HRS | CDR Health and Retirement Study
Contextual Data Resource

Documentation for

Environmental Quality Index Files

released January 2025

Jennifer Ailshire, Kate Vavra-Musser, and Calley E. Fisk

Requested Acknowledgment

If you used the Contextual Data Resource data in a written analysis, please include the following acknowledgement:

> This analysis uses data or information from the Contextual Data Resource (CDR): United States Environmental Protection Agency Environmental Quality Index Files by County, 2000/05 and 2006/10, as of January 2025, developed by Jennifer Ailshire, Kate Vavra-Musser, and Calley E. Fisk at the USC/UCLA Center on Biodemography and Population Health. The development of the CDR was funded by the National Institute on Aging (R21 AG045625, P30 AG017625).

Suggested Citation

Jennifer Ailshire, Kate Vavra-Musser, Calley E. Fisk. 2025. Contextual Data Resource (CDR): United States Environmental Protection Agency Environmental Quality Index Files by County, 2000/05 and 2006/10. Los Angeles, CA: USC/UCLA Center on Biodemography and Population Health.

Data Sources

United States Environmental Protection Agency (EPA). Environmental Quality Index (EQI), https://www.epa.gov/healthresearch/environmental-quality-index-eqi

Linking CDR Data with HRS Respondent Data

The Contextual Data Resource is designed to be linked with the Health and Retirement Study Cross-Wave Geographic Information (Detail) file, which is available as restricted data. See https://hrs.isr.umich.edu/data-products/restricted-data for more information about accessing the HRS restricted data. All geographic identifiers are in string format, and match the geographic identifiers in the HRS Cross-Wave Geographic Information (Detail) file.

- To link census tract level data, merge using the 11-digit LINKCEN2010 geographic identifier, which is the 2-digit state + 3-digit county + 6-digit census tract FIPS code. The LINKCEN2010 geographic identifier has been harmonized to the 2010 census tract boundaries.

Note on HRS Cross-Wave Geographic Information File

Respondent locations are carried forward to the next wave and they appear for waves in which the interview may not have occurred. xIWTYPE in the Tracker file indicates whether someone was actually interviewed in that wave, in which case, the location would have been confirmed.

If users want to limit analyses to waves in which the R was interviewed, keep waves of observations where xIWTYPE = 1. In summary, a location appearing on a given line does not indicate that an interview took place in that wave; it is simply the location that was on record at the time.

Dataset Overview

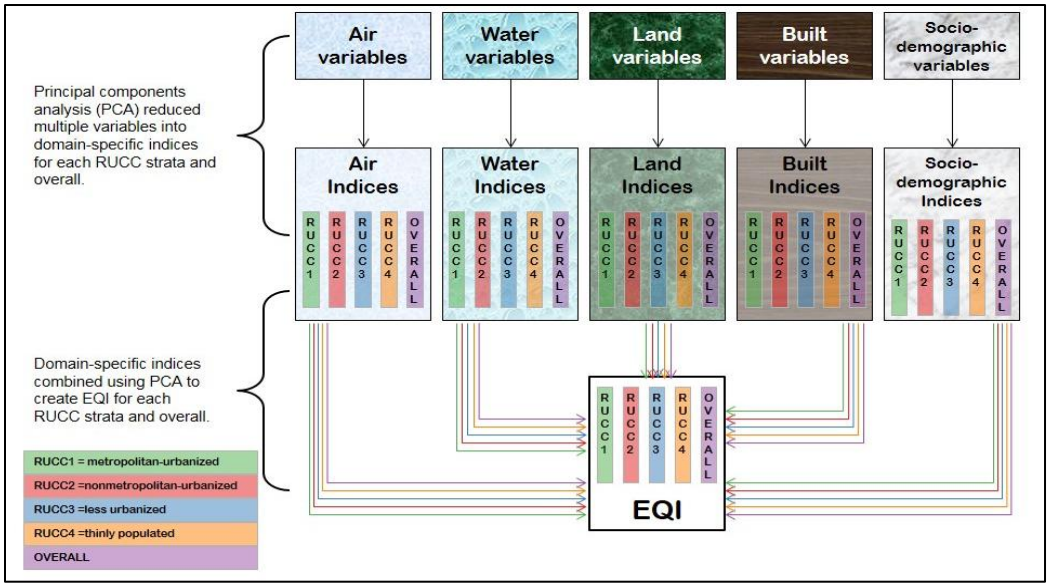
The US Environmental Protection Agency Environmental Quality Index (EQI) Files provide a snapshot of overall environmental quality using principal component analysis on a set of selected data in five domains: air, water, land, built, and sociodemographic environments. Data are available at the county level and includes an overall EQI value and EQI values for each of the five domains. EQI values are provided as 5-year estimates, and data are available for 2000-2005, and 2006-2010.

