Prevalence of Dementia and Mild Cognitive Impairment in the Health and Retirement Study: Figures, Tables, and Appendices

Richard N. Jones

Jennifer J. Manly

Douglas Tommet

Lindsay Kobayashi

Lindsay H. Ryan

Deborah A. Levine

Ryan J. McCammon

Steven G. Heeringa

David Weir

Kenneth M. Langa

2025-07-22

|  |
| --- |
| Figure 1: Algorithm for classifying dementia and mild cognitive impairment in the Health and Retirement Study |

[Figure 1](#fig-figure1) caption. The algorithm for classifying dementia and mild cognitive impairment in the core Health and Retirement Study (HRS). Cell entries for sample size refer to a subsample of HRS/HCAP participants who completed all of the HCAP cognitive tests. Severe and Moderate cognitive impairment are defined using an estimated performance score derived from a single factor model fit to Core cognitive tests, using impairment thresholds to match prevalence estimates from the Harmonized Cognitive Assessment Protocol (HCAP) substudy (Manly et al., 2022). ADL/IADL limitations are assessed using core measures, and any impairment in any assessment item defines impairment. Self-reported concerns about memory or thinking are assessed with two core items asking about change from 2 years ago and current functioning. For HRS participants who do not complete the cognitive assessment, a study partner (usually a spouse) is asked about changes in cognitive performance using Jorm’s Informant Questionnaire (IQCODE). Thresholds for impairment on the IQCODE are those used in Manly et al. (2022).

If this paper is about the HRS/Core cognitive classification, then the cell sample sizes based on N = 2993 are fine. However, if this paper is about the prevalence of dementia and MCI in the HRS, then the sample sizes should be based on the full HRS sample. This version of the flow diagram could then be included in appendix material that provides more detail on the derivation of the Core classification and relates it to the HCAP classification.

The text in red highlights differences from previous presentation, and will be removed in the final version.

# A tibble: 2 × 2  
 PA019\_group n  
 <chr> <int>  
1 <65 10940  
2 >=65 9972

| Table 1. Participant Characteristics | | |
| --- | --- | --- |
| Characteristic | HRS 2016 wave (N = 20,912) | HCAP subsample (N = 2,993) |
| Sex (N (%)) |  |  |
| Female | 12,245 (58.6%) | 1,795 (60.0%) |
| Male | 8,664 (41.4%) | 1,198 (40.0%) |
| Blank, partial interview | 3 (0.0%) |  |
| Age in years (Mean (SD)) | 65.7 (11.8) | 75.7 (7.5) |
| Race (N (%)) |  |  |
| White | 13,822 (66.1%) | 2,389 (79.8%) |
| Black | 4,532 (21.7%) | 455 (15.2%) |
| All other groups | 2,477 (11.8%) | 148 (4.9%) |
| Not collected | 78 (0.4%) | 1 (0.0%) |
| Blank, partial interview | 3 (0.0%) |  |
| Hispanic origin (N (%)) |  |  |
| Not Hispanic | 17,441 (83.4%) | 2,674 (89.3%) |
| Hispanic | 3,427 (16.4%) | 318 (10.6%) |
| Not obtained | 41 (0.2%) | 1 (0.0%) |
| Blank, partial interview | 3 (0.0%) |  |
| Years of education (Mean (SD)) | 12.8 (3.3) | 12.8 (3.2) |

Notes: The HCAP subsample includes 2,993 participants from the HRS 2016 wave. The full HRS/HCAP sample included 3,496 participants. The subsample of 2,993 we use includes those HCAP participants who participated in the neuropsychological assessment portion sufficiently such that none of their 5 derived cognitive domain scores used in the dementia and MCI classification algorithm had to be imputed. We restrict our HCAP comparator sample to those without imputed data to ensure that imputation error does not compound classification error. Race and ethnicity data were gathered via self-report at the time of first interview and are considered to be markers of exposure to evolving systems of racism, not as a proxy for genetic variation or any other biological variable. Race was self-selected by participants at the time of the first interview from a list of options defined by the 2000 US Census criteria. Other race includes a pooled group of participants who identified as American Indian or Alaska Native, Asian, Native Hawaiian or Pacific Islander, or another self-described race, consolidated due to small sample sizes and risk of identification.

