# HEALTH AND RETIREMENT STUDY
# RESTRICTED DATA
# DOCUMENTATION

## Spatial Splines

Version 1.0, released May 2025

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

Kayleigh Keller[1], Adam Szpiro[2], Meredith Pedde[3], Katherine Francois[3], Sara Adar[3]

[1] Colorado State University
[2] University of Washington School of Public Health
[3] University of Michigan 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 spatial splines may be made available to HRS users through the virtual enclave as an HRS user contributed product. This dataset is intended for use with the HRS survey 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 spatial splines (Spatial Splines, Version 1.0) generated at the centroids of US Census tracts 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 and Szpiro AA. Selecting a Scale for Spatial Confounding Adjustment. *Journal of the Royal Statistical Society: Series A Statistical Society*. 2020;183(3): 1121–1143. doi:10.1111/rssa.12556

Paciorek CJ. The importance of scale for spatial-confounding bias and precision of spatial regression estimators. *Statistical Science*. 2010;25(1):107–125. doi:10.1214/10-STS326

## Questions?

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

## Data Description

Environmental exposures are inherently spatial and thus there is a risk of these exposures being correlated with other features of place. While one can attempt to adjust for these characteristics in statistical analyses using personal and neighborhood features, there is the possibility of residual confounding by place. Therefore, we have developed variables that can be used to estimate spatial splines to flexibly account for place in statistical models. These continuous place-based indicators allow the user to model place flexibly. We provide these splines for 5 degrees of freedom (df), 10 df, and 15 df to allow the user to specify the number of degrees of freedom desired. These splines were generated at the centroids of all 2000 and 2010 US census tracts.

Users can link these data to HRS respondents by US census tract IDs (i.e., 'TRACTCE'). The locations for the 2000 and 2010 Census use the same underlying spatial map so they can be mixed and matched based on current address. In fact, matching should be performed based on the location at an individual exam. The averaging of spatial splines from different locations across time is not recommended.

To properly use these data in a model, the user should enter all *x* variables corresponding to the degrees of freedom specified into their model. For example, for 5 degrees of freedom, Y=spatialbasis_1of5 +spatialbasis_2of5+ spatialbasis_3of5 +spatialbasis_4of5 +spatalbasis_5of5. Be sure to use the connected pairs (i.e., 1of10 – 10of10 rather than mixing and matching 1of5 with 10of10 or using fewer than specified like 1of10-8of10). The degrees of freedom are not set, but we recommend 10 df as a default. The spline coefficients should not be interpreted.

For more information on these data sources, including how to interpret and describe, please review the recommended citations and/or the following manuscript where spatial splines 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: Centroids of US census tracts

Temporal Resolution: Estimated for 2000 and 2010 US Census tracts

Data Generation: Calculated for each US Census tract using a temporal model

Units: unitless

Variable Names: STATEFP
COUNTYFP

TRACTCE
YEAR
spatialbasis_1of5_epoch
spatialbasis_2of5_epoch
spatialbasis_3of5_epoch
spatialbasis_4of5_epoch
spatialbasis_5of5_epoch

spatialbasis_1of10_epoch
spatialbasis_2of10_epoch
spatialbasis_3of10_epoch
spatialbasis_4of10_epoch
spatialbasis_5of10_epoch
spatialbasis_6of10_epoch
spatialbasis_7of10_epoch
spatialbasis_8of10_epoch
spatialbasis_9of10_epoch
spatialbasis_10of10_epoch

spatialbasis_1of15_epoch
spatialbasis_2of15_epoch
spatialbasis_3of15_epoch
spatialbasis_4of15_epoch
spatialbasis_5of15_epoch
spatialbasis_6of15_epoch
spatialbasis_7of15_epoch
spatialbasis_8of15_epoch
spatialbasis_9of15_epoch
spatialbasis_10of15_epoch
spatialbasis_11of15_epoch
spatialbasis_12of15_epoch
spatialbasis_13of15_epoch
spatialbasis_14of15_epoch
spatialbasis_15of15_epoch

Summary Statistics:

| Variable | Description | N | Mean | Std Dev | Minimum | Maximum |
|---|---|---|---|---|---|---|
| **STATEFP** | US Census State FIPS code (2 digits) | 137274 | NA | NA | 01 | 56 |
| **COUNTYFP** | US Census County FIPS code (3 digits) | 137274 | NA | NA | 001 | 840 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **TRACTCE** | US Census Tract FIPS code (6 digits) | 137274 | NA | NA | 000100 | 998900 |
| **YEAR** | US Census Year | 137274 | NA | NA | 2000 | 2010 |
| **spatialbasis_1of5_epoch** | Predicted value of variable 1 (of 5) from a spatial curve function (df=5) | 137274 | 0.008928806 | 0.9975701 | -2.178893 | 1.572162 |
| **spatialbasis_2of5_epoch** | Predicted value of variable 2 (of 5) from a spatial curve function (df=5) | 137274 | 0.02939412 | 0.9927229 | -2.679709 | 2.28347 |
| **spatialbasis_3of5_epoch** | Predicted value of variable 3 (of 5) from a spatial curve function (df=5) | 137274 | 0.9939849 | 0.1081934 | 0.8353221 | 1.203657 |
| **spatialbasis_4of5_epoch** | Predicted value of variable 4 (of 5) from a spatial curve function (df=5) | 137274 | 0.2924456 | 0.9429315 | -1.14901 | 2.344247 |
| **spatialbasis_5of5_epoch** | Predicted value of variable 5 (of 5) from a spatial curve function (df=5) | 137274 | 0.5415529 | 0.85049 | -1.170267 | 1.670743 |
| **spatialbasis_1of10_epoch** | Predicted value of variable 1 (of 10) from a spatial curve function (df=10) | 137274 | 0.008928806 | 0.9975701 | -2.178893 | 1.572162 |
| **spatialbasis_2of10_epoch** | Predicted value of variable 2 (of 10) from a spatial curve function (df=10) | 137274 | 0.02939412 | 0.9927229 | -2.679709 | 2.28347 |
| **spatialbasis_3of10_epoch** | Predicted value of variable 3 (of 10) from a spatial curve function (df=10) | 137274 | 0.9939597 | 0.1084016 | 0.8348335 | 1.203732 |
| **spatialbasis_4of10_epoch** | Predicted value of variable 4 (of 10) from a spatial curve function (df=10) | 137274 | -0.2924999 | 0.9429146 | -2.344281 | 1.149015 |
| **spatialbasis_5of10_epoch** | Predicted value of variable 5 (of 10) from a spatial curve function (df=10) | 137274 | 0.5416961 | 0.8503993 | -1.169901 | 1.67074 |
| **spatialbasis_6of10_epoch** | Predicted value of variable 6 (of | 137274 | -0.6650223 | 0.7351141 | -2.526751 | 0.4053074 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | 10) from a spatial curve function (df=10) | | | | | |
| **spatialbasis_7of10_epoch** | Predicted value of variable 7 (of 10) from a spatial curve function (df=10) | 137274 | 0.7671249 | 0.6441013 | -0.9237949 | 1.778229 |
| **spatialbasis_8of10_epoch** | Predicted value of variable 8 (of 10) from a spatial curve function (df=10) | 137274 | 0.3587411 | 0.9330478 | -2.42042 | 1.994669 |
| **spatialbasis_9of10_epoch** | Predicted value of variable 9 (of 10) from a spatial curve function (df=10) | 137274 | -0.2629408 | 0.9657296 | -1.740768 | 1.909545 |
| **spatialbasis_10of10_epoch** | Predicted value of variable 10 (of 10) from a spatial curve function (df=10) | 137274 | -0.8261996 | 0.5619776 | -2.056828 | 0.2104132 |
| **spatialbasis_1of15_epoch** | Predicted value of variable 1 (of 15) from a spatial curve function (df=15) | 137274 | 0.008928806 | 0.9975701 | -2.178893 | 1.572162 |
| **spatialbasis_2of15_epoch** | Predicted value of variable 2 (of 15) from a spatial curve function (df=15) | 137274 | 0.02939412 | 0.9927229 | -2.679709 | 2.28347 |
| **spatialbasis_3of15_epoch** | Predicted value of variable 3 (of 15) from a spatial curve function (df=15) | 137274 | 0.9939593 | 0.1084044 | 0.8348235 | 1.203735 |
| **spatialbasis_4of15_epoch** | Predicted value of variable 4 (of 15) from a spatial curve function (df=15) | 137274 | 0.2925007 | 0.9429144 | -1.149012 | 2.344295 |
| **spatialbasis_5of15_epoch** | Predicted value of variable 5 (of 15) from a spatial curve function (df=15) | 137274 | 0.5417003 | 0.8503963 | -1.169881 | 1.670742 |
| **spatialbasis_6of15_epoch** | Predicted value of variable 6 (of 15) from a spatial curve function (df=15) | 137274 | -0.6650165 | 0.7351194 | -2.526764 | 0.4053208 |
| **spatialbasis_7of15_epoch** | Predicted value of variable 7 (of 15) from a | 137274 | 0.7671203 | 0.6441075 | -0.9238448 | 1.778233 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | spatial curve function (df=15) | | | | | |
| **spatialbasis_8of15_epoch** | Predicted value of variable 8 (of 15) from a spatial curve function (df=15) | 137274 | 0.3587313 | 0.9330516 | -2.420441 | 1.994681 |
| **spatialbasis_9of15_epoch** | Predicted value of variable 9 (of 15) from a spatial curve function (df=15) | 137274 | -0.262951 | 0.9657267 | -1.740777 | 1.909506 |
| **spatialbasis_10of15_epoch** | Predicted value of variable 10 (of 15) from a spatial curve function (df=15) | 137274 | -0.826194 | 0.5619864 | -2.056876 | 0.2104288 |
| **spatialbasis_11of15_epoch** | Predicted value of variable 11 (of 15) from a spatial curve function (df=15) | 137274 | -0.4712007 | 0.8786682 | -1.893658 | 1.920593 |
| **spatialbasis_12of15_epoch** | Predicted value of variable 12 (of 15) from a spatial curve function (df=15) | 137274 | -0.7741244 | 0.6460149 | -1.939465 | 0.715452 |
| **spatialbasis_13of15_epoch** | Predicted value of variable 13 (of 15) from a spatial curve function (df=15) | 137274 | 0.5623494 | 0.824963 | -1.201134 | 3.027113 |
| **spatialbasis_14of15_epoch** | Predicted value of variable 14 (of 15) from a spatial curve function (df=15) | 137274 | 0.1238012 | 0.9930972 | -2.970319 | 1.530796 |
| **spatialbasis_15of15_epoch** | Predicted value of variable 15 (of 15) from a spatial curve function (df=15) | 137274 | -0.5944845 | 0.7888014 | -2.496078 | 0.8591589 |