018
CMS								
DOT								
EPA	-	EJScreen	<a href="https://www.epa.gov/ej-screen">https://www.epa.gov/ej-screen</a>	EJScreen is an environmental justice mapping and screening tool that provides EPA with a nationally consistent dataset and approach for combining environmental and demographic indicators. EJScreen users choose a geographic area; the tool then provides demographic and environmental information for that area. All of the EJScreen indicators are publicly-available data. EJScreen simply provides a way to display this information and includes a method for combining environmental and demographic indicators into EJ indexes.	Low life expectancy, heart disease, asthma	12 environmental indicators including air quality, dust/lead paint, hazardous waste proximity, wastewater discharge; climate change data (wildfire, drought, flooding)	A basic user guide, extensive technical documentation, and data files are available; expert/advanced users can download raw data files	Depends on variable
EPA	-	Consolidated Human Activity Database (CHAD)	<a href="https://www.epa.gov/fera/consolidated-human-activity-database-chad">https://www.epa.gov/fera/consolidated-human-activity-database-chad</a>	The CHAD database includes nearly 180,000 individual study days of detailed human behavior, with each day broken down into individual hours and activity types. The data include demographic information such as age, sex, employment, and education level, which allows researchers to examine specific groups within the general population and how their unique behavior patterns influence their exposures to chemicals.	Various	Various	User guide, technical memo, and latest master data file available for download (in CSV or SAS dataset formats)	1982-2010
FEMA								
HRSA								
HUD								

Agency	Sub-Agency	Source	Link	Description	Health-Related Measures	Environment-Related Measures	Accessibility	Available Years
NASA	-	Multi-Angle Imager for Aerosols (MAIA) mission	<a href="https://maia.jpl.nasa.gov/">https://maia.jpl.nasa.gov/</a>	The MAIA investigation will seek to understand how different types of air pollution affect human health.	N/A	N/A	N/A	N/A
NIH	NIEHS	Chemical Effects in Biological Systems (CEBS) Database	<a href="https://manticore.niehs.nih.gov/cebssearch?mscLinkId=b49c27ccff311ec9d8ee79f4ff7d3b5">https://manticore.niehs.nih.gov/cebssearch?mscLinkId=b49c27ccff311ec9d8ee79f4ff7d3b5</a>	The CEBS database houses data of interest to environmental health scientists. CEBS is a public resource, and has received depositions of data from academic, industrial and governmental laboratories. Data in CEBS are housed in a relational database designed to display data in the context of biology and study design, and permit data integration for cross-study analysis, knowledge generation and novel meta analysis.	N/A	Various chemicals	N/A	N/A
NIH	NIEHS	Comparative Toxicogenomics Database (CTD)	<a href="http://ctdbase.org/">http://ctdbase.org/</a>	The CTD is a robust, publicly available database that aims to advance understanding about how environmental exposures affect human health. It provides manually curated information about chemical-gene/protein interactions, chemical-disease and gene-disease relationships. These data are integrated with functional and pathway data to aid in development of hypotheses about the mechanisms underlying environmentally influenced diseases.	Cardiovascular, endocrine system, musculoskeletal, respiratory tract diseases	Various chemicals	N/A	N/A
NIH	NCI	Surveillance, Epidemiology, and End Results (SEER) Program	<a href="https://seer.cancer.gov/">https://seer.cancer.gov/</a>	The SEER Program provides information on cancer statistics in an effort to reduce the cancer burden among the U.S. population. SEER is supported by the Surveillance Research Program (SRP) in NCI's Division of Cancer Control and Population Sciences (DCCPS).	Cancer	N/A	SEER has two data products available: SEER Research and SEER Research Plus. The Research Plus databases require a more rigorous process for access that includes user authentication through an Institutional Account or a multiple-step request process for Non-Institutional users.	1975-2019

Agency	Sub-Agency	Source	Link	Description	Health-Related Measures	Environment-Related Measures	Accessibility	Available Years
NOAA	NCEI	Data Catalog / OneStop Catalog	<a href="https://data.noaa.gov/datasetsearch/?msckid=962cb12ed0a711ecba4e951e5ad7f041">https://data.noaa.gov/datasetsearch/?msckid=962cb12ed0a711ecba4e951e5ad7f041</a>	The traditional NOAA Data Catalog for all data collections, and the new NOAA OneStop catalog which initially includes only the archived datasets but will eventually replace the traditional catalog.	N/A	Various weather and climate indicators	Depends on dataset	N/A
OCCHE								
SAMHSA								
VA								

# Appendix B. Data Quality Assessment

---

## I. Quality Assessment of the Environmental Data Sources Using the Framework for Data Quality

The Federal Committee on Statistical Methodology (FCSM) established the Framework for Data Quality to assist federal agencies with making decisions about the management of data and to support its appropriate and effective use. The FCSM defines data quality as “the degree to which data capture the desired information using appropriate methodology in a manner that sustains public trust.” The framework consists of three domains—utility, objectivity, and integrity—and 11 dimensions within these domains.

