The Lifetime Risk of Living with Dementia for Six Months, One, Two, or Five Years

Péter Hudomiet, PhD
RAND

Michael D. Hurd, PhD
RAND, NBER, NETSPAR

Susann Rohwedder, PhD
RAND, NETSPAR

Corresponding author: Michael D. Hurd, RAND, 1776 Main Street, Santa Monica, CA-90401, Tel: (310) 393-0411 x6945, Email: mhurd@rand.org.

Summary

Background

Dementia prevalence approaches 40% at advanced old age, but that high rate is not informative about the lifetime risk of dementia.. Yet individuals, policy makers, health care providers and insurance companies need to know the lifetime risk of dementia and its duration because it can be very costly depending on its duration. Due to substantial data requirements, only a limited number of research studies have estimated this risk, and they show conflicting results. This study produces nationally representative estimates of the lifetime risk of living with dementia for six months, one, two, or five years in the U.S. by gender, race, ethnicity, age-at-death, health, and socio-economic status.

Methods

The study sample includes 8,733 individuals who died after age 70 between 1998 and 2014 in a large, longitudinal, nationally representative U.S. survey, the Health and Retirement Study (HRS). A subsample of the HRS has undergone a clinical assessment for dementia (n=856). Using the dementia diagnosis from the subsample, and cognitive, functional status and health predictors in the larger sample, probabilities of dementia were estimated for the full study sample. Then a longitudinal latent variable model was fit on the wave-by-wave probabilities to estimate the probability of dementia at several discrete times prior these individuals' death. The average of these probabilities by subgroups provides unbiased estimates of the average risk by the groups.

Findings

43·2% of those who died after age 70, had dementia six months before death. 38·6%, 31·7%, and 18·7% had dementia one, two, and five years before death, respectively. The risk was higher for females, the less educated, African Americans, those with lower lifetime earnings, and those who died at older ages. Comorbidities decreased the risk except for stroke and psychiatric problems, which increased it.

Interpretation

The lifetime risk of dementia is higher than found in earlier U.S. studies, and poses a very significant risk to individuals and families. The lifetime risk of dementia is a health disparity in that the risk is higher for those with lower socio-economic status.

Funding

National Institute on Aging.

Introduction

Dementia is one of the most expensive diseases in the United States and many other countries, with growing financial and non-financial impacts on older adults, their families, and government programs.[1-3] Living with dementia for an extended period of time is a significant risk for middle-aged individuals and their families who have to make financial and non-financial arrangements about their care needs at older ages, as well as for policy makers, health care providers, and insurance companies that provide and finance care for these individuals.

Providing nationally representative estimates of the lifetime risk and the duration of dementia is complicated by many factors, such as the difficulty of accurately measuring dementia in large population-representative surveys;[1] properly accounting for the correlation between the risk of dementia and mortality;[4-9] and the need for longitudinal follow-up observations. Consequently, there is only a limited number of research studies in the U.S.[10-13] and in other countries[14,15] with conflicting results.

The Health and Retirement Study (HRS) provides an unparalleled opportunity to estimate the lifetime risk and the duration of dementia: it is a large, nationally representative and longitudinal dataset (~20,000 individuals per wave); it has a wide range of information about individuals' socio-economic status (SES), health, and cognitive state[16] ; and a subsample of 856 individuals went through a clinical assessment for dementia within the Aging, Demographics and Aging Study (ADAMS), which enables properly calibrating the cognitive variables in the HRS. The HRS has been used to estimate the prevalence of dementia in the U.S.,[1,17-19] but this is the first study to estimate the *lifetime* risk and the duration of dementia using the HRS. Further, because of the richness of the HRS data, the study provides estimates of the risk and duration by subpopulations, addressing whether dementia is a health inequality aligned with socio-economic status. High SES persons have better age-adjusted health including dementia; but

they are more likely to live to advanced ages where dementia is highly prevalent. Thus, it is not *a priori* evident whether high SES persons have a lower or a high lifetime risk of dementia.

Methods

Data and Study Sample

The HRS is a nationally representative, biennial, longitudinal survey of the U.S. population age 51 or older. It is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan. The survey collects a wide range of information about participants, and it is unusually detailed about health and cognitive ability compared with other large, general, social science surveys.

The HRS began in 1992. New cohorts of 51-56-year-old individuals have been added every six years to maintain population representativeness. Individuals are followed into nursing homes, which is crucial to estimate accurately dementia prevalence.[20,21] Whenever a study member is unable or unwilling to participate in a survey wave, HRS attempts to interview a proxy respondent (typically a spouse or a child). Around half of the sample with dementia is interviewed by proxy.[1]

Our study period is the 1998 through 2014 survey waves, and our main study sample is the 8,733 individuals who died after the age of 65 during the study period. Vital status and the date of death are primarily identified from the linked National Death Index. When the linked data were not available, we used proxy reports of vital status from the survey data. When properly weighted, the HRS provides a population representative sample of deaths during the study period. In addition to the study sample, we used data on as many as 21,462 individuals aged 65 or older year old to estimate several auxiliary variables such as a measure of lifetime earnings.

Table 1 shows descriptive statistics for the study sample. The first two columns show unweighted counts and fractions; and the last column shows weighted fractions using the survey weight from the last core HRS interview. The weighted and unweighted fractions are similar except that the unweighted fractions of blacks and Hispanics are higher than the weighted fractions because the HRS oversamples minorities to increase statistical precision by race. 54% of the sample are female; 88% are white; 10% are black and about 5% are Hispanic. More than two thirds have high school education or less; 60% died single; and 79% died between ages 75 and 95. About two thirds of the sample reported at least once in the HRS to have had hypertension and arthritis; about half reported a heart problem; a quarter had diabetes or cancer; and about a fifth reported a lung disease, stroke, or psychiatric problems at least once in the HRS.

