Prevalence of Dementia and Mild Cognitive Impairment in the Health and Retirement Study

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

Abstract

**Importance:** Dementia and mild cognitive impairment (MCI) are common in older adults, but their prevalence is not well understood and challenging to assess in large population-based studies. **Objective:** Describe the prevalence of MCI and dementia in the Health and Retirement Study (HRS), using a novel approach to classifying dementia informed by the recently completed wave 1 of the Harmonized Cognitive Assessment Protocol (HCAP) study. **Design:** Cross-sectional analysis of the 2016 HRS Core sample, validation using existing consensus panel classifications. **Main Outcomes and Measures:** The prevalence of dementia and MCI is XYZ and PDQ. Our algorithm shows poor agreement with consensus standard classifications (weighted kappa = 0.33), but this is due to difficulties in reaching agreement regarding mild cognitive impairment, and the agreement statistic is excellent for dementia versus less impaired cognition (kappa = 0.77). **Conclusions:** The new algorithm shows very promising validity evidence for dementia classification in the HRS. The accurate classification of MCI remains a challenge in field research settings.

# Outline

## Introduction

* Overview of dementia, mild cognitive impairment (MCI), and field studies
* Health and Retirement Study (HRS) approaches:
  + Langa-Weir: count-and-cut method
  + Hurd & Hudomiet: predictive models
  + HCAP: actuarial classification[[1]](#footnote-20)
* This study develops a new actuarial classification using only Core HRS measures, following Manly et al. (2022)
* We describe:
  + The derivation and validation of the new classification
  + The prevalence of dementia and MCI in the Core HRS sample

## Methods

* Sample: HRS & HCAP subsample (N = 2,993)
  + Rationale and construction of the HRS/HCAP Normative Reference Group
  + Validation subsample within HRS/HCAP
* Cognitive and functional measures
* Brief description of derivation of actuarial algorithm (full details in Appendix 1)
* Methods for assessing agreement with comparator measures
* Prevalence estimation procedures

## Results

* **Figure 1**: Schematic of algorithm
* **Table 1**: Participant characteristics (N = 20,912; subsample N = 2,993)
* **Table 2**: Descriptive statistics (cognitive, functional, self-report, proxy measures)
* **Table 3**: Reliability and validity
  + HRS/HCAP validation sample (N = 50), 3-category weighted kappa: κ = 0.33
  + HRS/HCAP actuarial classification (N = 2,992): κ = 0.51
  + HRS Langa-Weir classification (N = 2,992): κ = 0.57
  + Hurd model (N = 2,992): [to be added]
  + Hudomiet model (N = 2,992): [to be added]
  + Dementia vs. non-dementia classification (validation sample, N = 50): κ = 0.77
* **Table 4**: Participant characteristics by diagnostic group

## Discussion

* Three-way agreement (normal, MCI, dementia) is poor-to-fair across classifications
* Binary agreement for dementia diagnosis with HRS/HCAP consensus panel is excellent (κ = 0.77)
* Comparison with previous methods: Langa-Weir, Hurd, Hudomiet, Graves (NACC), Kasper (NHATS/ADAMS)
* Strengths:
  + Large representative sample
  + Rigorously defined normative reference group
  + Access to a reference standard diagnosis in a field study is rare
* Limitations:
  + Small validation sample (N = 50)
* Implications for future research

## Appendices

* Appendix 1: Derivation of the actuarial algorithm

**How well do we do compared to reference standard diagnoses × other classification systems?**

This is for the discussion. The comparisons below tell me that *the HRS/HCAP actuarial classification as reported in Farron et al (2025) is really good* (very adequate for research (>=.60 target); almost good enough for diagnostic decision making (>=0.80 target); Farron et al (2025) report a confidence interval of 0.49 to 0.91, so adequacy for diagnostic decision making is not out of the question for the HRS/HCAP algorithm), but that is not the point of this paper.