The FCSM provides the following definitions for each domain:

**Utility** reflects the usefulness of the information to the intended users.

**Objectivity** refers to whether the information is accurate, reliable, and unbiased.

**Integrity** refers to the maintenance of rigorous scientific standards and protection of information from manipulation or influence as well as unauthorized access or revision.

Because this environmental scan included only published federal data sources, and federal agencies are required to issue guidelines on upholding data quality<sup>7</sup>, this data quality assessment focuses on the utility domain and its dimensions: relevance, accessibility, timeliness and punctuality, and granularity. Specifically, this assessment rated the utility of federal environmental datasets for health services research and patient-centered outcomes research.

Table 1 below includes the FCSM definitions of each dimension of utility.

**Table 1. Dimensions of the Utility Domain in the Framework for Data Quality**

<b>Relevance</b>	Relevance refers to whether the data product is targeted to meet current and prospective user needs
<b>Accessibility</b>	Accessibility relates to the ease with which data users can obtain an agency’s products and documentation in forms and formats that are understandable to data users
<b>Timeliness</b>	Timeliness is the length of time between the event or phenomenon the data describe and their availability
<b>Punctuality</b>	Punctuality is measured as the time lag between the actual release of the data and the planned target date for data release

---

<sup>7</sup> Data.gov. Data Quality and Documentation. Retrieved from: <https://data.gov/ocean/data-quality-and-documentation-subpage/#data-quality-standards>

## Granularity

Granularity refers to the amount of disaggregation available for key data elements. Granularity can be expressed in units of time, level of geographic detail available, or the amount of detail available on any number of characteristics (e.g., demographic, socio-economic)

Federal Committee on Statistical Methodology. 2020. FCSM: A Framework for Data Quality. *Table ES1. Dimensions of Data Quality*, p.4. Retrieved from:

[https://nces.ed.gov/fcsm/pdf/FCSM.20.04\\_A\\_Framework\\_for\\_Data\\_Quality.pdf](https://nces.ed.gov/fcsm/pdf/FCSM.20.04_A_Framework_for_Data_Quality.pdf)

To be considered high-quality for the purposes of this e-scan, the data must be:

*Relevant*, or include both health and environmental indicators within the dataset. It may also be considered relevant if it includes either health or environmental indicators and allow for linkage with other datasets. This dimension aims to address the following research questions of this environmental scan:

- What are the prioritized HSR questions involving environmental data?
- What environmental data is suitable for PCOR?

*Accessible*, or in a format that can be easily used by NORC, ASPE, and other researchers. This dimension addresses the following research question:

- What is the status of federal environmental data access, sharing, and linkage activities to other HHS data appropriate for PCOR?
- Possible example: air quality, blood lead levels/exposure, heat-related illness

*Timely and punctual*, or collected since 2012.

*Granular*, or able to obtain estimates for populations of interest to ASPE, specifically at the state or county levels to allow linkage with other data sources.

Based on our analysis, we have determined the following data sources to be of highest utility for ASPE's goal to answer select HSR and PCOR research questions:

**The National Health and Nutrition Examination Survey (NHANES), CDC.** This dataset includes a variety of up-to-date health and environment measures that can be used by themselves or together with other datasets to examine how the environment may impact health services and outcomes.

**Adult Blood Lead Epidemiology and Surveillance (ABLES), CDC.** Although limited in its topical area, this dataset includes state-level data for blood lead levels in adults. ASPE could use this data combined with other datasets to assess environmental impacts on health.

**Surveillance, Epidemiology, and End Results (SEER) Program, NIH.** SEER collects cancer data from hospital and healthcare registries, which are made available at the state level. SEER has been linked with other federal databases, including Medicare data to provide detailed information about Medicare beneficiaries with cancer. It also uses GIS methods to explore contextual influences on cancer risk, such as environmental exposures. The data tools available through this program's website allow for the connection of health and environmental measures that can inform HSR.

- **Social Determinants of Health Database, AHRQ.** This up-to-date and highly accessible dataset includes several environmental measures that impact health, such as air quality and particulate matter. Data are provided at the county-level, allowing ASPE and other users to link it to other datasets in order to assess HSR questions.
- **National Environmental Public Health Tracking Network, CDC.** This dataset includes many health and environment measures, making a clear connection between how the environment influences health.
- **EJScreen, EPA.** This dataset provides census-block level data on environmental and demographic indicators, as well as a few health indicators. Given its high granularity, ASPE can use this dataset in combination with health and healthcare datasets to answer HSR research questions.

The tables below describe all data sources that were included in our analysis as well as our assessment of each source for its utility for ASPE. *Table 2* focuses on single datasets while *Table 3* covers federal data hubs that include the compilation of multiple datasets.

