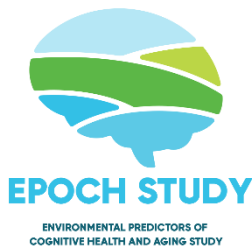


HEALTH AND RETIREMENT STUDY RESTRICTED DATA DOCUMENTATION

Individual-Level Greenspace Restricted Data

Version 1.0, released May 2025



Prepared by:

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

Jennifer D'Souza¹, Alan Fossa¹, 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 greenspace 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 greenspace (Individual-Level Greenspace, Version 1.0) estimates 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. 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

NDVI:

Vermote E, Justice C, Csiszar I, Eidenshink J, Myneni RB, Baret F, Masuoka E, Wolfe RE, Claverie M. NOAA CDR Program. (2014): NOAA Climate Data Record (CDR) of Normalized Difference Vegetation Index (NDVI), Version 6. Accessed December 2023. NOAA National Centers for Environmental Information. doi:[10.7289/V5PZ56R6](https://doi.org/10.7289/V5PZ56R6)

Didan, K. (2015). *MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006* [Data set]. NASA EOSDIS Land Processes Distributed Active Archive Center. Accessed 2025-01-09 from doi:[10.5067/MODIS/MOD13Q1.006](https://doi.org/10.5067/MODIS/MOD13Q1.006)

Parks:

Browning MHEM, Rigolon A, Ogletree S, et al. The PAD-US-AR dataset: Measuring accessible and recreational parks in the contiguous United States. *Sci Data* **9**, 773 (2022). doi:[10.1038/s41597-022-01857-7](https://doi.org/10.1038/s41597-022-01857-7)

Zanaga D, Van De Kerchove R, De Keersmaecker W, et al. ESA WorldCover 10 m 2020 v100. Zenodo (2021), doi:[10.5281/zenodo.5571936](https://doi.org/10.5281/zenodo.5571936)

Questions?

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

Data Description

Greenspace is used to characterize one aspect of exposure to nature and natural settings. It is hypothesized that greenspace can have important impacts on health through pathways including increased exercise, social engagement, and reduced stress. It can also provide protection from environmental contaminants and shade from sunlight and heat. To assess greenspace, the EPOCH study linked two different measures to the HRS survey using respondents' exact address locations throughout the study.

The first measurement was Normalized Difference Vegetation Index (NDVI) calculated using pre-processed images taken by the Moderate Resolution Imaging Spectroradiometer aboard NASA's Terra satellite (MODIS-Terra) obtained from Google Earth Engine ([Gorelick et al., 2017](#)). MODIS-Terra images are available at a 250 meter (m), 16-day resolution. Greater detail of these data sources and the methods used can be found in the citations above, but, briefly, the NDVI theoretically ranges from -1 to 1, with densely vegetated areas like forests approaching 1 and water, ice, pavement, bare soil, and roads close to or less than 0. We extracted the annual maximum NDVI to reflect the presence of vegetation around the home as a broader indicator of natural settings that can influence health beyond just greenness alone. This measure has an added benefit that it is less sensitive to the definition of season since the peak and length of growing seasons vary by place. As a secondary measure, we also extracted the annual minimum NDVI to indicate the variation in greenness across the year.

We estimated average annual maximum and minimum NDVI levels within 250 m and 1 km buffers of each participant's residential address at the time of their surveys. A 1 km distance was selected as a feasible walking distance for individuals over 50 years old ([Camarri et al., 2006](#)) within a normative 10 minute period ([Yang and Diez-Roux, 2012](#)). This duration has been adopted as a benchmark for park accessibility by U.S. policymakers and non-profit organizations ([10-Minute Walk—Improving Park & Green Space Access](#)).