### Table 2. Summary of cognitive, functional, and self-report variables in the HRS 2016 subsample sample (n = 2993)

### Table 3.1 Diagnostic Comparison: HRS/Core Actuarial Classification versus HCAP Consensus Panel diagnosis (N=50)

Cell entries are expected percentages of the total HRS/HCAP sample.

|  | **Consensus** |  |  |  |
| --- | --- | --- | --- | --- |
| **HRS Actuarial** | **Dementia** | **MCI** | **Normal** | **Total** |
| Dementia | 3.3 | 0.9 | 0.4 | 4.7 |
| MCI | 0.2 | 5.4 | 19.9 | 25.5 |
| Normal | 0.4 | 16.7 | 52.6 | 69.8 |
| **Total** | 4 | 23.1 | 72.9 | 100 |

Note: The validation sample over-represents persons with high diagnostic uncertainty. Table 3 accounts for the sample selection probabilities to generate agreement statistics that would be expected if we had drawn a simple random sample of HRS/HCAP participants to the validation study.

The weighted kappa statistic is 0.33, indicating a “poor” level of agreement between the HRS/Core Actuarial classification and the HCAP Consensus Panel diagnosis.

From: Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological assessment, 6(4), page 286:

| R | Agreement |
| --- | --- |
| <0.4 | “Poor” |
| 0.40 - 0.59 | “Fair” |
| 0.60 - 0.74 | “Good” |
| 0.75 - 1.00 | “Excellent” |

**But the “problem” is MCI**

The agreement (unweighted Cohen’s kappa) for dementia vs MCI or Normal is 0.77 indicating a “excellent” level of agreement between the HRS/Core Actuarial classification and the HRS/HCAP validation sample consensus panel classification.

By way of comparison, the HRS/HCAP Actuarial classification of Dementia vs MCI or Normal, relative to the HRS/HCAP validation subsample consensus diagnosis of Dementia vs MCI or Normal, has agreement characterized by a kappa statistic of 0.70 (Farron et al 2025).

|  | **Consensus** |  |  |
| --- | --- | --- | --- |
| **HRS/HCAP** | **Dementia** | **MCI or Normal** | **Total** |
| Dementia | 4.4 | 3.3 | 7.7 |
| MCI or Normal | 0.2 | 92.0 | 92.2 |
| **Total** | 4.6 | 95.3 | 99.9 |

Note: Results summarized from Farron et al 2025.

### Table 3.2 Diagnostic Comparison: HRS/Core Actuarial Classification versus HRS/HCAP Actuarial Classification (N=2,992)

Cell entries are observed percentages of the total HRS/HCAP subsample.

|  | **HRS/HCAP** |  |  |  |
| --- | --- | --- | --- | --- |
| **HRS Actuarial** | **Dementia** | **MCI** | **Normal** | **Total** |
| Dementia | 4.2 | 2.9 | 1.8 | 9 |
| MCI | 3.3 | 6.4 | 7.2 | 16.8 |
| Normal | 1.9 | 12.5 | 59.8 | 74.2 |
| **Total** | 9.4 | 21.8 | 68.8 | 100 |

Note: cell entries are do not take into account sampling weights.

The weighted kappa statistic is 0.51, indicating a “fair” level of agreement between the HRS/Core Actuarial classification and the HCAP actuarial classification.

The agreement (unweighted Cohen’s kappa) for dementia vs MCI or Normal is 0.4 indicating a “fair” level of agreement between the HRS/Core Actuarial classification and the HRS/HCAP actuarial classification.