<Table 1 about here>

Because the HRS core lacks a direct measure of dementia status, 856 HRS respondents of age 70 or older underwent in 2001-2003 a detailed in-home clinical assessment for dementia as part of ADAMS. The duration of the assessment was 3-4 hours, and it concluded in a final diagnosis of 1) dementia; 2) cognitively impaired but not dementia (CIND); or 3) normal cognitive function.[22]

We used an extensive set of cognition measures to predict dementia status in the full HRS sample. A five-point subjectively assessed memory scale (varying from excellent to poor) was available in both self and proxy interviews. Other cognition variables differed by interview type. For self-interviews, we used available measures adapted from the Telephone Interview for Cognitive Status: a 10-word immediate recall test, a 10-word delayed recall test, a serial-7 subtraction test (subtracting 7 five times starting from 100), backward counting (from 20 to 10), question on date of the interview, the name of the president, and being able to name “cactus” and “scissor” after paraphrasing. For proxy interviews, we used two cognition measures. Jorm IQCODE is a 16-item battery about recent improvements/declines in various aspects of memory, all using 5-point scales from (1) “memory much improved” to (5) “memory much

worse." We took the person-specific mean of the 16 items. Interviewers also assessed whether cognitive limitations were a reason for conducting a proxy interview.

We used two measures of Socioeconomic Status (SES). The first is an individual's highest level of education (less than high school, high school, some college, college or more). Second, we classify individuals into quintiles of an index of lifetime earnings of the household. (see Appendix A2 for details).

To increase the precision of the dementia probabilities, we used several additional demographic variables as predictors of dementia status: gender, age, race (white or non-white), ethnicity (Hispanic or non-Hispanic), and marital status. We also used the following list of health variables: self-rated health (5-point scale), number of limitations in activities of daily living (from zero to five), number of limitations in instrumental activities of daily living (from zero to five), CES-D depression score (from zero to eight), and ever having been diagnosed with hypertension, diabetes, cancer, lung disease, heart problems, stroke, psychiatric problems, or arthritis. Some of these predictor variables had missing values, which were imputed by logical or single imputations models. Sections A1 and A2 in the online appendix provide details.

We used the RAND HRS variables whenever available.[23] The RAND HRS Data file is an easy to use longitudinal data set based on the most commonly used HRS variables. It was developed at RAND with funding from the National Institute on Aging and the Social Security Administration.

Statistical analysis

We developed a two-step procedure to estimate the probabilities of living with dementia for various lengths of time. In step 1 we estimated the probabilities of dementia for each individual in the main sample in each survey wave. We estimated ordered probit models of the ADAMS diagnosis (dementia, CIND, normal) over the 856 ADAMS participants as functions of HRS predictors, and then predicted the probability of dementia to the full HRS sample in all waves. The HRS predictor variables were cognition, demographics, functional limitations, health, and the time-to-death (defined as the number of months

before death). Because different sets of predictor variables are available in self and proxy interviews, we estimated two separate models by interview type. This procedure is similar to the one used by Hurd et al. (2013), but we also included measures of health and mortality (time-to-death) to improve the predictions of dementia status in the sample of individuals close to death. The step 1 dementia probabilities were estimated for all individuals who died after age 65 (N =9,479) in each survey wave. We used the age 65 rather than the age 70 age cutoff of the main sample because we want to estimate the dementia probabilities up to five years before death. The step 1 probabilities are cross-sectional estimates, because they are only functions of the predictor variables from the contemporaneous HRS wave. The appendix discusses further details of the procedure and it shows the outcomes of the models.

In step 2, we fitted a flexible, longitudinal statistical model on the step 1 dementia probabilities. The model included gender, education, race, ethnicity, flexible functions of age, time-to-death, and a stochastic Brownian motion. Then, we used this model to predict the probabilities of dementia for each individual at 6, 12, 24, and 60 months before death. These estimates show the probabilities of dementia as functions of the model parameters, the step 1 dementia probabilities (from all survey waves), demographics, and individuals' current age and age-at-death. Figure 1 illustrates the approach, and the appendix provides more details, including the output of the model and diagnostics.

<Figure 1 about here>

We emphasize probabilities at least six months before death rather than at death to avoid confusing dementia with terminal cognitive decline.[24] Six months is a meaningful length of time to live with dementia, and it is comparable to dementia criteria Seshadri et al. (1997) used in the Framingham Study.

To obtain population probabilities of dementia and hence risk, we calculated weighted averages of the estimated probabilities using the HRS survey weights from the last core HRS interview.

We tested the validity of this statistical methodology on synthetic data with realistic parameter values (See the supplementary appendix). We developed a bootstrap procedure to provide standard errors, as discussed in the appendix.

Results

We first show lifetime risk estimates by population subgroups in Figures 2-7. Table A10 in the appendix shows all the underlying numbers.

Figure 2 shows the percentage of people who live with dementia for at least 6, 12, 24, and 60 months before death, and the 95% confidence intervals that reflect sampling variation and the uncertainty of the step 1 and step 2 dementia probability models. (Table A10 in the appendix shows all the underlying numbers) The uncertainty coming from the step 1 model was non-trivial, because of the relatively small sample size of the HRS-ADAMS subsample that was used to estimate the step 1 dementia probabilities.

We found that 18·7% (CI: 12·4% to 25%) of the sample had dementia five years prior to their deaths, 31·7·% (CI: 26·7% to 36·7%) had dementia two years before deaths, 38·6·% (CI: 33·5% to 43·7%) had dementia one year before death, and 43·2·% (CI: 37·4% to 48·9%) had dementia six months before death.

