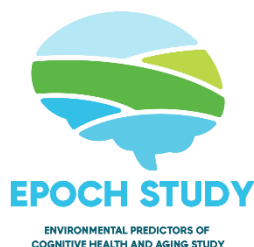


# HEALTH AND RETIREMENT STUDY RESTRICTED DATA DOCUMENTATION

## Individual-Level Fine Particulate Matter (PM<sub>2.5</sub>) Restricted Data

Version 1.0, released May 2025



Prepared by:

The Environmental Predictors of Cognitive Health and Aging  
(EPOCH) Study

Jennifer D'Souza<sup>1</sup>, Michael Young<sup>2</sup>, Adam Szpiro<sup>2</sup>, Meredith  
Pedde<sup>1</sup>, Sara Adar<sup>1</sup>

---

<sup>1</sup> University of Michigan School of Public Health

<sup>2</sup> University of Washington School of Public Health

The EPOCH Study was funded by the National Institutes of  
Environmental Health Sciences and National Institutes of  
Aging (NIEHS/NIA, R01ES028694, Adar)

The EPOCH air pollution estimates may be made available to HRS users through the virtual enclave as an HRS restricted geographic data product through a restricted data agreement. This restricted dataset is intended for exclusive use by persons specified in an Agreement for Use of Restricted Data from the Health and Retirement Study and/or a Supplemental Agreement with Research Staff for Use of Restricted Data from the Health and Retirement Study. Please go to <https://hrs.isr.umich.edu/data-products/restricted-data/vdi> to start a new restricted data agreement application. By requesting access to these data, you agree to cite the EPOCH study (NIEHS/NIA, R01ES028694, Adar) and its data sources in your publications.

## Requested Acknowledgment

When using these data for published products, please include the following acknowledgment:

This analysis uses air pollution estimates (Individual-Level Fine Particulate Matter, Version 1.0) generated at participant addresses for the Health and Retirement Study by the Environmental Predictors of Cognitive Health and Aging (EPOCH) Study developed by researchers at the University of Michigan School of Public Health and the University of Washington School of Public Health. The EPOCH Study was funded by the National Institutes of Environmental Health Sciences and the National Institutes of Aging (NIEHS/NIA, R01ES028694, Adar).

## Recommended Citations

Kirwa K, Szpiro AA, Sheppard L, et al. Fine-scale air pollution models for epidemiologic research: insights from approaches developed in the Multi-ethnic Study of Atherosclerosis and Air Pollution (MESA Air). *Curr Environ Health Rep*. 2021;8(2):113-126. doi:[10.1007/s40572-021-00310-y](https://doi.org/10.1007/s40572-021-00310-y)

Keller JP, Olives C, Kim S-Y, et al. A unified spatiotemporal modeling approach for predicting concentrations of multiple air pollutants in the multi-ethnic study of atherosclerosis and air pollution. *Environ Health Perspect*. 2015;123(4):301-309. doi:[10.1289/ehp.1408145](https://doi.org/10.1289/ehp.1408145)

Kim S-Y, Olives C, Sheppard L, et al. Historical prediction modeling approach for estimating long-term concentrations of PM<sub>2.5</sub> in cohort studies before the 1999 implementation of widespread monitoring. *Environ Health Perspect*. 2017;125(1):38-46. doi:[10.1289/EHP131](https://doi.org/10.1289/EHP131)

Wang M, Young M, Marshall J, Piepmeier L, Bi J, Kaufman JD, and Szpiro AA. National PM<sub>2.5</sub> spatiotemporal model integrating intensive monitoring data and land use regression in a likelihood-based universal kriging framework in the United States: 2000-2019. *Environ Poll*. 2025;366. doi:[10.1016/j.envpol.2024.125405](https://doi.org/10.1016/j.envpol.2024.125405)

## Questions?

For scientific questions on these variables, please contact [gatewayexposome@umich.edu](mailto:gatewayexposome@umich.edu).

## Data Description

Particulate matter of aerodynamic diameter less than 2.5 microns (PM<sub>2.5</sub>) is a measure of extremely small particles that can be found in air. These particles are generated from a wide variety of sources ranging from natural emissions of windblown dust to anthropogenic emissions from fossil fuel combustion, traffic, industry, and fires. PM<sub>2.5</sub> is ubiquitous and poses public health risks as their inhalation can initiate inflammation, oxidative stress, vascular changes, and autonomic imbalance that can ultimately lead to increased morbidity and mortality.