Table 2. Federal Databases

Source + Agency	Description	Relevance <sup>1</sup> as an Independent Data Source	Relevance if Linked with Other Data Sources <sup>2</sup>	Accessibility <sup>3</sup>	Timeliness and Punctuality <sup>4</sup>	Granularity <sup>5</sup>
<p><b><u>Early Head Start Family and Child Experiences Study (Baby FACES 2009)</u></b></p>	<p>Baby FACES continues a series of ongoing descriptive studies aimed at maintaining an up-to-date, extensive knowledge base to support Early Head Start policies and programs.</p>	<p> Low</p> <p>The dataset includes primary caregiver health status variables and the health status of early head start children. Specific measures include self-reported overall health, access to health insurance, smoking, premature birth, birth weight, and immunization status.</p> <p>The only environmental data is whether lead exposure in the home was discussed during a home visit.</p>	<p> Low</p> <p>The low granularity level of this dataset will make it difficult to link with other data sources.</p>	<p> Moderate</p> <p>Data available through the Child and Family Data Archive, and can search by data format, time period, and restriction. There are no online summaries or analysis capabilities.</p>	<p> High</p> <p>Data are available through 2018, with 2022 data expected in 2023.</p>	<p> Low</p> <p>The 2018 Baby FACES summary notes that the smallest geographic unit is the United States.<sup>8</sup></p>

<sup>8</sup> <https://www.childandfamilydataarchive.org/cfda/archives/cfda/studies/37666/summary>

Source + Agency	Description	Relevance <sup>1</sup> as an Independent Data Source	Relevance if Linked with Other Data Sources <sup>2</sup>	Accessibility <sup>3</sup>	Timeliness and Punctuality <sup>4</sup>	Granularity <sup>5</sup>
<b>National Health Interview Survey (NHIS), CDC</b>	Administered through the Center for Disease Control and Prevention's National Center for Health Statistics, the NHIS monitors a broad range of health topics among the U.S. population, collected through personal household interviews.	 Low The dataset includes many health topics. Those most relevant to the environment include asthma and food security. There are no environment-related measures included in the NHIS dataset.	 Moderate This data could be linked with environmental data sources at the regional level.	 High Through the interactive data query system, user can view data tables for variables, grouped by key demographics (e.g., sex, race, age, education, MSA) and can also access CSV, ASCII data files, plus SAS, SPSS, and STATA input statements.	 High Data are available from 2004-2020.	 Moderate Data are available at the regional level.
<b>Behavioral Risk Factor Surveillance System (BRFSS), CDC</b>	With assistance from CDC, all state health departments, the District of Columbia, Puerto Rico, the Virgin Islands, and Guam collect this telephone survey to track critical health problems, which informs the development and evaluation of public health programs. The BRFSS is the primary source of information on the health-related behaviors of adults in the United States.	 Low The BRFSS focuses on health. Some health conditions covered that are connected to the environment include general health status, chronic health conditions such as asthma and heart disease, and cancer. Childhood asthma prevalence is an optional module. There are no environment-specific variables included in the survey.	 Moderate The environment-related health indicators as well as the high accessibility and granularity of this dataset makes it possible to link with other environmental or health services data sources.	 High Annual data files and codebook are available for download. GIS maps and online data visualization tools are also available.	 High Data are available from 1984-2022.	 High Data are available at the city and county levels.

Source + Agency	Description	Relevance <sup>1</sup> as an Independent Data Source	Relevance if Linked with Other Data Sources <sup>2</sup>	Accessibility <sup>3</sup>	Timeliness and Punctuality <sup>4</sup>	Granularity <sup>5</sup>
<b><u>Pregnancy Risk Assessment Monitoring System (PRAMS), CDC</u></b>	Developed in 1987, PRAMS collects site-specific, population-based data on maternal attitudes and experiences before, during, and shortly after pregnancy.	 <p>While some environmental health conditions (e.g., asthma) and exposures (e.g., lead and mercury) are asked about on PRAMS, they are not consistently measured and only asked about in certain states.</p>	 <p>Because PRAMS only asks about environmental health conditions in 21 states, it will be minimally relevant even when linked with other sources.</p>	 <p>Users must apply to access this data and will receive the files within 6 weeks of their application being reviewed.</p>	 <p>Data is available from 1988-2020.</p>	 <p>Granularity varies by year. Data is available at the national level while state-level data is only available if it meets sample size thresholds.</p>

Source + Agency	Description	Relevance <sup>1</sup> as an Independent Data Source	Relevance if Linked with Other Data Sources <sup>2</sup>	Accessibility <sup>3</sup>	Timeliness and Punctuality <sup>4</sup>	Granularity <sup>5</sup>
<p><b><u>National Health and Nutrition Examination Survey (NHANES), CDC</u></b></p>	<p>NHANES assesses the health and nutritional status of adults and children across the United States through a series of cross-sectional nationally representative health examination surveys that use a combination of interviews and physical examinations.</p>	<p style="text-align: center;"> Moderate</p> <p>Many health issues are covered in the survey, including cardiovascular disease, blood pressure, cholesterol, diabetes, nutrition, immunization, kidney conditions, pain, oral health, physical activity, respiratory health, sleep disorders, smoking, alcohol and drug use, mental health, pubertal maturation, and reproductive health.</p> <p>Additionally, the following environmental factors are included: pesticide exposure, carbon monoxide in blood, and volatile toxicants (e.g., paint fumes, cooking with natural gas).</p>	<p style="text-align: center;"> Moderate</p> <p>Researchers can apply to gain access to the restricted variables, that include geocodes to allow linking at the state-level.</p>	<p style="text-align: center;"> High</p> <p>Data for each variable is available for download in XPT format. Codebooks are available online. NHANES provides analytic guidance and online tutorials for how to use the data and users can easily create dashboards on the website.</p>	<p style="text-align: center;"> High</p> <p>Data is currently available through 2020 and 2020-2022 data expected shortly.</p>	<p style="text-align: center;"> Low</p> <p>Data below the national level are restricted.</p>