<Figure 2 about here>

Figure 3 shows the same risks by gender. These estimates are not adjusted by gender specific mortality differences or any other factors, which will be done later. 23·7% (CI: 16·5% to 31·0%) of females lived with dementia for at least five years, while only 12·5% (CI: 6·3% to 18·6%) of males did. The gender difference was statistically significant at 1%. (Our modeling procedure produces more precise estimates of group differences than of group levels: group differences are primarily identified by group differences in the dementia predictor variables, which are available in the full sample, whereas group levels are primarily identified in the relative small sample of ADAMS) The gender differences in the risks two, and one year before death are slightly smaller, but still statistically significant. The gender difference six

month prior to death is not statistically significant at 5%, but the point estimates are still substantial: 46·4% (CI: 39·6% to 53·3%) for females and 39·1% (CI: 31·8% to 46·5%) for males.

<Figure 3 about here>

The differences in the dementia probabilities between whites and blacks at the four time points are all statistically significant at 1%. Figure 4 shows that 26·8% (CI: 19·1% to 34·6%) of blacks lived with dementia for at least five years, while only 17·8% (CI: 11·4% to 24·2%) of whites did. The racial differences are even wider closer to death. Six months prior to death, for example, 54·1% (CI: 47·3% to 61·0%) of blacks and 42% (CI: 35·9 to 48·0%) of whites had dementia.

<Figure 4 about here>

The differences in the dementia probabilities between Hispanics and non-Hispanics are small and not statistically significant (Figure 5).

<Figure 5 about here>

Figure 6 shows the risks of living with dementia for five years, and for six months by the highest level of education of the sample members. The risks are significantly higher for high school dropouts compared to the other three education groups. Namely, high school dropouts have a 23·3% chance (CI: 17·1% to 29·6%) of living with dementia for at least half a year. The corresponding numbers for high school graduates, college dropouts, and college graduates are 17·1% (CI: 10·4% to 23·9%), 14·6% (CI: 6·8% to 22·3%), and 13·2% (CI: 4·0% to 22·3%), respectively. The differences in the dementia probabilities six months prior to death are slightly higher.

<Figure 6 about here>

Figure 7 demonstrates that lifetime earnings negatively correlate with the probabilities of dementia six months and five years before death. The association is strong and monotonic in the bottom four earnings

quintiles, and levels out in the top two. The probability of living with dementia for 5 years is 25·1% (CI: 19·0% to 31·2%) in the bottom earnings quintile, while it is 15·1% (CI: 8·5% to 21·7%) in the top quintile. The corresponding numbers for the dementia probability six months before death are 50·3% (CI: 44·8% to 55·8%) and 38·3% (CI: 31·5% to 45·0%).

<Figure 7 about here>

Table 2 shows OLS regressions of the probability of living with dementia for at least six months and for at least five years on demographics, SES, age at death, and co-morbidities. The differences by gender, race, ethnicity, education, and lifetime earnings are qualitatively similar to the unadjusted differences in Figures 2-7. Females have a significantly higher chance of living with dementia for five years, but the gender differential is small and not statistically significant six months prior to death. Blacks are more likely to live with dementia for six months and for five years than whites, and the differences between whites in individuals of other race (including Asians) are small and not statistically significant. There is no difference in the dementia probabilities of Hispanics and non-Hispanics. The differentials by education and lifetime earnings are large and statistically significant.

Individuals who die at older ages face substantially higher chances of living with dementia. For example, those who die after age 95 have 21·5 percentage points higher chance (CI: 14·1 to 28·9) of living with dementia for five years compared to those who die between age 70 and 75. The corresponding differential six months prior to death is 48·0 percentage points (CI: 34·5 to 61·5).

Co-morbidities tend to decrease the chance of dying with dementia, probably because comorbidities increase the chances of dying prior to developing dementia. We found, however, that strokes increase the chances of living with dementia for five years by 20·0 percentage points (CI: 14·5 to 25·5), likely due to a causal channel. Strokes also increase the chance of living with dementia for six months by 29·2 percentage points (CI: 23·1 to 35·3). Having been diagnosed with psychiatric problems also increased the

chances of dying with dementia. It is unclear if this latter correlation reflects a causal channel or if having a psychiatric problem results from dementia or if both pathways are in operation.

<Table 2 about here>

Discussion

A large fraction of the U.S. population goes through a phase of living with dementia before they die. In the sample of HRS participants who were observed to die at an age above 70, 43·2% lived with dementia for at least half a year, 38·6% lived with dementia for at least a year, 31·7% lived with dementia for at least two years, and 18·7% lived with dementia for at least five years. We found large differences across demographic groups. Females are more likely to live with dementia especially for longer durations. Blacks, the less educated, and individuals in families with less lifetime earnings are more likely to live with dementia for short and for long durations. We found negligible differences between Hispanics and non-Hispanics.

The Alzheimer's Associations (2014) reported results based on a methodology similar to ours. They found that one-third of all seniors who died in their sample (based on the Medicare Current Beneficiary Survey, MCBS) had been previously diagnosed with Alzheimer's or another dementia. Their method is likely to underestimate the lifetime risk of dementia: the assessment is based on linked Medicare data which would miss undiagnosed dementia.

