

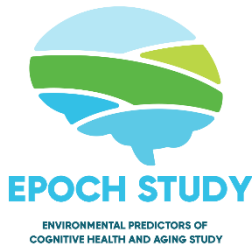
HEALTH AND RETIREMENT STUDY  
RESTRICTED DATA  
DOCUMENTATION

Individual-Level Source-Specific 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>, Katherine Francois<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 Source-Specific 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

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)

McDuffie EE, Martin RV, Spadaro JV, Burnett R, Smith, SJ, O'Rourke P, Hammer M, van Donkelaar A, Bindle L, Shah V, Jaegle L, Luo G, Yu F, Adeniran J, Lin J, Brauer M. Source Sector and Fuel Contributions to Ambient PM<sub>2.5</sub> Attributable Mortality Across Multiple Spatial Scales. *Nature Communications.* 2021;12:3594. doi:[10.1038/s41467-021-23853-y](https://doi.org/10.1038/s41467-021-23853-y)

## 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 generated by a 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 its inhalation can initiate inflammation, oxidative stress, vascular changes, and autonomic imbalance that can ultimately lead to increased morbidity and mortality. The originating source of the PM<sub>2.5</sub> has importance both for interventions but also for health since the emissions source can alter the physical and chemical properties of the particles.

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).

To estimate the PM<sub>2.5</sub> from each source, we combined our spatiotemporal estimates of total mass with the spatial estimates of the fraction of PM<sub>2.5</sub> concentrations attributable to 9 different sources as well as 3 fuel categories. These fractions have been predicted with a resolution of 0.5° x 0.625° in North America by serially running the GEOS Chem atmospheric chemical-transport model with all emission sources but one to isolate the unique contribution of that source to the total mixture of PM<sub>2.5</sub>. Although these source fraction data reflect emissions in 2017, these temporally fixed values are assumed to be representative of other years. The sources captured by this model and included in this data product are as follows:

List of Sources with Descriptions of Primary Emissions:

AGR	<i>Agriculture</i> - manure management, soil fertilizer emissions, rice cultivation, enteric fermentation, and other agricultural activities
ENEcoal	<i>Energy Production (coal combustion only)</i> - electricity and heat production, fuel production and transformation, oil and gas fugitive/flaring, and fossil fuel fires
ENEother	<i>Energy Production (all non-coal combustion)</i> - electricity and heat production, fuel production and transformation, oil and gas fugitive/flaring, and fossil fuel fires
GFEDoburn	<i>Other Open Fires</i> - deforestation, boreal forest, peat, savannah, and temperate forest fires (from the GFED fires inventory)
INDcoal	<i>Industry (coal combustion only)</i> – industrial combustion (iron and steel, non-ferrous metals, chemicals, pulp and paper, food and tobacco, non-metallic minerals, construction, transportation equipment, machinery, mining and quarrying, wood products, textile and leather, and other industry combustion) and non-combustion industrial processes and product use (cement production, lime production, other minerals, chemical industry, metal production, food, beverage, wood, pulp, and paper, and other noncombustion industrial emissions)
INDother	<i>Industry (all non-coal combustion)</i> – industrial combustion (iron and steel, non-ferrous metals, chemicals, pulp and paper, food and tobacco, non-metallic minerals, construction, transportation equipment, machinery, mining and quarrying, wood products, textile and leather, and other industry combustion) and non-combustion industrial processes and product use (cement production, lime production, other minerals, chemical industry, metal production, food, beverage, wood, pulp, and paper, and other noncombustion industrial emissions)
NRTR	<i>Non-Road/Off-Road Transportation</i> – rail, domestic navigation, other transportation
Other	<i>All Remaining Sources</i> - Includes anthropogenic fugitive, combustion, and industrial dust; agricultural waste burning; commercial combustion, other combustion, residential combustion (solid biofuel only, coal only, all non-coal and non-solid biofuel); international shipping; solvents; waste; volcanic SO <sub>2</sub> ; lightning Nox; biogenic soil NO; ocean emissions; biogenic emissions; very short-lived iodine and bromine species; decaying plants (misc. inventories)
ROAD	<i>Road Transportation</i> – cars, motorcycles, heavy and light duty trucks and buses
WDUST	<i>Windblown Dust</i>

List of Fuel Categories with Primary Emissions (Note: These do not sum to 100% for each grid cell since they only include combustion sources of PM<sub>2.5</sub>)

BIOFUEL	<i>Solid Biofuel (or Biomass) Combustion</i> - Solid biofuel
COAL	<i>Total Coal Combustion</i> – Hard coal, brown coal, coal coke
OILGAS	<i>Liquid Oil and Natural Gas Combustion</i> –light and heavy oil, diesel oil, and natural gas

Using these two data sources, we calculated the absolute contributions of PM<sub>2.5</sub> attributable to each source for each respondent by multiplying the location-specific total mass by the location-specific fractional source contributions. We then 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.