Source + Agency	Description	Relevance <sup>1</sup> as an Independent Data Source	Relevance if Linked with Other Data Sources <sup>2</sup>	Accessibility <sup>3</sup>	Timeliness and Punctuality <sup>4</sup>	Granularity <sup>5</sup>
<p><b><u>National Occupational Mortality Surveillance (NOMS)</u></b></p>	<p>NOMS is a federal-state partnership that monitors changes in cause of death by usual occupation or industry in the United States over time. NOMS data are collected from state vital records offices and can be used to conduct occupational mortality surveillance by calculating proportionate mortality ratios (PMRs), which indicate whether the proportion of deaths due to a specific cause appears to be high or low for a particular occupation or industry.</p>	<div data-bbox="690 237 806 350" style="text-align: center;">  <p>Low</p> </div> <p>Health-related measures include age-adjusted underlying cause of death PMRs* while the only somewhat environment-related measures include industry and occupation.</p> <p>*Note: PMRs are calculated by comparing the proportion of deaths from a specified cause within a specified occupation or industry group with the proportion of deaths due to that cause among all decedents and age-adjusts after stratifying on race.</p>	<div data-bbox="1005 237 1121 350" style="text-align: center;">  <p>Low</p> </div> <p>The low granularity of this data does not allow for linking usable data.</p>	<div data-bbox="1283 237 1398 350" style="text-align: center;">  <p>Moderate</p> </div> <p>Users can query the data by selecting race, industry or occupation, and cause of death. The query returns the proportionate mortality ratio, deaths, and significance/CI.</p>	<div data-bbox="1560 237 1675 350" style="text-align: center;">  <p>High</p> </div> <p>Data are available for the following years: 1985-1998; 1999; 2003-2004; 2007-2014.</p>	<div data-bbox="1818 237 1934 350" style="text-align: center;">  <p>Low</p> </div> <p>Geographic level is not available through the query system.</p>

Source + Agency	Description	Relevance <sup>1</sup> as an Independent Data Source	Relevance if Linked with Other Data Sources <sup>2</sup>	Accessibility <sup>3</sup>	Timeliness and Punctuality <sup>4</sup>	Granularity <sup>5</sup>
<p><b><u>Work-Related Lung Disease Surveillance System (eWoRLD)</u></b></p>	<p>eWoRLD provides up-to-date summary tables, figures, and maps on various occupationally-related respiratory diseases, including data on morbidity, mortality, and associated workplace exposures by geographic region, industry and occupation, time period, and demographic group.</p>	<p style="text-align: center;"> Moderate</p> <p>Health-related measures include pneumoconiosis mortality, asthma morbidity, malignant mesothelioma mortality, hypersensitivity pneumonitis, and smoking status by occupation and industry.</p> <p>Environmental measures include exposures most frequently reported by work-related asthma cases and environment-related occupations (mining, agriculture). Additionally, coal miner’s pneumoconiosis is covered, drawing the connection between environment and health.</p>	<p style="text-align: center;"> Moderate</p> <p>Some variables in this data may be linked with other data sets at the state and county level to further explore connections between environment, health, and health services.</p>	<p style="text-align: center;"> Low</p> <p>Downloadable CSV files are available for data summaries only.</p>	<p style="text-align: center;"> Moderate</p> <p>Most datasets are available from 1990-1999. Some variables, including those listed in the relevance column, have data after 2012. The most recent dataset is percentage of examined underground miners with coal workers' pneumoconiosis from 2017.</p>	<p style="text-align: center;"> Moderate</p> <p>Most data are at the national level, but some variables are available at the state and county levels.</p>

