

Gianattasio-Power Predicted Dementia Probability Scores and Dementia Classifications

This data file (hrs dementia_2021_1109.sas7bdat) contains predicted dementia probabilities and classifications for 2000-2016 HRS respondents aged 70+ with self-reported race/ethnicity non-Hispanic white, non-Hispanic black, or Hispanic, using three newly developed algorithms: a modified version of an algorithm originally developed by Hurd and colleagues¹ (Modified Hurd Model), a new expert-informed logistic model (Expert Model), and a new LASSO-reduced logistic model (LASSO Model). Algorithms were trained and evaluated using HRS data and data from all four waves of the Aging, Demographics, and Memory Study (ADAMS; <https://hrsdata.isr.umich.edu/data-products/aging-demographics-and-memory-study-adams-wave>), and achieve 77-83% sensitivity, 92-94% specificity, and 90-92% accuracy in overall out-of-sample performance.

The algorithms use different combinations of sociodemographic characteristics, health and physical functioning variables, social engagement indicators, and cognitive indicators (i.e. cognition test item scores and proxy-reports of cognition) to estimate a predicted dementia probability, which are then used to classify dementia status using race/ethnicity-specific probability thresholds. Each algorithm was developed to minimize differences in predictive performance across race/ethnicity groups, achieving pairwise differences of ≤ 3 percentage points for sensitivity and ≤ 5 percentage points for specificity, and are therefore adequate for use in race/ethnicity disparities research. Further details on the development and performance of the algorithms are available in our paper.²

This data file (hrs dementia_20211109.sas7bdat) was created using the 2018 RAND V1 HRS longitudinal file ("randhrs1992_2018v1") and core HRS data; code for reproducing this dataset is available in the following Github repository, and is dated 2021_1109.

Note that there are small differences in the probabilities and classifications for all years in this dataset compared to the previously distributed dataset (hrs dementia_20191028.sas7bdat) due to differences in the source data. This previously distributed dataset covered 2000 to 2014, and was created using the 2014 RAND HRS longitudinal V2 file ("randhrs1992_2014v2") and core HRS data; code for reproducing this prior version of the dataset is available in the following Github repository and is dated 2019_0529: https://github.com/powerpilab/AD_algorithm_development.

Variables list

- **HHID**: HRS household ID number
- **PN**: HRS person number
- **hrs_year**: the survey year from which predictions are made
- **expert_p**: predicted probability of dementia using the Expert Model
- **expert_dem**: dementia classification (0=no, 1=yes) using Expert Model
- **LASSO_p**: predicted probability of dementia using the LASSO Model
- **LASSO_dem**: dementia classification (0=no, 1=yes) using LASSO Model
- **hurd_p**: predicted probability of dementia using the Modified Hurd Model
- **hurd_dem**: dementia classification (0=no, 1=yes) using Modified Hurd Model

Please note that the authors are not responsible for errors resulting from the use of this dataset or referenced SAS code.

This work was funded by the National Institute on Aging, grant R03 AG055485, awarded to Dr. Melinda C. Power.

References

1. Hurd MD, Martorell P, Delavande A, Mullen KJ, Langa KM. Monetary Costs of Dementia in the United States. *N Engl J Med*. 2013;368(14):1326-1334. doi:10.1056/NEJMsa1204629
2. Gianattasio KZ, Ciarleglio A, Power MC. Development of algorithmic dementia ascertainment for racial/ethnic disparities research in the U.S. Health and Retirement Study. *Epidemiology*. 2020;31(1):126-133. doi:10.1097/EDE.0000000000001101