**Data Generation:** Calculated for each respondent based on their residential locations using one spatiotemporal prediction model and one physical-chemical transport 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\_AGR\_epoch\_1y

pm25\_ENEcoal\_epoch\_1y

pm25\_ENEother\_epoch\_1y

pm25\_GFEDoburn\_epoch\_1y

pm25\_INDcoal\_epoch\_1y

pm25\_INDoother\_epoch\_1y

pm25\_NRTR\_epoch\_1y

pm25\_ROAD\_epoch\_1y  
 pm25\_WDUST\_epoch\_1y  
 pm25\_OTHER\_epoch\_1y  
 pm25\_BIOFUEL\_epoch\_1y  
 pm25\_COAL\_epoch\_1y  
 pm25\_OILGAS\_epoch\_1y

pm25\_AGR\_epoch\_5y  
 pm25\_ENEcoal\_epoch\_5y  
 pm25\_ENEother\_epoch\_5y  
 pm25\_GFEDoburn\_epoch\_5y  
 pm25\_INDcoal\_epoch\_5y  
 pm25\_INDothier\_epoch\_5y  
 pm25\_NRTR\_epoch\_5y  
 pm25\_ROAD\_epoch\_5y  
 pm25\_WDUST\_epoch\_5y  
 pm25\_OTHER\_epoch\_5y  
 pm25\_BIOFUEL\_epoch\_5y  
 pm25\_COAL\_epoch\_5y  
 pm25\_OILGAS\_epoch\_5y

pm25\_AGR\_epoch\_10y  
 pm25\_ENEcoal\_epoch\_10y  
 pm25\_ENEother\_epoch\_10y  
 pm25\_GFEDoburn\_epoch\_10y  
 pm25\_INDcoal\_epoch\_10y  
 pm25\_INDothier\_epoch\_10y  
 pm25\_NRTR\_epoch\_10y  
 pm25\_ROAD\_epoch\_10y  
 pm25\_WDUST\_epoch\_10y  
 pm25\_OTHER\_epoch\_10y  
 pm25\_BIOFUEL\_epoch\_10y  
 pm25\_COAL\_epoch\_10y  
 pm25\_OILGAS\_epoch\_10y

### Summary Statistics:

Variable	Description	N	Mean	Std Dev	Minimum	Maximum
<b>hhidpn</b>	HRS Participant ID: Household ID + Person Number (numerical as HHIDPN=1000*HHID+PN)	4101845	223172462	230043222	2010	959738010
<b>HHID</b>	Household ID (6- character)	4101845	NA	NA	000002	959738
<b>PN</b>	Person Number (3- character)	4101845	NA	NA	010	043

<b>end_date</b>	Index date for exposure average	4101845	NA	NA	1/15/2000	12/15/2016
<b>pm25_AGR_epoch_1y</b>	AGR PM <sub>2.5</sub> , 1-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3961600	1.0352	0.5805	0.0084	3.2410
<b>pm25_ENEcoal_epoch_1y</b>	ENEcoal PM <sub>2.5</sub> , 1-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3961600	0.7496	0.4783	0.0006	3.0449
<b>pm25_ENEother_epoch_1y</b>	ENEother PM <sub>2.5</sub> , 1-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3961600	0.5045	0.2289	0.0025	1.7174
<b>pm25_GFEDoburn_epoch_1y</b>	GFEDoburn PM <sub>2.5</sub> , 1-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3961600	1.0515	0.8080	0.1927	10.5766
<b>pm25_INDcoal_epoch_1y</b>	INDcoal PM <sub>2.5</sub> , 1-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3961600	0.2138	0.1223	-0.0483	1.2625
<b>pm25_INDoother_epoch_1y</b>	INDoother PM <sub>2.5</sub> , 1-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3961600	0.9291	0.5818	0.0051	5.9196
<b>pm25_NRTR_epoch_1y</b>	NRTR PM <sub>2.5</sub> , 1-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3961600	0.3484	0.1716	0.0037	0.9332
<b>pm25_ROAD_epoch_1y</b>	ROAD PM <sub>2.5</sub> , 1-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3961600	1.2416	0.4256	0.0217	2.9122
<b>pm25_WDUST_epoch_1y</b>	WDUST PM <sub>2.5</sub> , 1-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3961600	0.1947	0.4199	0.0046	6.4682
<b>pm25_OTHER_epoch_1y</b>	OTHER PM <sub>2.5</sub> , 1-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3961600	3.8928	1.4668	0.0682	15.4613
<b>pm25_BIOFUEL_epoch_1y</b>	BIOFUEL PM <sub>2.5</sub> , 1-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3961600	1.2860	0.8857	0.0079	8.4406
<b>pm25_COAL_epoch_1y</b>	COAL PM <sub>2.5</sub> , 1-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3961600	1.0253	0.5322	-0.0034	3.6608
<b>pm25_OILGAS_epoch_1y</b>	OILGAS PM <sub>2.5</sub> , 1-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3961600	2.9046	0.9866	0.0366	9.5973
<b>pm25_AGR_epoch_5y</b>	AGR PM <sub>2.5</sub> , 5-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3990274	1.094627	0.5973326	0.0108581	2.999032
<b>pm25_ENEcoal_epoch_5y</b>	ENEcoal PM <sub>2.5</sub> , 5-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3990274	0.7918212	0.4965094	0.000766961	2.971033
<b>pm25_ENEother_epoch_5y</b>	ENEother PM <sub>2.5</sub> , 5-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3990274	0.5336323	0.2330019	0.003011034	1.897295
<b>pm25_GFEDoburn_epoch_5y</b>	GFEDoburn PM <sub>2.5</sub> , 5-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3990274	1.114338	0.8490541	0.2130968	11.06741
<b>pm25_INDcoal_epoch_5y</b>	INDcoal PM <sub>2.5</sub> , 5-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3990274	0.2270501	0.127934	-0.04439187	1.136027