Several studies estimated the lifetime risk of dementia, defined as the fraction of persons cognitively normal at particular ages (e.g. age 70) who would develop the condition before they die. Our estimates are related, but not the same. The first difference is that we report the probabilities of dementia at different durations, up to five years. Second, the lifetime risk of dementia in these studies is typically estimated for single birth cohorts, while our sample included individuals from different cohorts. The average birth year in our sample was 1923, so our results may be closest to the lifetime risk of the 1923 cohort. As younger

generations live longer, we may expect higher lifetime risks of dementia in younger cohorts. There is some evidence in Europe and the U.S., however, that age-specific rates of dementia may be decreasing over time, [18-20, 25-30] which may moderate or even reverse the effect of increasing life expectancy on the lifetime risk of dementia. The third deviation of our estimates from this definition of the lifetime risk is that we did not condition the sample to be cognitively normal at the age of 70.

Based on the large sample size of HRS and the richness of the HRS data, we related dementia risk to a number of SES indicators. We found that dementia is a health disparity where the more educated and those with higher lifetime earnings have a smaller lifetime risk of dementia despite their greater longevity.

The lifetime risk of dementia at age 65 was found to be 42·6% in Canada in a recent study.[14] Our estimates are close to this number, but they exceed earlier estimates from the U.S.[10-13] and Europe.[15] Seshadri et al. (2007), using the community-based Framingham Study, found that 14·3% of 65-year old men and 21·7% of 65-year old women should expect to live with moderate to severe dementia (a subset of all dementia) for at least six months. They did not report the lifetime risk of other dementias. Fishman (2017) found a lifetime risk of 24% for 70-year old males, and 32% for 70-year old females. These estimates are based on observed dementia status some months prior to death and do not permit any progression to dementia in the last months of life. However, the probability of dementia increases substantially in those last months, as shown in Figure 2 and by our estimates of dementia status at death, so not including those late cases will lead to an under-estimate of lifetime risk. Indeed, in the Fishman study the average time between the last observation of dementia status and death was about 20 months; our estimates of dementia probability 24 months prior to death are more comparable, 25·9% for males, and 36·4% for females. There are other estimates of the lifetime risk of AD dementia[10,31] that are not comparable to our estimate of the lifetime risk of any dementia.

A major strength of this study stems from its use of the HRS: the HRS is nationally representative of the population over age 50; it has consistently measured cognitive performance longitudinally for more than

20 years; it clinically assessed a large subsample for dementia status. Thus this study was able to use an established model to estimate the probability of dementia of individuals over time.

A main limitation of this study is that, despite the very long longitudinal reach of the HRS, it has not followed any cohort from age 70 to the death of the last survivor. Thus, we combined estimated age-specific dementia probabilities of several cohorts to estimate risk from age 70 onward, much as a period life table uses moment-in-time mortality rates rather than cohort-specific mortality rates.

Another limitation is that the progression of dementia from the last HRS interview to death is not directly observed. However, the date of death is random with respect to the last HRS interview, and we observe the cognitive status of different people at different durations such as one month prior to death, two months prior and so forth. Our statistical model permits the progression of the probability of dementia to vary smoothly and flexibly with age and duration until death. Furthermore, we explored other models of the probability of dementia. Table 11 in the supplementary appendix shows very similar predictions to the preferred numbers discussed here.

The lifetime risk of dementia is higher than previously estimated; yet, only about 10%-15% of older Americans have long term care insurance,[32,33] leaving the majority of the population dependent on Medicaid or exposed to a large uninsured risk. Further, because of the strong increase in dementia prevalence with age, projected increases in life expectancy are likely to increase the lifetime risk of dementia.

Contributors

All authors contributed to the design of the study and data analyses while Péter Hudomiet took primary responsibility. The manuscript was drafted by the three authors. Jessica Hayes derived the Lifetime Earnings Index with the supervision of the three authors.

Declaration of interests

None of the authors has any conflict of interest for this study.

Acknowledgment

Research support from the National Institute on Aging under grant R01AG053972 (PI: Hurd) is gratefully acknowledged.

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.
2. Prince M, Bryce R, Albanese E, Wimo A, Ribeiro W, Ferri CP. The global prevalence of dementia: A systematic review and metaanalysis. *Alzheimers Dement*. 2013;9(1):63-75. e62.
3. Wimo A, Jonsson L, Bond J, Prince M, Winblad B. (2013), The worldwide economic impact of dementia 2010. *Alzheimers Demen.* 2013;9(1):1-11.
4. Brookmeyer R, Corrada MM, Curriero FC, Kawas C. Survival following a diagnosis of Alzheimer disease. *Arch Neurol*. 2002;59(11):1764-1767.
5. Ganguli M, Dodge HH, Shen C, Pandav RS, DeKosky ST. Alzheimer disease and mortality: a 15-year epidemiological study. *Arch Neurol*. 2005;62(5):779-84.
6. Helzner EP, Scarmeas N, Cosentino S, Tang MX, Schupf N, Stern Y. Survival in Alzheimer disease: a multiethnic, population-based study of incident cases. *Neurology*. 2008;71(19):1489-1495.
7. Larson EB, Shadlen MF, Wang L, McCormick WC, Bowen JD, Teri L, et al. Survival after initial diagnosis of Alzheimer disease. *Ann Intern Med*. 2004;140(7):501-9.
8. Rait G, Walters K, Bottomley B, Petersen I, Iliffe S, Nazareth I. Survival of people with clinical diagnosis of dementia in primary care: cohort study. *Bmj*. 2010;341:c3584.
9. Tom SE, Hubbard RA, Crane PK, Haneuse SF, Bowen J, McCormick WC, et al. Characterization of dementia and Alzheimer's disease in an older population: updated incidence and life expectancy with and without dementia. *Am J of Public Health*. 2015;105(2):408–413.

