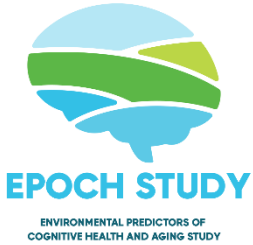
HEALTH AND RETIREMENT STUDY
RESTRICTED DATA
DOCUMENTATION

Individual-Level Nitrogen Dioxide (NO_2)
Restricted Data

Version 1.0, released May 2025



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

Jennifer D'Souza[1], Michael Young[2], Adam Szpiro[2], Meredith Pedde[1], Sara Adar[1]

[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 acknowledgement:

> This analysis uses air pollution estimates (Individual-Level Nitrogen Dioxide, 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

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

Kim S-Y, Bechle M, Hankey S, Sheppard L, Szpiro AA, Marshall JD. Concentrations of criteria pollutants in the contiguous U.S., 1979 – 2015: Role of prediction model parsimony in integrated empirical geographic regression. *PLoS ONE* 2020;15(2): e0228535. doi:10.1371/journal.pone.0228535

Young MT, Bechle MJ, Sampson PD, Szpiro AA, Marshall JD, Sheppard L, Kaufman JD. Satellite-Based NO_2 and Model Validation in a National Prediction Model Based on Universal Kriging and Land-Use Regression. *Environ Sci Technol*. 2016;50(7):3686-94. doi:10.1021/acs.est.5b05099

Questions?

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

Data Description

Nitrogen dioxide, or NO_2, is a pollutant that is commonly formed by the combustion of fossil fuels, most notably from the transportation sector but also from power plants and industry. In urban areas, NO_2 is commonly used as an indicator of a mixture of pollutants generated by the transportation sector and traffic. It has been associated with adverse health effects in epidemiology studies.

The EPOCH study has generated time-varying estimates of NO_2 concentrations in parts per billion, or ppb, for all HRS respondents using their exact address locations over the course of the study. Specifically, we leveraged an existing spatiotemporal model that fuses measurements from the Environmental Protection Agency regulatory networks and non-regulatory monitors that captured fine-scale residential and roadway gradients with over 300 geographic covariates characterizing local transportation, land cover and use, population density, emissions sources, vegetation, satellite-based information, 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 universal kriging approach to estimate concentrations at all homes.

Using this national spatiotemporal model, we predicted individual-level estimates of NO_2 for each participant's home address at a two-week time resolution between 1990 and 2016. 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

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
no2_epoch_1y
no2_epoch_5y
no2_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
no2_epoch_1y	NO_2, 1-year average (ppb), indexed to 'end_date'	3961600	9.585391	6.16275	0.8190564	54.94265
no2_epoch_5y	NO_2, 5-year average (ppb), indexed to 'end_date'	3990306	10.25037	6.484701	0.8816212	55.06686
no2_epoch_10y	NO_2, 10-year average (ppb), indexed to 'end_date'	3973127	10.92586	6.804651	0.9081966	56.65188