<b>pm25_INDoth</b> _epoch_5y	INDother PM <sub>2.5</sub> , 5-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3990274	0.9891633	0.6098424	0.005931477	5.486178
<b>pm25_NRTR</b> _epoch_5y	NRTR PM <sub>2.5</sub> , 5-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3990274	0.3685481	0.1757422	0.004753547	0.9177981
<b>pm25_ROAD</b> _epoch_5y	ROAD PM <sub>2.5</sub> , 5-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3990274	1.31347	0.427334	0.02809597	2.775055
<b>pm25_WDUST</b> _epoch_5y	WDUST PM <sub>2.5</sub> , 5-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3990274	0.2039475	0.4300192	0.006004481	7.025232
<b>pm25_OTHER</b> _epoch_5y	OTHER PM <sub>2.5</sub> , 5-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3990274	4.132381	1.506823	0.08846596	13.90127
<b>pm25_BIOFUEL</b> _epoch_5y	BIOFUEL PM <sub>2.5</sub> , 5-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3990274	1.370121	0.930661	0.01023264	7.660152
<b>pm25_COAL</b> _epoch_5y	COAL PM <sub>2.5</sub> , 5-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3990274	1.084647	0.5492245	- 0.003090989	3.572003
<b>pm25_OILGAS</b> _epoch_5y	OILGAS PM <sub>2.5</sub> , 5-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3990274	3.079068	0.9986282	0.04746042	8.628977
<b>pm25_AGR</b> _epoch_10y	AGR PM <sub>2.5</sub> , 10-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3973055	1.159155	0.6184098	0.01169798	2.973337
<b>pm25_ENEcoal</b> _epoch_10y	ENEcoal PM <sub>2.5</sub> , 10-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3973055	0.8365891	0.5177105	0.000817115	3.071146
<b>pm25_ENEother</b> _epoch_10y	ENEother PM <sub>2.5</sub> , 10-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3973055	0.5654807	0.2399481	0.003207935	2.011016
<b>pm25_GFEDoburn</b> _epoch_10y	GFEDoburn PM <sub>2.5</sub> , 10-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3973055	1.185947	0.9082834	0.2254004	11.66757
<b>pm25_INDcoal</b> _epoch_10y	INDcoal PM <sub>2.5</sub> , 10-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3973055	0.2416039	0.1346575	-0.04448466	1.056756
<b>pm25_INDoth</b> _epoch_10y	INDother PM <sub>2.5</sub> , 10-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3973055	1.055717	0.6434352	0.006592339	5.147542
<b>pm25_NRTR</b> _epoch_10y	NRTR PM <sub>2.5</sub> , 10-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3973055	0.3904087	0.1811757	0.005121238	0.9547974
<b>pm25_ROAD</b> _epoch_10y	ROAD PM <sub>2.5</sub> , 10-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3973055	1.391648	0.4330919	0.03026921	2.87101
<b>pm25_WDUST</b> _epoch_10y	WDUST PM <sub>2.5</sub> , 10-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3973055	0.2140835	0.4444451	0.006468932	7.341024
<b>pm25_OTHER</b> _epoch_10y	OTHER PM <sub>2.5</sub> , 10-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3973055	4.396321	1.564795	0.09530887	12.95418
<b>pm25_BIOFUEL</b> _epoch_10y	BIOFUEL PM <sub>2.5</sub> , 10-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3973055	1.463552	0.983837	0.01102414	7.187326

<b>pm25_COAL_epoch_10y</b>	COAL PM <sub>2.5</sub> , 10-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3973055	1.148225	0.5696626	-0.00309745	3.692367
<b>pm25_OILGAS_epoch_10y</b>	OILGAS PM <sub>2.5</sub> , 10-year average (µg/m <sup>3</sup> ), indexed to 'end_date'	3973055	3.269945	1.021844	0.05113151	8.041084