10. Seshadri S, Wolf PA, Beiser A, Au R, McNulty K, White R, et al. Lifetime risk of dementia and Alzheimer's disease: The impact of mortality on risk estimates in the Framingham Study. *Neurology*. 1997;49(6):1498-1504.
11. Seshadri S, Wolf PA. Lifetime risk of stroke and dementia: current concepts, and estimates from the Framingham Study. *The Lancet Neurol*. 2007;6(12):1106-1114.
12. Alzheimer's Association. 2014 Alzheimer's disease facts and figures. *Alzheimers Dement*. 2014;10(2):e47-e92.
13. Fishman E. Risk of developing dementia at older ages in the United States. *Demography*. 2017;54(5):1897-1919.
14. Carone M, Asgharian M, Jewell NP. Estimating the lifetime risk of dementia in the Canadian elderly population using cross-sectional cohort survival data. *J Am Stat Assoc*. 2014;109(505):24-35.
15. Lobo A, Lopez-Anton R, Santabárbara J, de-la-Cámara C, Ventura T, Quintanilla MA, et al. Incidence and lifetime risk of dementia and Alzheimer's disease in a southern European population. *Acta Psychiatr Scand*. 2011;124(5):372–383.
16. Sonnega A, Faul JD, Ofstedal MB, Langa KM, Phillips JW, Weir DR. Cohort profile: The Health and Retirement Study (HRS). *Int J Epidemiol*. 2014;43(2):576-585.
17. Crimmins EM, Kim JK, Langa KM, Weir DR. Assessment of cognition using surveys and neuropsychological assessment: The Health and Retirement Study and the Aging, Demographics, and Memory Study. *J Gerontol B Psychol Sci Soc Sci*. 2011;66(suppl 1):i162-i171.
18. Hudomiet P, Hurd MD, Rohwedder S. Dementia prevalence in the United States in 2000 and 2012: estimates based on a nationally representative study. *J Gerontol B Psychol Sci Soc Sci*. 2018;73(S1):S10-S19.
19. Langa KM, Larson EB, Crimmins EM, Faul JD, Levine DA, Kabeto MU, et al. A comparison of the prevalence of dementia in the United States in 2000 and 2012. *JAMA Intern Med.* 2017;177(1):51-58.

20. Magaziner J, German P, Zimmerman SI, Hebel JR, Burton L, Gruber-Baldini Al, et al. The prevalence of dementia in a statewide sample of new nursing home admissions aged 65 and older: diagnosis by expert panel. *Gerontologist.* 2000;40(6):663-672.
21. Yaffe K, Fox P, Newcomer R, Sands L, Lindquist K, Dane K, Covinsky KE. Patient and caregiver characteristics and nursing home placement in patients with dementia. *JAMA*. 2002;287(16):2090-2097.
22. Heeringa SG, Fisher GG, Hurd MD, Langa KM. Aging, Demographics and Memory Study (ADAMS) sample design, weighting and analysis for ADAMS. University of Michigan; 2007.
23. RAND HRS Data, Version P. Produced by the RAND Center for the study of aging, with funding from the National Institute on Aging and the Social Security Administration; 2016.
24. Wilson RS, Beck TL, Bienias JL, Bennett JA. Terminal cognitive decline: accelerated loss of cognition in the last years of life. *Psychosom Med*. 2007;69(2):131-137.
25. Larson EB, Yaffe K, Langa KM. New insights into the dementia epidemic. *N Engl J Med.* 2013;369(24): 2275–2277.
26. Matthews FE, Arthur A, Barnes LE, Bond J, Jagger C, Robinson L, et al. Medical research council cognitive function and ageing collaboration. A two-decade comparison of prevalence of dementia in individuals aged 65 years and older from three geographical areas of England: results of the Cognitive Function and Ageing Study I and II. *Lancet*. 2013;382(9902):1405–12.
27. Rocca WA, Petersen RC, Knopman DS, Hebert LE, Evans DA, Hall KS, et al. Trends in the incidence and prevalence of Alzheimer's disease, dementia, and cognitive impairment in the United States. *Alzheimers Dement*. 2011;7(1):80-93.
28. Satizabal CL, Beiser AS, Chouraki V, Chêne G, Dufouil C, Seshadri S. Incidence of dementia over three decades in the Framingham Heart Study. *N Engl J Med*. 2016;375(1):93-94.
29. Schrijvers EM, Verhaaren BF, Koudstaal PJ, Hofman A, Ikram MA, Breteler MM. Is dementia incidence declining? Trends in dementia incidence since 1990 in the Rotterdam Study. *Neurology.* 2012;78(19):1456–63.

30. Wu YT, Fratiglioni L, Matthews FE, Lobo A, Breteler MM, Skoog I, Bryne C. Dementia in western Europe: epidemiological evidence and implications for policy making. *Lancet Neurol.* 2015; 4422:92-97.
31. Brookmeyer R, Abdalla N. Estimation of lifetime risks of Alzheimer's disease dementia using biomarkers for preclinical disease. *Alzheimers Dement*. In press.
32. Brown JR, Goda GS, McGarry K. Long-term care insurance demand limited by beliefs about needs, concerns about insurers, and care available from family. *Health Affairs*. 2012;31(6):1294-1302.
33. Coe NB, Skira MM, Van Houtven CH. Long-term care insurance: Does experience matter? *J Health Econ*. 2015;40:122-131.