As a second measure, we identified accessible parks near respondents' homes using the Parks and Protected Areas Database of the United States Accessible Recreation (PADUS-AR V2.1). The PAD-US-AR is a refined version of the US Geological Survey's national repository of public and private spaces that aims to identify publicly accessible places whose primary purpose is recreation ([Browning et al., 2022](#)). The version of the data linked to the HRS includes 248,871 recreational open spaces covering a total area of 1,866,564 km² across the United States. We calculated the total, grassy, tree-covered, and non-vegetated park area by the respective proportion of each landcover type as defined by the WorldCover 2020 dataset developed by the European Space Agency. The WorldCover 2020 dataset classifies land at a 1 m resolution, achieving an accuracy of 75% ([Zanaga et al., 2021](#)). For this exposure, we did not restrict our estimates to the park area within 1 km but rather identified all accessible parks within 1 km of a participant's home at the time of the survey and then assigned the participant the total, grassy, tree-covered, and non-vegetated area of those parks that intersected with their 1 km buffer.

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:

For NDVI: Fossa AJ, D'Souza J, Bergmans R, Zivin K, Adar SD. Residential greenspace and major depression among older adults living in urban and suburban areas with different climates across the United States. *Environ. Res.* 2024;243. doi: [10.1016/j.envres.2023.117844](https://doi.org/10.1016/j.envres.2023.117844).

For Parks: Fossa AJ, D'Souza J, Bergmans R, Zivin K, Adar SD. Different types of greenspace within urban parks and depressive symptoms among older U.S. adults living in urban areas. *Environ. Int.* 2024;192. doi: [10.1016/j.envint.2024.109016](https://doi.org/10.1016/j.envint.2024.109016).

Data Details

Spatial Resolution: The NDVI data have a 250 meter resolution whereas the parks data reflect exact locations.

Temporal Resolution: Data are generated to reflect the year of the survey (2000-2016), however we note that the parks dataset is extrapolated from information only reported in 2020. Therefore, the user should consider if this temporal resolution is sufficient for their scientific question of interest.

Data Generation: NDVI is generated by satellite imagery while the PAD-US-AR dataset is a curated dataset by a team of researchers and outdoor recreation experts. EPOCH linked these data to HRS participants based on residential addresses and survey dates.

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 park data available for those locations.

Units: None for NDVI, km² for park area

Variable Names: hhidpn

HHID

PN

wave

year

ndvi_max_1km_epoch

ndvi_max_250m_epoch

ndvi_min_1km_epoch

ndvi_min_250m_epoch

parks_total_area_1km_epoch

parks_tree_area_1km_epoch

parks_grassy_area_1km_epoch

parks_built_area_1km_epoch

Summary Statistics:

Variable	Description	N	Mean	Std Dev	Minimum	Maximum
hhidpn	HRS Participant ID: Household ID + Person Number (numerical as HHIDPN=1000*HHID+PN)	184263	227757152	233481887	2010	959738010
HHID	Household ID (6- character)	184263	NA	NA	000002	959738
PN	Person Number (3- character)	184263	NA	NA	010	043
wave	Wave-alphabetic, g thru p	184263	NA	NA	g	p
year	WAVE YEAR (2000-2016)	184263	2007.991	5.183	2000	2016
ndvi_max_1km_epoch	average annual maximum NDVI within a 1 km buffer	184249	0.793	0.144	-0.038	1.000
ndvi_max_250m_epoch	average annual maximum NDVI within a 250 m buffer	184056	0.673	0.169	-0.030	1.000
ndvi_min_1km_epoch	average annual minimum NDVI within a 1 km buffer	184249	0.062	0.124	-0.200	0.594
ndvi_min_250m_epoch	average annual minimum NDVI within a 250 m buffer	184058	0.165	0.154	-0.200	0.739
parks_total_area_1km_epoch	Total area of parks with boundaries within 1 km (km ²)	128927	122.662	1014.713	1.00E-06	21754.110
parks_tree_area_1km_epoch	Total tree-covered area of parks with boundaries within 1 km (km ²)	128927	38.499	336.429	0	7928.618
parks_grassy_area_1km_epoch	Total grassy area of parks with boundaries within 1 km (km ²)	128927	36.810	495.517	0	9244.085
parks_built_area_1km_epoch	Total non-vegetated area of parks with boundaries within 1 km (km ²)	128927	18.070	284.779	0	10243.500