### Table 3.3 Diagnostic Comparison: HRS/Core Actuarial Classification versus HRS Langa-Weir Classification (N=2,992)

Cell entries are observed percentages of the total HRS/HCAP subsample.

|  | **Langa-Weir** |  |  |  |
| --- | --- | --- | --- | --- |
| **HRS Actuarial** | **Dementia** | **MCI** | **Normal** | **Total** |
| Dementia | 3.6 | 3.8 | 1.5 | 9 |
| MCI | 2.1 | 7.5 | 7.2 | 16.8 |
| Normal | 1.2 | 9.2 | 63.8 | 74.2 |
| **Total** | 6.9 | 20.5 | 72.6 | 100 |

Note: cell entries are do not take into account sampling weights.

The weighted kappa statistic is 0.57, indicating a “fair” level of agreement between the HRS/Core actuarial classification and the HRS Langa-Weir Classification.

The agreement (unweighted Cohen’s kappa) for dementia vs MCI or Normal is 0.41 indicating a “fair” level of agreement between the HRS/Core actuarial classification and the Langa-Weir classification.

**Table 4. Group Differences in Prevalence of Dementia and Mild Cognitive Impairment (MCI) Among Health and Retirement Study (HRS, 2016, N = 20,912)**

## Appendix: Derivation and operationalization of the HRS/Core Actuarial Cognitive Classification

### Overview

* Adapted version Jak/Bondi approach to MCI and dementia classification, using cognitive data, informant reports, functional data, and self-reports. Similar to how implemented in HRS/HCAP (Manly et al 2022). Key features include:
  + Cognitive performance normalized with respect to internal reference group and standardized with respect to demographic factors (age, sex, race, ethnicity, education, and their two-way interactions).
  + Dementia classification requires cognitive impairment and functional impairment
  + MCI classification requires cognitive impairment that is either severe but without marked functional impairment, or moderate with functional impairment or self-concerns. Moderate impairment without functional impairment and without self-concerns is considered normal cognition.
* Our procedure had the following ***similarities*** with respect to Manly et al (2022)
  + We use a subsample of the HRS/HCAP sample to derive the classification (N=2993), defined as persons who had none of their HRS/HCAP cognitive domains imputed.
  + We use the same normative reference group as Manly et al (2022) to standardize cognitive performance.
  + We compare to the same reference standard (consensus diagnoses in N=50 subsample, over sampled of persons with high diagnostic uncertainty).
* Our operationalization had the following ***differences*** with respect to
  + Instead of 5 cognitive domains, we measure a single cognitive domain (global cognition) using the cognitive measures available in the Core in the 2016 wave. However, the cognitive impairment thresholds (after normalization and standardization) are selected to match the marginal distributions of impairment categories observed in the HRS/HCAP sample: these end-up being at T-score values of 36.4 and 43.8 , about the 9th and 27th percentiles of the HRS/HCAP sample.
  + Whereas Manly et al (2022) used the Jorm and Blessed to identify functional impairment, we only had access to the Core HRS ADL/IADL items. We attempted to match marginal distributions of functional impairment, but had little flexibility in choosing thresholds. Our thresholds for impairment are between no ADL/IADL difficulties and any ADL/IADL difficulties, and we still under-identify functional impairment relative to what we were able to identify using the HRS/HCAP measures.
  + HRS Core does have access to information from the Jorm, but this is only available for those participants that did not participate in the cognitive assessment. We use the Jorm to classify persons who did not participate in the cognitive assessment, but we do not use it to classify persons who did participate in the cognitive assessment (as we do in HRS/HCAP).
  + We use two self-report items to identify cognitive concerns, PD101 and PD102. In HCAP, we only used PD102. HRS/HCAP only considers self-concerns of worsening (PD102), the Core algorithm also includes self-report of poor functioning.

The appendix ***will*** include details on the specific cognitive variables used, the results of factor analysis, standardization and normalization, threshold identification. This information is currently available [at this link](https://quantsci.s3.amazonaws.com/Work/HCAP/HRS_cognition_2025-07-22.html).

### Appendix 2: HRS 2016 subsamples

|  |
| --- |
| Figure 2 |

Thanks for reading.