Navigating Data Linkages in Environmental Health and Disaster Research

Aubrey K. Miller, MD, MPH
Ann Liu, PhD, MPH

July 17, 2025
50th Annual Natural Hazards Research and Applications Workshop and Researchers Meeting
Broomfield, CO

Data Needs in Environmental Health and Disaster Research

Problem: Ability to identify, access, harmonize, and use available earth and atmospheric observations, environmental monitoring, and local data sources WITH health data to understand a range of health impacts, especially for at-risk populations

Needs:

- **Exposure data at geographic levels and timescales** that can be linked to health data (e.g. EHRs, cohort studies, surveillance systems)
- **Models and datasets** for use by health scientists for timely and effective health research
- **Standardized metadata**, ontologies, common data elements to reduce a myriad of inefficiencies in comparing and understanding health impacts and generalizability of findings
- **Governance frameworks and ethical standards** to protect personal information while promoting data access and use by diverse stakeholders, such as researchers, health officials, policymakers, communities
- **Training** in use of GIS data, models, datasets needed to conduct health research

OSTP Science for Disaster Reduction (SDR):

Geo and Health Data Integration Working Group: 48 participants from 16 organizations

Wildfire scenario in Sonoma County, CA, used to:

- Evaluate capacity to collect, harmonize, share, and integrate different data streams
- Strengthen data-focused collaborations to address priority issues
- Map data gaps, challenges, and opportunities for reducing acute and longer-term health impacts

Foundation for subsequent work products:

- *Data Linkages for Wildfire Exposures and Human Health Studies: A Scoping Review*
- *Conceptual GeoHealth Framework for Disaster Response Research: Case Study for Wildland Urban Interface (WUI) Fires and Data Integration*
- NIH CHORDS project

HHS Patient-Centered Outcomes Research Trust Fund (PCORTF) Project: NIH CHORDS

Connecting Health Outcomes Research and Data Systems (CHORDS)

Goal: Strengthen data infrastructure to facilitate research connections between environmental exposures and health outcomes so researchers can 1) identify, analyze, and reduce the health effects associated with disaster-related events (e.g., wildfires) and 2) improve patient and population health outcomes

CHORDS Objectives and Deliverables

Objective 1:
Web-based data catalog

Objective 2:
Standardized, linked datasets

Objective 3:
Toolkit of software, tools, and tutorials

Objective 4:
Use case

Objective 5:
Addressing end user needs

CHORDS Data Platform

About CHORDS

The CHORDS project aims to build and strengthen data infrastructure for patient-centered outcomes research on environment and health.

Browse Catalog

View a catalog of over 100 CHORDS Data Resources.

Highlighted Research

Collection of case studies intended to provide examples of research articles that examine the health effects of wildfire-related exposures and highlight the key environment and health data sets used in these studies.