Figures and tables

Figure 1. The logic of the longitudinal model used for estimating dementia probabilities at various times prior to death

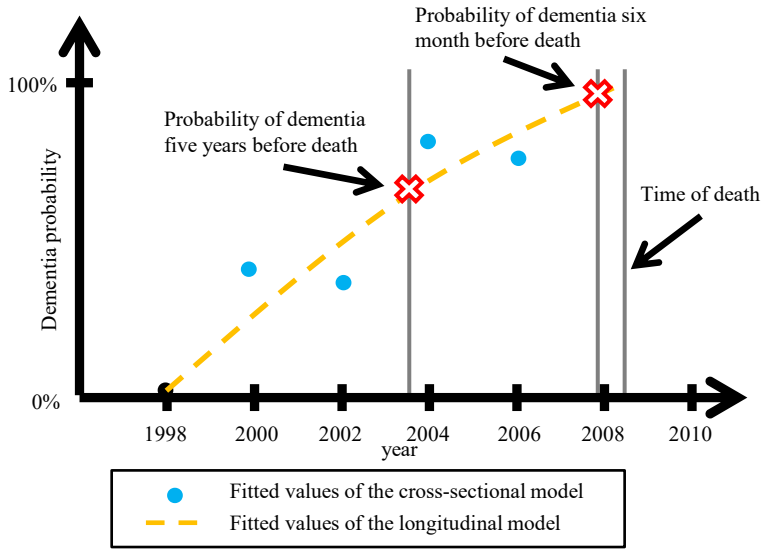
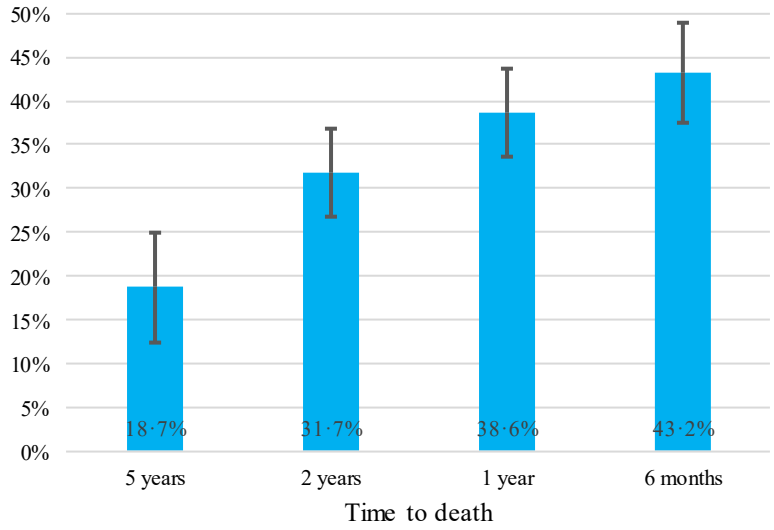
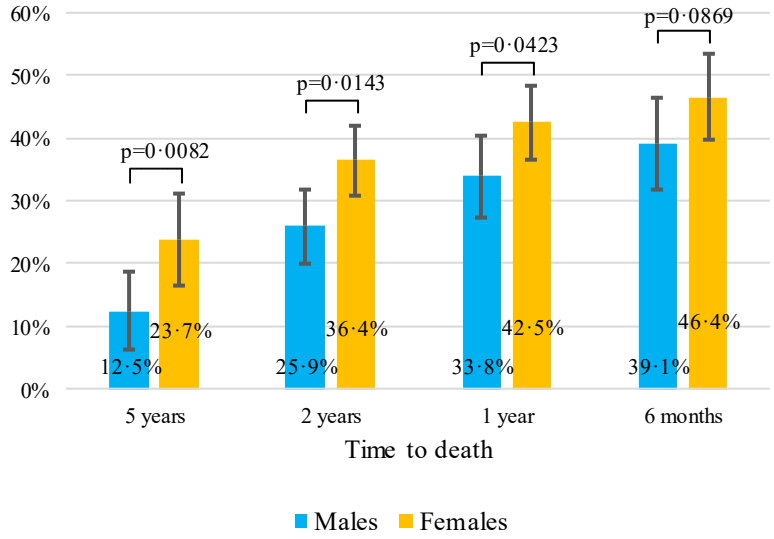


Figure 2. The lifetime risk of living with dementia for six months, one, two, or five years



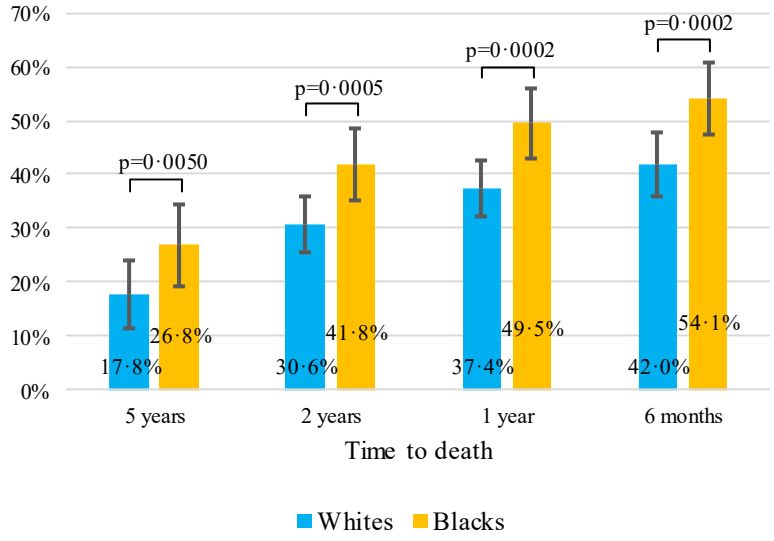
Sample: HRS, 1998-2014, age 70+, dies before 2014. N=8,733. Statistics are weighted. The error bars show the 95% confidence intervals, and they reflect sampling variation and the uncertainty of the dementia calibration models.