Source + Agency	Description	Relevance <sup>1</sup> as an Independent Data Source	Relevance if Linked with Other Data Sources <sup>2</sup>	Accessibility <sup>3</sup>	Timeliness and Punctuality <sup>4</sup>	Granularity <sup>5</sup>
<b><u>Adult Blood Lead Epidemiology and Surveillance (ABLES)</u></b>	ABLES examines U.S. trends in adult workplace lead exposure to identify where exposures occur. ABLES helps state health departments monitor workplace lead exposure trends within their state and find ways to prevent these lead exposures.	 <p>Moderate</p> <p>ABLES includes adult blood lead levels collected from laboratories and physicians through mandatory reporting.</p>	 <p>Moderate</p> <p>Data can be linked with environmental datasets at the state-level to identify connections between environmental factors and blood lead levels.</p>	 <p>Moderate</p> <p>Data summary tables are available for download in .xlsx format.</p>	 <p>Moderate</p> <p>Data is available for most years from 1994 to 2016. Data is only available for 26 states in 2016.</p>	 <p>Moderate</p> <p>Data is reported at the state-level.</p>
<b><u>Multi-Angle Imager for Aerosols (MAIA) Mission, NASA</u></b>	The MAIA investigation will seek to understand how different types of air pollution affect human health. It's the first time NASA has partnered with epidemiologists and health organizations on a satellite mission to study human health and improve lives.	 <p>High</p> <p>MAIA will view a set of Primary Target Areas from space and combine the data gathered with other information, such as measurements of air pollution, to create daily maps of airborne particulate matter (PM) amounts in the Primary Target Areas. Epidemiologists will then use the data to conduct health studies to understand which types of PM are most harmful.</p>	Not able to assess: Data from this instrument are not available yet.	Not able to assess: Data from this instrument are not yet available.	Not able to assess: Data from this instrument are not yet available.	Not able to assess. Data from this instrument are not yet available.

Source + Agency	Description	Relevance <sup>1</sup> as an Independent Data Source	Relevance if Linked with Other Data Sources <sup>2</sup>	Accessibility <sup>3</sup>	Timeliness and Punctuality <sup>4</sup>	Granularity <sup>5</sup>
<p><b><u>Chemical Effects in Biological Systems (CEBS) Database, NIH</u></b></p>	<p>The CEBS database houses data of interest to environmental health scientists. CEBS has received depositions of data from academic, industrial and government laboratories. Data in CEBS are housed in a relational database designed to display data in the context of biology and study design, and permit data integration for cross-study analysis.</p>	<p style="text-align: center;"> Low</p> <p>CEBS includes chemical and biological studies, rather than environmental and population health measures.</p>	<p style="text-align: center;"> Low</p> <p>Data available through CEBS would be difficult to link to other datasets for the purpose of this project.</p>	<p style="text-align: center;"> Low</p> <p>Some data is publicly available for download. It is unclear in what format. There are no summaries or visualizations available online.</p>	<p style="text-align: center;"> Low</p> <p>The years of accessible data are not clear on their website.</p>	<p style="text-align: center;"> Low</p> <p>Because this does not include population environment or health measures, the geographic granularity is not clear.</p>

Source + Agency	Description	Relevance <sup>1</sup> as an Independent Data Source	Relevance if Linked with Other Data Sources <sup>2</sup>	Accessibility <sup>3</sup>	Timeliness and Punctuality <sup>4</sup>	Granularity <sup>5</sup>
<p><b><u>Comparative Toxicogenomics Database (CTD), NIH</u></b></p>	<p>CTD aims to advance understanding about how environmental exposures affect human health. It provides manually curated information about chemical–gene/protein interactions, chemical–disease and gene–disease relationships. These data are integrated with functional and pathway data to aid in development of hypotheses about the mechanisms underlying environmentally influenced diseases.</p>	<p> Moderate</p> <p>CTD covers cardiovascular, endocrine, musculo-skeletal, and respiratory tract diseases and includes data on various chemicals as they relate to these conditions. For example, in their “environmental illness” category, they connect vehicle emissions and tobacco smoke pollution to gene A2M, noting how each of these chemicals alters the expression of this gene.</p>	<p> Low</p> <p>Data available through CTD would be difficult to connect to other datasets.</p>	<p> Moderate</p> <p>The datanase allows you to search by chemicals, diseases, organisms and genes as well as query the data based on a variety of factors.</p>	<p> Moderate</p> <p>Data has been updated in 2022, but it is not clear when the data was collected.</p>	<p>Not able to assess: data does not specify geographic granularity.</p>

Source + Agency	Description	Relevance <sup>1</sup> as an Independent Data Source	Relevance if Linked with Other Data Sources <sup>2</sup>	Accessibility <sup>3</sup>	Timeliness and Punctuality <sup>4</sup>	Granularity <sup>5</sup>
<p><b><u>Surveillance, Epidemiology, and End Results (SEER) Program, NIH</u></b></p>	<p>SEER provides information on cancer statistics in an effort to reduce the cancer burden among the U.S. population.</p>	<p style="text-align: center;"> Moderate</p> <p>SEER includes data on cancer, specifically its registries routinely collect data on patient demographics, primary tumor site, tumor morphology and stage at diagnosis, first course of treatment, and follow-up for vital status.</p> <p>Additionally, SEER uses GIS methods to explore contextual influences on cancer risk, such as environmental exposures.</p>	<p style="text-align: center;"> High</p> <p>SEER has been linked with other federal databases, including Medicare data and Consumer Assessment of Healthcare Providers and Systems.</p>	<p style="text-align: center;"> High</p> <p>SEER has two data products available: SEER Research and SEER Research Plus. Research Plus databases require user authentication to access through an Institutional Account or a multiple-step request process. Numerous data query and interactive tools are also available.</p>	<p style="text-align: center;"> High</p> <p>Data is available through 2019. The specific years of available data vary between states, starting as early as 1975 and as late as 2000.</p>	<p style="text-align: center;"> Moderate</p> <p>Data is collected from registries at hospitals and health systems. Summaries are available at the state-level.</p>