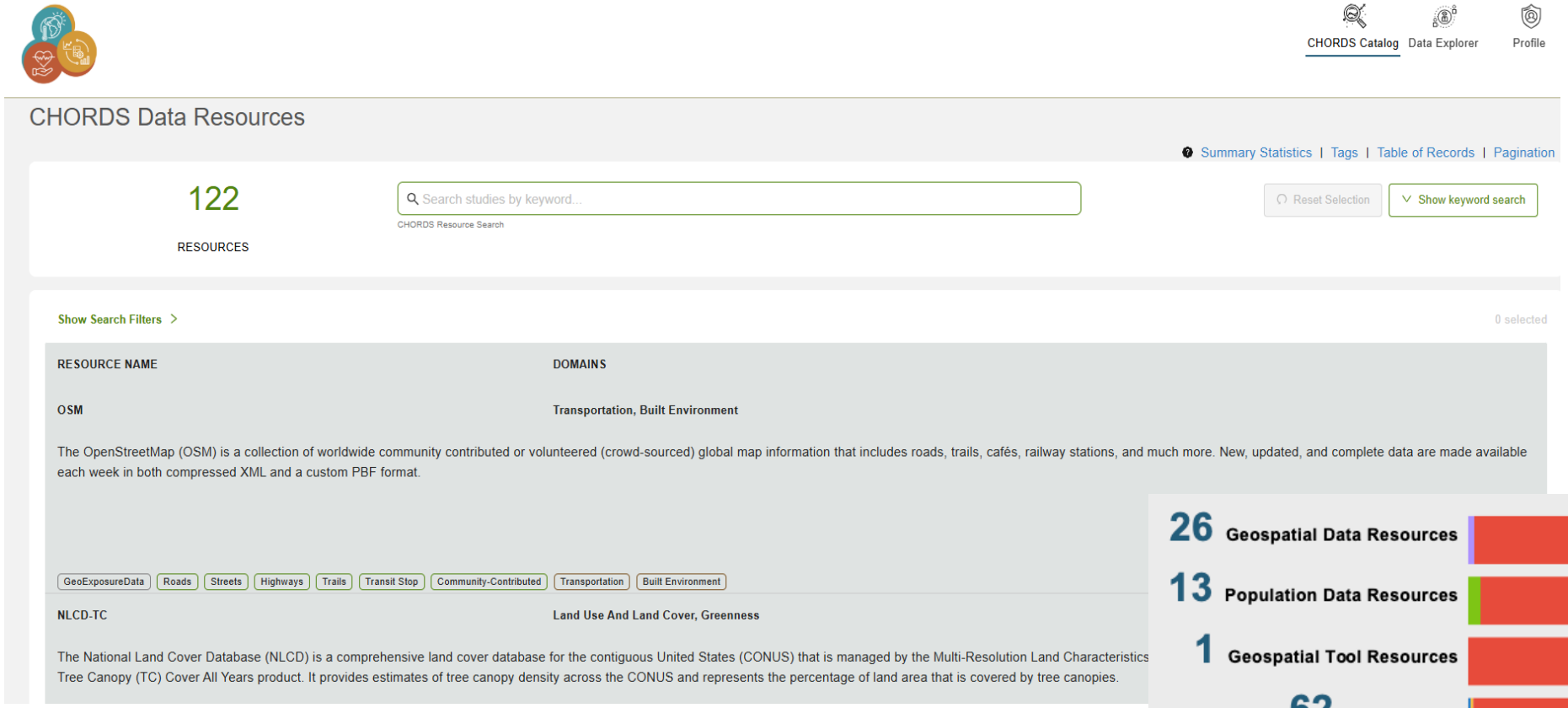
Software

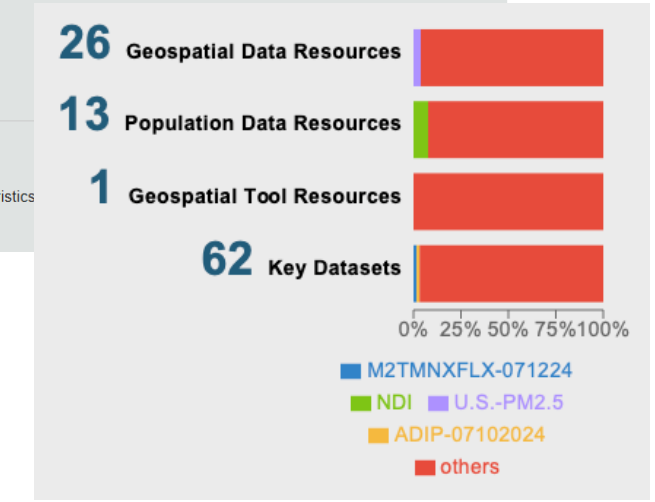
Find out more and download the CHORDS software.

Training and Use Cases

The CHORDS project seeks to connect researchers with guides, tutorials, and example code.

Data Catalog: Manually Curated Searchable Population and Geospatial Data Resources





Metadata provided:

- Project information
- Measures
- Spatial characteristics
- Temporal characteristics
- Data access information

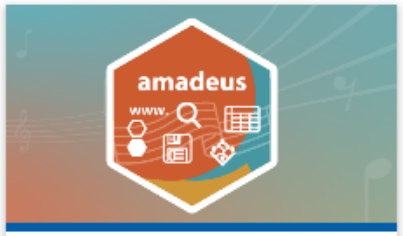
Other:

Keywords
Citations
Publications

Open-Source Software and Tools for Simplifying Environmental Health Data Analysis

CHORDS-specific tools

- ***amadeus:*** Downloads large-scale environmental and weather data and gets it ready for analysis
- ***beethoven:*** A reproducible and extensible pipeline for air pollution exposure designed for updated and timely releases
- ***chopin:*** Simplifies parallelization, running many calculations or processes simultaneously, of big environmental exposure data



Amadeus

repo status Active

Machine for Data, Environments, and User Setup for common environmental and climate health datasets is an R package developed to improve and expedite users' access to large, publicly available geospatial datasets.

Beethoven

repo status WIP

Building an Extensible, Reproducible, Test-driven, Harmonized, Open-source, Versioned, Ensemble model for air quality is an R package developed to facilitate the development of ensemble models for air quality.