Figure 3. The lifetime risk of living with dementia for six months, one, two, or five years by gender



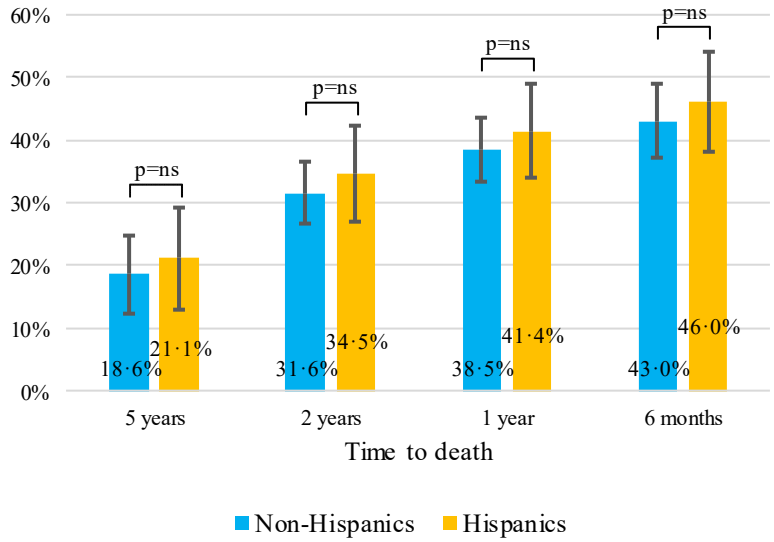
Sample: HRS, 1998-2014, age 70+, dies before 2014. N=8,733. Statistics are weighted. The error bars show the 95% confidence intervals. The confidence intervals and the p-values reflect sampling variation and the uncertainty of the dementia calibration models. Group differences in the risk are more precisely estimated than the levels.

Figure 4. The lifetime risk of living with dementia for six months, one, two, or five years by race



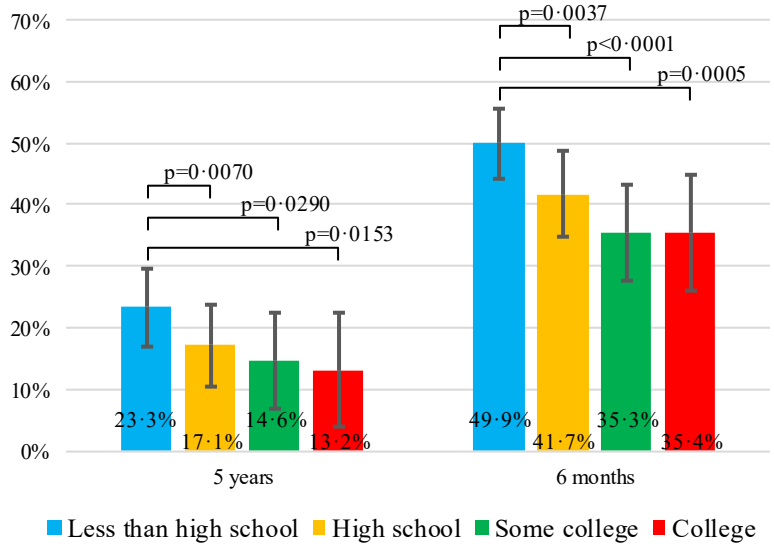
Sample: HRS, 1998-2014, age 70+, dies before 2014, whites and blacks. N=8,511. Statistics are weighted. The error bars show the 95% confidence intervals. The confidence intervals and the p-values reflect sampling variation and the uncertainty of the dementia calibration models. Group differences in the risk are more precisely estimated than the levels.

Figure 5. The lifetime risk of living with dementia for six months, one, two, or five years by ethnicity



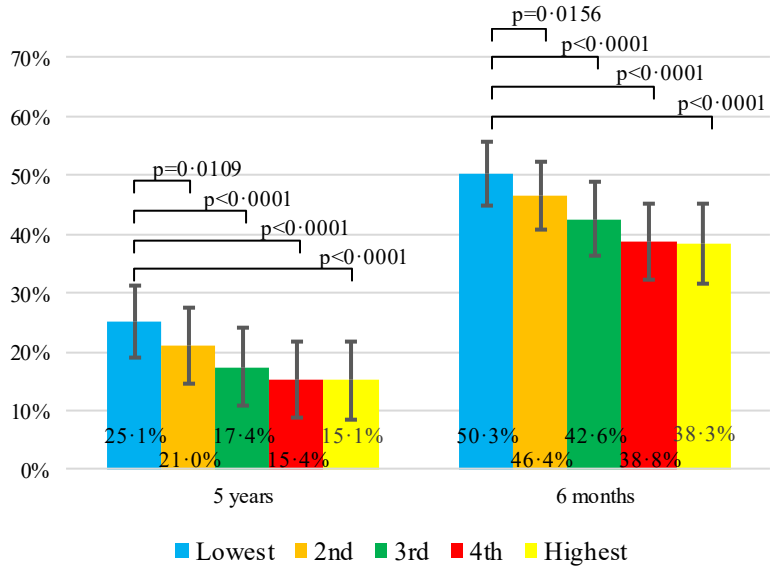
Sample: HRS, 1998-2014, age 70+, dies before 2014. N=8,733. Statistics are weighted. The error bars show the 95% confidence intervals. The confidence intervals and the p-values reflect sampling variation and the uncertainty of the dementia calibration models. Group differences in the risk are more precisely estimated than the levels.

Figure 6. The lifetime risk of living with dementia for six months or five years by education



Sample: HRS, 1998-2014, age 70+, dies before 2014. N=8,733. Statistics are weighted. The error bars show the 95% confidence intervals. The confidence intervals and the p-values reflect sampling variation and the uncertainty of the dementia calibration models. Group differences in the risk are more precisely estimated than the levels.