Source + Agency	Description	Relevance <sup>1</sup> as an Independent Data Source	Relevance if Linked with Other Data Sources <sup>2</sup>	Accessibility <sup>3</sup>	Timeliness and Punctuality <sup>4</sup>	Granularity <sup>5</sup>
<b><u>Environmental Agents Service (EAS) Registry System of Records, VA</u></b>	The Environmental Agent Service (EAS) Registries is the information system encompassing the Ionizing Radiation Registry (IRR), the Agent Orange Registry (AOR), and the Gulf War Registry (GWR) which also includes related Depleted Uranium (DU) exams.	 <p>The IRR contains results of questionnaires completed by Veterans who may have been exposed to ionizing radiation while on active military duty and have had an IRR examination at a VA medical facility.<sup>6</sup></p> <p>The AOR contains results of the VA's Agent Orange Registry health exam, which alerts Veterans to possible long-term health problems that may be related to Agent Orange exposure during their military service.</p> <p>Similarly, the GWR monitors health conditions among veterans that may be related to environmental exposures in the Persian Gulf Region.</p>	Not able to assess: the data is not accessible and therefore it is difficult to understand how it could be linked with other data sources.	 <p>This registry is not publicly available. To access the data, users need to contact the dataset owner.</p>	 <p>Data is available for 1998-2014</p>	Not able to assess: the data are not accessible and therefore we are unable to assess granularity.

<sup>1</sup> High relevance: dataset makes the connection between environmental and health measures; moderate relevance: dataset includes both environment and health measures; low relevance: dataset only includes health or environment measures

<sup>2</sup> High linkage relevance: dataset has been linked with other datasets or has multiple health and/or environmental measures that could be linked with other datasets at a moderate or high granularity<sup>4</sup> level; moderate linkage relevance: dataset offers at least one health and/or environmental measure that can be linked with other datasets at a moderate or high granularity level; low linkage relevance: the dataset is unable to be linked at a moderate or high granularity level or it is unclear how it will be linked with other datasets to answer research questions related to health and the environment.

<sup>3</sup> High accessibility: all data is available for download in XLSX or other common format, codebook is available, and users can view summaries or interact with the data online without download; moderate accessibility: data is available for download, but there is limited interaction or summaries available; low accessibility: data is not available for download or only some variables are available for download

<sup>4</sup> High timeliness and punctuality: data is available for years 2012 and later; moderate timeliness and punctuality: some variables are available for years 2012 and later, but not all; low timeliness and punctuality: data is not available after 2012.

<sup>5</sup> High granularity: available at the county level and lower; moderate granularity: available at the state level; low granularity: available at the national level

<sup>6</sup> <https://www.data.va.gov/dataset/Ionizing-Radiation-Registry-IRR-The-Environmental-/anm5-vwhs>

Table 3. Federal Data Hubs

Source + Agency	Description	Relevance as an Independent Data Source <sup>1</sup>	Relevance if Linked with Other Data Sources <sup>2</sup>	Accessibility <sup>3</sup>	Timeliness and Punctuality <sup>4</sup>	Granularity <sup>5</sup>
<p><b><u>Social Determinants of Health Database, AHRQ</u></b></p>	<p>These SDOH beta data files are curated from existing Federal datasets and other publicly available data sources. The purpose of the files is to make it easier to find a range of well documented, readily linkable SDOH variables across domains without having to access multiple source files, facilitating SDOH research and analysis.</p>	<div style="text-align: center;">  <p>Moderate</p> </div> <p>The database includes the following variables related to the environment and health:</p> <ul style="list-style-type: none"> <li>• Percentage of days with good air quality</li> <li>• Annual average ambient concentrations of PM2.5 micrograms per cubic meter</li> <li>• Person-days with PM2.5 levels over national ambient air quality standard</li> <li>• Percentage of days with PM2.5 levels over national ambient air quality standard</li> </ul>	<div style="text-align: center;">  <p>High</p> </div> <p>This database contains its own data linkages from existing publicly-available data sources. It also can be linked to other health or environmental datasets at the county-level.</p>	<div style="text-align: center;">  <p>High</p> </div> <p>The AHRQ website has links to data files from 2011 to 2018 in XLSX format. There is an accompanying Codebook with all variables in XLSX format.</p>	<div style="text-align: center;">  <p>High</p> </div> <p>The database includes data collected between 2011 and 2018.</p>	<div style="text-align: center;">  <p>High</p> </div> <p>Data are reported at the county level.</p>