For this paper, the **HRS Core** actuarial classification, the 3-way, with weighted kappa of 0.33, is not good at all relative to other algorithms. But, the classification of dementia vs MCI or Normal is excellent in the HRS Core, ***so far the best we have seen*** among field survey validation exercises, with a weighted kappa of 0.77, and positive predictive value of 0.70, and sensitivity of 0.83. Still forthcoming is the comparison with the Langa-Weir classification, and the Hurd and Hudomiet classifications. Regardless, the HRS Core actuarial classification is not very good at classifying MCI, and the problem is distinguishing MCI from Normal, not distinguishing MCI from Dementia. ***Maybe including PD101 in the self-concerns was a mistake, we’ll have to look at that.*** See Table 3.1 in the tables and figures attachment.

| **Source** | **Sample, N** | **Classification** | **Criterion** | **Statistic** |
| --- | --- | --- | --- | --- |
| **3-way** |  |  |  |  |
| Farron et al (2025) | HRS/HCAP validation (N=50) | Normal, MCI, Dementia (actuarial) | HCAP Consensus Panel | =0.75 |
|  |  |  |  |  |
| Kasper et al (2013) | ADAMS subsample (N=121) | Probable, Possible, Unlikely Dementia (NHATS algorithm) | ADAMS Consensus diagnoses | =0.67 |
|  |  |  |  |  |
| Graves et al (2020) | NACC (N=1524) | Normal, MCI, Dementia (actuarial) | NACC Consensus diagnoses | =0.54 |
|  |  |  |  |  |
| This paper (2026) | HRS Core (HCAP validation, N=50) | Normal, MCI, Dementia (actuarial) | HRS/HCAP Consensus Panel | =0.33 |
|  |  |  |  |  |
| This paper (2026) | HRS Core (HCAP validation, N=50) | Normal, MCI, Dementia (Langa-Weir) | HRS/HCAP Consensus Panel | =? |
|  |  |  |  |  |
| **2-way (dementia)** |  |  |  |  |
| This paper (2026) | HRS Core (HCAP validation, N=50) | Dementia (actuarial) | HRS/HCAP Consensus | =0.77 |
|  |  |  |  | PPV=0.70 |
|  |  |  |  | SN=0.83 |
|  |  |  |  |  |
| Graves et al (2020) | NACC (N=1524) | Dementia (actuarial) | NACC Consensus | =0.72 |
|  |  |  |  | PPV=0.43 |
|  |  |  |  | SN=0.96 |
|  |  |  |  |  |
| Farron et al (2025) | HRS/HCAP validation (N=50) | Dementia (actuarial) | HCAP Consensus | =0.70 |
|  |  |  |  | PPV=0.57 |
|  |  |  |  | SN=0.96 |
|  |  |  |  |  |
| Kasper et al (2013) | ADAMS subsample (N=121) | Probable dementia (NHATS algorithm) | ADAMS Consensus | =0.53 |
|  |  |  |  | PPV=0.69 |
|  |  |  |  | SN=0.66 |
|  |  |  |  |  |
| This paper (2026) | HRS Core (HCAP validation, N=50) | Dementia (Langa-Weir) | HRS/HCAP Consensus | ? |
|  |  |  |  |  |
| This paper (2026) | HRS Core (HCAP validation, N=50) | Dementia (Hurd) | HRS/HCAP Consensus | ? |
|  |  |  |  |  |
| This paper (2026) | HRS Core (HCAP validation, N=50) | Dementia (Hudomiet) | HRS/HCAP Consensus | ? |

The details of this table may need to be it’s own appendix. The data from Kasper et al (2013) and Graves et al (2020) are modified from their as-published form. Neither of these authors provide 3-way agreement statistics or confusion matrix. Graves was not a representative sample. I re-constructed 3-way agreement tables (confusion matrix) from the data provided by these authors, and for both re-weighted to estimated population marginal distributions for Normal, MCI, and Dementia with respect to their study’s reference standard diagnostic classification.

Thanks for reading.

1. Jak and Bondi use the term “actuarial diagnosis” to describe a rule-based approach to classifying cognitive impairment based on objective neuropsychological test performance. The “actuarial” qualifier is used to set the method apart from classifications that rely on clinical judgment or consensus diagnosis, and emphasizes the method applies explicit decision rules that can be replicated and automated. This approach improves transparency and reliability, especially in research and population-based settings (Jak et al., 2009; Bondi & Jak, 2014). [↑](#footnote-ref-20)