Data Summary

Dataset Name: Environmental Quality Index (EQI) Files
Data Source: United States Environmental Protection Agency (EPA)
Data Source URL: https://www.epa.gov/healthresearch/environmental-quality-index-eqi
Data Collection Method: Data from multiple primary sources compiled by the EPA.
Years Collected: Final 5-year estimates are provided for 2000-2005 and 2006-2010. Domain-specific measures include measures from specific years between 2000 and 2010.
Geographic Level: County; CDR data uses the 2010 county boundaries to be comparable across the years.
Geographic Coverage: EQI Files cover the contiguous U.S, Alaska, and Hawaii.

Technical Information about Environmental Quality Index

The EQI principal components analysis uses multiple variables to create domain-specific EQI indices. These indices are further combined using principal components analysis to estimate overall EQI values.



Visual description of the process used to generate domain-specific and overall EQI values; reproduced from original EQI technical documentation

The principal component analysis creates EQI values at the overall county level and stratified by rural-urban continuum codes. The CDR EQI files provide overall county level EQI values only.

Variable-specific information for each domain, including data source and years, can be found in the technical documentation for each survey year:

> U.S. EPA. Creating an Overall Environmental Quality Index - Technical Report (2000-2005) (Final, 2014). U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-14/304, 2014.
>
> U.S. EPA. Environmental Quality Index - Technical Report (2006-2010) (Final, 2020). U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-20/367, 2020.

Due to changes in variable selection, the EPA does not advise direct comparisons between the 2000-2005 and 2006-2010 EQI estimates.

For further information on the environmental quality index files, please see the following published journal papers:

Lobdell, D. T., J. S. Jagai, K. Rappazzo, and L. C. Messer. (2011). Data sources for environmental assessment: determining availability, quality and utility. American Journal of Public Health. American Public Health Association, Washington, DC, 101(Supp 1):S277-285. DOI: 10.2105/AJPH.2011.300184

Messer, L.C., Jagai, J.S., Rappazzo, K.M., & Lobdell, D.T. (2014). Construction of an environmental quality index for public health research. Environmental Health, 13(1):39. DOI: 10.1186/1476-069X-13-39

U.S. EPA. Environmental Quality Index - Overview Report. (Final, 2014) U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-14/305.

Variable List

Name	Definition	Geography	Years
	Geographic Identifiers		
STCTYFIPS10	2010 County FIPS code (5 digits: 1-2 state, 3-5 county)		
	Overall EQI County Estimate		
eqco	Overall EQI	County	2000-2005, 2006-2010
	Domain-Specific EQI County Estimates		
aico	Air EQI	County	2000-2005, 2006-2010
beco	Built Environment EQI	County	2000-2005, 2006-2010
laco	Land EQI	County	2000-2005, 2006-2010
seco	Sociodemographic EQI	County	2000-2005, 2006-2010
waco	Water EQI	County	2000-2005, 2006-2010

Variable Summary

Variable	Label	Mean	SD	Min	Max	N
STCTYFIPS10	2010 County FIPS code	.	.			3147
eqco20002005	Overall EQI Estimate	0.000	1.000	-6.786419	2.831005	3143
aico20002005	Air EQI Estimate	-0.000	1.000	-6.715335	3.712864	3143
waco20002005	Water EQI Estimate	0.000	1.000	-1.461017	2.047879	3143
laco20002005	Land EQI Estimate	-0.000	1.000	-4.544748	1.84174	3143
seco20002005	Sociodemographic EQI Estimate	-0.000	1.000	-5.128991	2.761724	3143
beco20002005	Built Environment EQI Estimate	0.000	1.000	-4.708277	5.65637	3143
eqco20062010	Overall EQI Estimate	-0.000	1.000	-5.879856	2.84909	3141
aico20062010	Air EQI Estimate	0.000	1.000	-3.258012	2.78984	3141
waco20062010	Water EQI Estimate	-0.000	1.000	-1.641266	1.478177	3141
laco20062010	Land EQI Estimate	0.000	1.000	-5.209604	2.094526	3141
seco20062010	Sociodemographic EQI Estimate	0.000	1.000	-4.80999	3.979472	3141
beco20062010	Built Environment EQI Estimate	0.000	1.000	-6.085963	3.883786	3141