<p><b>National Environmental Public Health Tracking Network, CDC</b></p>	<p>The CDC’s Environmental Public Health Tracking Network (Tracking Network) brings together health data and environmental data from national, state, and city sources and provides supporting information to make the data easier to understand. The Tracking Network has data and information on environments and hazards, health effects, and population health.</p>	<div style="text-align: center;">  <p>High</p> </div> <p>Health-related variables include:</p> <ul style="list-style-type: none"> <li>• Asthma</li> <li>• Birth defects</li> <li>• Cancer</li> <li>• Carbon monoxide poisoning</li> <li>• Childhood lead poisoning</li> <li>• COPD</li> <li>• Developmental disabilities</li> <li>• Heart disease</li> <li>• Hormone disorders</li> <li>• Reproductive and birth outcomes</li> </ul> <p>Environment-related variables include:</p> <ul style="list-style-type: none"> <li>• Air quality</li> <li>• Drought</li> <li>• Environmental justice</li> <li>• Pesticide exposures</li> <li>• Precipitation &amp; flooding</li> <li>• Sunlight &amp; UV</li> <li>• Radon testing</li> </ul> <p>The data hub also includes variables that make a clear connection between the environment and health, including levels of environmental chemicals in blood and urine, such as tobacco smoke, metals, volatile organic compounds, chloroform, PFAs, chemicals found in personal care products such as BPA and methyl paraben, pesticide metabolites, hydrocarbons, and phthalate metabolites. Additionally, the hub covers children’s environmental health (e.g., lead poisoning) and heat-related ED visits, hospitalizations, and mortality.</p>	<div style="text-align: center;">  <p>High</p> </div> <p>The Tracking Network already makes connections between multiple datasets to drive its data explorer. Users can also connect the data files for select variables with other datasets on health services or the environment.</p>	<div style="text-align: center;">  <p>Moderate</p> </div> <p>Through the data explorer, users can select their content area, indicator, and measure of interest to create a data visualization and summary tables are also available. Raw data file download is available only for air pollution, solar radiation, and drought.</p>	<div style="text-align: center;">  <p>High</p> </div> <p>Most variables are available through 2020.</p>	<div style="text-align: center;">  <p>High</p> </div> <p>Almost all variables are available at the national, state, and county level, while some are also available at the census-tract level.</p>
--	---	---	--	--	--	---

Source + Agency	Description	Relevance as an Independent Data Source <sup>1</sup>	Relevance if Linked with Other Data Sources <sup>2</sup>	Accessibility <sup>3</sup>	Timeliness and Punctuality <sup>4</sup>	Granularity <sup>5</sup>
<p><b><u>Social Vulnerability Index, CDC</u></b></p>	<p>The Social Vulnerability Index uses 15 U.S. census variables to help local officials identify communities that may need support before, during, or after disasters.</p>	<p style="text-align: center;"> Low</p> <p>The Social Vulnerability Index does not include specific health or environment-related measures.</p>	<p style="text-align: center;"> Moderate</p> <p>This database’s high granularity may allow it to be combined with environment and health datasets to identify social vulnerability factors as they relate to environmental health and health outcomes.</p>	<p style="text-align: center;"> High</p> <p>Data is available through a downloadable .csv file and their site also includes an interactive map users can use to explore the data.</p>	<p style="text-align: center;"> High</p> <p>Data are available for 2000, 2010, 2014, 2016, and 2018.</p>	<p style="text-align: center;"> High</p> <p>Data is available for the census-tract, county, state-level.</p>

Source + Agency	Description	Relevance as an Independent Data Source <sup>1</sup>	Relevance if Linked with Other Data Sources <sup>2</sup>	Accessibility <sup>3</sup>	Timeliness and Punctuality <sup>4</sup>	Granularity <sup>5</sup>
<b>EJScreen, EPA</b>	EJScreen is an environmental justice mapping and screening tool that provides EPA with a nationally consistent dataset and approach for combining environmental and demographic indicators. EJScreen users choose a geographic area; the tool then provides demographic and environmental information for that area. All of the EJScreen indicators are publicly-available data. EJScreen provides a way to display this information and includes a method for combining environmental and demographic indicators into EJ indexes.	 <p>Health-related measures included in the EJScreen tool include low life expectancy, heart disease, and asthma.</p> <p>EJ screen covers 12 environmental indicators including air quality, dust/lead paint, hazardous waste proximity, wastewater discharge; and climate change data (wildfire, drought, flooding).</p> <p>On the map, users can overlay environmental, health, and sociodemographic factors.</p>	 <p>EJScreen already makes some connections between datasets and the high granularity could allow it to be combined with additional datasets focusing on health services, outcomes, and environmental factors.</p>	 <p>A basic user guide, extensive technical documentation, and data files are available; advanced users can download raw data files.</p>	 <p>Environmental indicators are available from the years 2017-2021.</p>	 <p>Data is available at the census block group level.</p>

