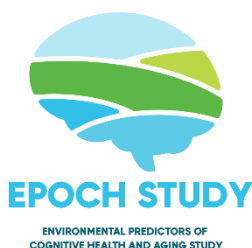


HEALTH AND RETIREMENT STUDY RESTRICTED DATA DOCUMENTATION

Individual-Level Ozone (O₃) Restricted Data

Version 1.0, released May 2025



Prepared by:

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

Jennifer D'Souza¹, Michael Young², Adam Szpiro², Meredith
Pedde¹, Katherine Francois¹, Sara Adar¹

¹ University of Michigan School of Public Health

² 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 Ozone, 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

Wang M, Keller JP, Adar SD, et al. Development of long-term spatiotemporal models for ambient ozone in six metropolitan regions of the United States: the MESA Air Study. *Atmos Environ (1994)*. 2015;123(A):79-87.
doi:[10.1016/j.atmosenv.2015.10.042](https://doi.org/10.1016/j.atmosenv.2015.10.042)

Keller JP, Olives C, Kim SY, 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)

Questions?

For scientific questions on these variables, please contact gatewayexposome@umich.edu.

Data Description

Ground-level ozone, or O₃, is a component of smog that is formed when emissions from sources, including cars and industry, react together in the presence of sunlight and heat. As a highly reactive gas, ozone can cause a variety of health problems, especially to the respiratory system. For example, ozone has been shown to cause coughing, shortness of breath, asthma exacerbation, and increase the susceptibility to respiratory infections.

The EPOCH study has generated time-varying estimates of O₃ concentrations in parts per billion, or ppb, 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 fuses chemical transport modeling and a spatiotemporal land-use regression model to estimate concentrations.

Using this national spatiotemporal model, we predicted individual-level estimates of O₃ for each participant's home address at a two-week time resolution between 1990 and 2016. These estimates included predictions for all hours of the day and all seasons. 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 using a spatiotemporal 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: ppb

Variable names: hhidpn

HHID

PN

end_date

o3_epoch_1y

o3_epoch_5y

o3_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
o3_epoch_1y	O ₃ , 1-year average (ppb), indexed to 'end_date'	3961600	26.9823	3.706485	10.51207	48.20406
o3_epoch_5y	O ₃ , 5-year average (ppb), indexed to 'end_date'	3990306	26.93115	3.600055	11.95983	46.62007
o3_epoch_10y	O ₃ , 10-year average (ppb), indexed to 'end_date'	3973127	26.70476	3.588058	14.32205	46.27654