The EPOCH study has generated time-varying estimates of PM<sub>2.5</sub> concentrations in µg/m<sup>3</sup> for all HRS respondents using their exact address locations throughout the study. Specifically, we leveraged an existing spatiotemporal model that fuses measurements from the Environmental Protection Agency regulatory networks and several research studies, over 300 geographic covariates characterizing local transportation, land cover and use, population density, emissions sources, vegetation, and spatial correlations. Greater detail of these data sources and the methods used can be found in the citations above, but, briefly, the approach uses a hierarchical model that accounts for three components of variations across space and time: (1) a spatially varying long-term average, (2) spatially varying time trends, and (3) spatially correlated but temporally independent residuals. While originally developed for the six metropolitan areas of the Multi-Ethnic Study of Atherosclerosis, this model was later extended to provide coverage across the full United States.

Using this national spatiotemporal model, we predicted individual-level estimates of PM<sub>2.5</sub> for each participant's home address at a two-week time resolution between 1998 and 2016 (Keller et al., 2015). Between 1990 and 1997, we predicted these estimates with less temporal resolution (i.e., an annual average) since the ground-level monitoring data during this period was more limited. In the earlier periods, our estimates were derived using a combination of extrapolation from more recent EPA PM<sub>2.5</sub> measurements as well as PM<sub>2.5</sub> sulfate data from non-regulatory EPA stations and visibility data across the Weather Bureau Army Navy Network (Kim et al., 2017).

We generated time-weighted concentrations averaged over 1-, 5-, and 10-year periods and updated them monthly between 2000 and 2016 for the time period between a participant's first interview and their last interview (or time of death). For participants who relocated during the follow-up period, our estimates were weighted by the duration of time a participant spent at a given address. If the information was missing, we assumed that the move occurred midway between the current and the previous wave. The survey item related to how long the participant resided at their address was used to estimate the time they were at their first address. We required that there was 75% complete information for the averaging period of interest to have a value reported.

For more information on these data sources, please review the recommended citations and/or the following manuscript where these data linked with HRS were first published:

Zhang B, Weuve J, Langa KM, et al. Comparison of Particulate Air Pollution from Different Emission Sources and Incident Dementia in the US. *JAMA Intern Med.* 2023;183(10):1080–1089. [doi:10.1001/jamainternmed.2023.3300](https://doi.org/10.1001/jamainternmed.2023.3300)

## Data Details

**Spatial Resolution:** Exact address location (time-weighted by residential history, as described above)

**Temporal Resolution:** Data have been averaged over the previous 1-, 5-, and 10-year periods for each HRS respondent before the monthly index dates between 2000 and 2016 for the time period between a participant's first interview and their last interview (or time of death).

**Data Generation:** Calculated for each respondent based on their residential locations using a spatiotemporal prediction model

**Missing Data:** Exposure data is missing when the prediction model could not be estimated due to addresses being missing or incomplete (0.3% of survey addresses). All others were missing because there were no model estimates available for those locations (i.e., outside of design space or missing predictors required to estimate exposures).

**Units:** µg/m<sup>3</sup>

**Variable Names:** hhidpn

HHID

PN

end\_date

pm25\_epoch\_1y

pm25\_epoch\_5y

pm25\_epoch\_10y

**Summary Statistics:**

Variable	Description	N	Mean	Std Dev	Minimum	Maximum
hhidpn	HRS Participant ID: Household ID + Person Number (numerical as HHIDPN=1000*HHID+PN)	4101845	223172462	230043222	2010	959738010
HHID	Household ID (6-character)	4101845	NA	NA	000002	959738
PN	Person Number (3-character)	4101845	NA	NA	010	043
end_date	Index date for exposure average	4101845	NA	NA	1/15/2000	12/15/2016
pm25_epoch_1y	PM <sub>2.5</sub> , 1-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3961600	10.1611	3.004083	1.749139	31.19463

<b>pm25_epoch_5y</b>	PM <sub>2.5</sub> , 5-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3990274	10.76898	3.013132	2.322188	28.04712
<b>pm25_epoch_10y</b>	PM <sub>2.5</sub> , 10-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3973055	11.43695	3.061057	2.377636	26.13627