Chopin

repo status WIP

Computation of Spatial Data by Hierarchical and Objective Partitioning of Inputs for Parallel Processing.

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: www.elsevier.com/locate/envsoft

Amadeus: Accessing and analyzing large scale environmental data in R

Mitchell Manware [a], Insang Song [b,1], Eva S. Marques [a,c], Mariana Alifa Kassien [a], Lara P. Clark [a], Kyle P. Messier [a,c,*]

Toolkit of Code, Data, and Educational Materials

Open-source resources: Tutorials and educational materials to support different types of users (e.g., students, epidemiologists, clinicians, data managers) in accessing, processing, and integrating geospatial exposure data into health research

- **Geospatial Data Foundations**
  - Working with point, polygon, and raster data
- **Health Data Integration**
  - Linkage to Census Units
  - FHIR PIT Tutorial
- **Use Case Studies**
  - AHRQ HCUP Analysis

8 HCUP and Amadeus Smoke Plume Use Case

Clinician/Medical Professional | Clinical Data Manager | Community Health Worker | Student

Integrating HCUP databases with Amadeus Exposure data

Date Modified: April 29, 2025

Author: Darius M. Bost

Programming Language: R

8.1 Motivation

Understanding the relationship between external environmental factors and health outcomes is critical for guiding public health strategies and policy decisions. Integrating individual patient records from the Healthcare Cost and Utilization Project (HCUP) with data from environmental datasets allows researchers to examine how elements such as air quality, wildfire emissions, and extreme temperatures impact hospital visits and healthcare utilization patterns.

Ultimately, linking HCUP and environmental exposure data enhances public health monitoring and helps researchers better quantify environmental health risks.

Use Case: Healthcare Cost and Utilization Project (HCUP)

Background and objectives

- Understanding the relationship between environmental factors and health outcomes is critical for guiding public health strategies and policy decisions.
- The Agency for Healthcare Research and Quality (AHRQ) **Healthcare Cost and Utilization Project (HCUP)** databases of state data organizations, hospital associations, private data organizations, and the federal government to create a national information resource of encounter-level healthcare data.
- Integrating individual patient records with data from environmental datasets allows researchers to examine how elements such as air quality, wildfire emissions, and extreme temperatures impact hospital visits and healthcare utilization patterns.

Use Case: Healthcare Cost and Utilization Project (HCUP) Spatial Analysis

Spatial analysis: bivariate map of asthma prevalence and heavy smoke exposure across ZIP codes in Oregon

- Map shows areas with overlapping smoke & health risks
- **Logistic regression model: exposure to medium and heavy smoke significantly increase the odds of asthma**
  - Findings can support targeted public health interventions

Legend:

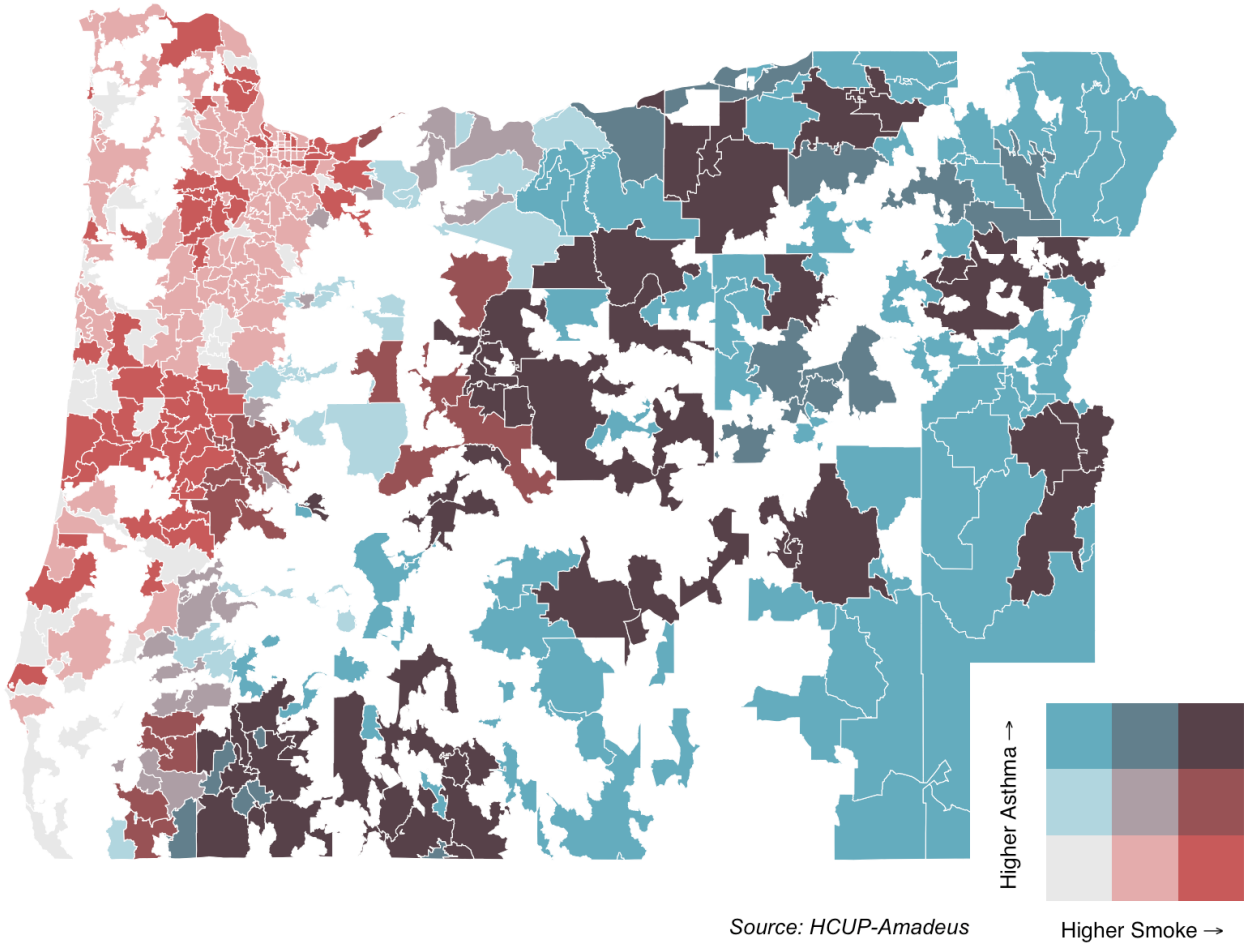
X-axis (**red**): higher smoke exposure
Y-axis (**blue**): higher asthma

Color interpretations:

- **Dark red**: high smoke, low asthma
- **Dark blue**: high asthma, low smoke
- **Dark purple**: high smoke, high asthma (compounded health and environmental burden)
- **Light gray**: low smoke, low asthma

Asthma Prevalence vs Heavy Smoke Exposure by ZIP Code

Bivariate map showing intersection of health and environmental burden



Source: HCUP-Amadeus

NIH Research Coordinating Center (RCC): CAFÉ Data Management Function
NSF GeoCAFÉ: Promoting Geo and Health Science Collaborations

Harvard Dataverse: Open-source data repository for the community of practice

- Collection of commonly-used climate and health data and linkages, including spatial data, curated and added by CAFÉ team as well as open to the research community to add their data sets (over 700 datasets currently available)

Tutorials/code walkthroughs

- CAFÉ GitHub with code and software tools for processing, linking, and analyzing data
- CAFÉ YouTube with tutorials and videos on spatial aggregations, data joins, census data, how to contribute data to Dataverse

CAFÉ

BU School of Public Health
Center for Climate & Health

HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH
Department of Biostatistics

CHORDS Team and Contributors

Core Team

- Adam Burkholder
- Alison Motsinger-Reif
- Ann Liu
- Aubrey Miller
- Charles Schmitt
- David Fargo
- David Reif
- Kyle Messier
- Trisha Castranio

NIH

- Asif Rizwan (NCI)
- Erin Iturriaga (NHLBI)
- Regina Bures (NHLBI)
- Richard Kwok (NIA)
- Umit Tokac (NHLBI)
- Cristina Justice (All of Us)

Other Agencies

- Pamela Owens (AHRQ)
- Patricia Keenan (AHRQ)
- NASA representatives

NIEHS

- Darius Bost
- Deep Patel
- Jennifer Fostel
- Lara Clark
- Marcus Jackson
- Maria Shatz
- Mariana Kassien
- Mike Conway
- Mitchell Manware
- Rupali Gupta
- Skylar Marvel
- Sue Nolte

Technical Expert Panel (TEP)

- Angela Werner (CDC)
- Caleb Dresser (Harvard)
- Cavin Ward-Caviness (EPA)
- Cole Brokamp (University of Cincinnati)
- Genee Smith (Johns Hopkins)
- Kevin Lane (BU/CAFE)
- Patrick Wall (CDC)
- Philip Awadalla (University of Toronto)
- Rima Habre (USC)
- Yang Liu (Emory)

Thank You!

Connecting Health Outcomes Research and Data Systems (CHORDS)