Source + Agency	Description	Relevance as an Independent Data Source <sup>1</sup>	Relevance if Linked with Other Data Sources <sup>2</sup>	Accessibility <sup>3</sup>	Timeliness and Punctuality <sup>4</sup>	Granularity <sup>5</sup>
<p><b><u>Consolidated Human Activity Database (CHAD), EPA</u></b></p>	<p>The CHAD database includes nearly 180,000 individual study days of detailed human behavior, with each day broken down into individual hours and activity types. The data include demographic information such as age, sex, employment, and education level, which allows researchers to examine specific groups within the general population and how their unique behavior patterns influence their exposures to chemicals.</p>	<p style="text-align: center;"> Moderate</p> <p>Health related variables include having asthma and smoking. Environment-related measures include average and maximum temperatures in a specific geographic area and pesticide exposure.</p>	<p style="text-align: center;"> Moderate</p> <p>This database could be combine some variables with other health or environment-related datasets.</p>	<p style="text-align: center;"> Moderate</p> <p>User guide, technical memo, and 2019-2020 master data file available for download (in CSV or SAS dataset formats).</p>	<p style="text-align: center;"> Low</p> <p>Depending on the specific data source, data are available for the years 1982-2013. Few studies included go beyond 2012.</p>	<p style="text-align: center;"> Moderate</p> <p>Data are available at various levels, including state, city, and national levels, based on the studies from which they were collected.</p>

Source + Agency	Description	Relevance as an Independent Data Source <sup>1</sup>	Relevance if Linked with Other Data Sources <sup>2</sup>	Accessibility <sup>3</sup>	Timeliness and Punctuality <sup>4</sup>	Granularity <sup>5</sup>
<p><b><u>Data Catalog / OneStop Catalog, NOAA/NCEI</u></b></p>	<p>The traditional NOAA Data Catalog is an inventory of all NOAA data collections. The new NOAA OneStop catalog initially includes only the archived datasets but will eventually replace the traditional catalog. There are currently over 90,000 datasets within the Catalog.</p>	<div data-bbox="779 237 896 350" style="text-align: center;">  <p>Low</p> </div> <p>Topics covered in the Catalog include weather, climate, satellites, fisheries, coasts, and oceans with a variety of specific datasets within these topics. There are no clear human health measures included in the database. Given the vast number of environmental datasets and limited connection to human health, this has little relevance for answering health services research questions.</p>	<div data-bbox="1161 237 1278 350" style="text-align: center;">  <p>Low</p> </div> <p>Some datasets within this Catalog may be able to be linked with other health data sources to draw a connection between the environment and health outcomes. However, given the vast number of datasets with varying granularity and timeliness, it would be difficult to make this connection.</p>	<div data-bbox="1402 237 1520 350" style="text-align: center;">  <p>Moderate</p> </div> <p>Accessibility varies based on the specific dataset.</p>	<p>Not able to assess: The data collection years vary between each dataset.</p>	<p>Not able to assess: Granularity varies between each dataset.</p>

Source + Agency	Description	Relevance as an Independent Data Source <sup>1</sup>	Relevance if Linked with Other Data Sources <sup>2</sup>	Accessibility <sup>3</sup>	Timeliness and Punctuality <sup>4</sup>	Granularity <sup>5</sup>
<b><u>Disaster Research Response (DR2) Resources Portal, NIH/NIEHS<sup>6</sup></u></b>	The DR2 resources portal is a repository of data collection tools and related resources to empower human health research in response to disasters and public health emergencies. The repository includes tools developed by government, university, and other research organizations.	 <p>This repository does not include data, only the resources used to collect the data.</p>	 <p>This repository does not include data, only the resources used to collect the data.</p>	Not able to assess: this repository does not include data.	Not able to assess: this repository does not include data.	Not able to assess: this repository does not include data.

<sup>1</sup> High relevance: dataset makes the connection between environmental and health measures; moderate relevance: dataset includes both environment and health measures; low relevance: dataset only includes health or environment measures

<sup>2</sup> High linkage relevance: dataset has been linked with other datasets or has multiple health and/or environmental measures that could be linked with other datasets at a moderate or high granularity<sup>4</sup> level; moderate linkage relevance: dataset offers at least one health and/or environmental measure that can be linked with other datasets at a moderate or high granularity level; low linkage relevance: the dataset is unable to be linked at a moderate or high granularity level or it is unclear how it will be linked with other datasets to answer research questions related to health and the environment.

<sup>3</sup> High accessibility: all data is available for download in XLSX or other common format, codebook is available, and users can view summaries or interact with the data online without download; moderate accessibility: data is available for download, but there is limited interaction or summaries available; low accessibility: data is not available for download or only some variables are available for download

<sup>4</sup> High timeliness and punctuality: data is available for years 2012 and later; low timeliness and punctuality: data is not available after 2012.

<sup>5</sup> High granularity: available at the county level and lower; moderate granularity: available at the state level; low granularity: available at the national level

<sup>6</sup> Included at ASPE's request; note that this does not include analyzable data, just resources that can be used for conducting research related to disasters