Figure 7. The lifetime risk of living with dementia for six months or five years by lifetime earnings



Sample: HRS, 1998-2014, age 70+, dies before 2014. N=8,733. Statistics are weighted. The error bars show the 95% confidence intervals. The confidence intervals and the p-values reflect sampling variation and the uncertainty of the dementia calibration models. Group differences in the risk are more precisely estimated than the levels. Lifetime earnings are estimated using a procedure detailed in the Appendix.

Table 1. Descriptive statistics

	Unweighted frequency	Unweighted fraction	Weighted fraction
Males	3987	0·457	0·447
Females	4746	0·543	0·553
Less than HS	3744	0·429	0·404
High school	2534	0·290	0·303
Some college	1415	0·162	0·166
College or more	1040	0·119	0·127
White	7265	0·832	0·879
Black	1246	0·143	0·096
Other race	222	0·025	0·025
Non-Hispanic	8167	0·935	0·949
Hispanic	566	0·065	0·051
Lifetime earnings quintiles			
Lowest quintile	1747	0·200	0·188
2nd quintile	1747	0·200	0·199
3rd quintile	1747	0·200	0·204
4th quintile	1746	0·200	0·205
Highest quintile	1746	0·200	0·203
Single at death	5161	0·591	0·602
Married at death	3572	0·409	0·398
Died at age 70-75	1275	0·146	0·144
Died at age 75-80	1613	0·185	0·188
Died at age 80-85	1884	0·216	0·230
Died at age 85-90	1915	0·219	0·221
Died at age 90-95	1382	0·158	0·150
Died at age 95+	664	0·076	0·067
Had hypertension before death	5632	0·645	0·645
Had diabetes before death	2256	0·258	0·252
Had cancer before death	2182	0·250	0·256
Had a lung disease before death	1660	0·190	0·197
Had a heart problem before death	4087	0·468	0·472
Had stroke before death	1965	0·225	0·224
Had psychiatric problems before death	1760	0·202	0·203
Had arthritis before death	5890	0·674	0·676
last HRS wave was 1998	1182	0·135	0·130
2000	1217	0·139	0·135
2002	1075	0·123	0·124
2004	1080	0·124	0·122
2006	1066	0·122	0·126
2008	1262	0·145	0·148
2010	914	0·105	0·106
2012	937	0·107	0·109

Sample: HRS, 1998-2014, age 70+, dies before 2014. N=8,733. Lifetime earnings are estimated using a procedure detailed in the Appendix.

Table 2. OLS regressions of the lifetime risk of living with dementia for six months or five years

	5 years		6 months	
	Coef.	95% CI	Coef.	95% CI
Females	0·076**	[0·015; 0·137]	0·01	[-0·078; 0·098]
Less than HS	ref.	ref.	ref.	ref.
High school	-0·038*	[-0·077; 0·001]	-0·045	[-0·098; 0·008]
Some college	-0·062**	[-0·117; -0·007]	-0·109***	[-0·172; -0·046]
College	-0·056	[-0·125; 0·013]	-0·096**	[-0·176; -0·016]
White	ref.	ref.	ref.	ref.
Black	0·073**	[0·014; 0·132]	0·109***	[0·046; 0·172]
Other race	0·004	[-0·067; 0·075]	0·012	[-0·066; 0·09]
Hispanic	0·01	[-0·047; 0·067]	0·006	[-0·059; 0·071]
Lifetime earnings quintiles				
Lowest quintile	ref.	ref.	ref.	ref.
2nd quintile	-0·032***	[-0·056; -0·008]	-0·032**	[-0·059; -0·005]
3rd quintile	-0·050***	[-0·074; -0·026]	-0·052***	[-0·079; -0·025]
4th quintile	-0·038***	[-0·063; -0·013]	-0·046***	[-0·075; -0·017]
Highest quintile	-0·040***	[-0·064; -0·016]	-0·061***	[-0·092; -0·03]
Married	0·005	[-0·028; 0·038]	0·031	[-0·008; 0·07]
Died at age 70-74	ref.	ref.	ref.	ref.
Died at age 75-79	0·047**	[0·01; 0·084]	0·110***	[0·065; 0·155]
Died at age 80-84	0·091***	[0·04; 0·142]	0·214***	[0·153; 0·275]
Died at age 85-89	0·131***	[0·074; 0·188]	0·323***	[0·254; 0·392]
Died at age 90-94	0·165***	[0·106; 0·224]	0·414***	[0·332; 0·496]
Died at age 95+	0·215***	[0·141; 0·289]	0·480***	[0·345; 0·615]
Had hypertension before death	-0·048***	[-0·07; -0·026]	-0·053***	[-0·09; -0·016]
Had diabetes before death	-0·009	[-0·034; 0·016]	0·01	[-0·031; 0·051]
Had cancer before death	-0·053***	[-0·069; -0·037]	-0·072***	[-0·094; -0·05]
Had a lung disease before death	-0·040***	[-0·06; -0·02]	-0·058***	[-0·082; -0·034]
Had a heart problem before death	-0·021***	[-0·035; -0·007]	-0·030***	[-0·048; -0·012]
Had stroke before death	0·200***	[0·145; 0·255]	0·292***	[0·231; 0·353]
Had psychiatric problems before death	0·144***	[0·109; 0·179]	0·187***	[0·14; 0·234]
Had arthritis before death	-0·047***	[-0·065; -0·029]	-0·070***	[-0·092; -0·048]
last HRS wave was 1998	ref.	ref.	ref.	ref.
2000	-0·017	[-0·046; 0·012]	0·005	[-0·03; 0·04]
2002	-0·038*	[-0·077; 0·001]	-0·006	[-0·041; 0·029]
2004	-0·048**	[-0·089; -0·007]	0·009	[-0·026; 0·044]
2006	-0·036	[-0·081; 0·009]	0·012	[-0·023; 0·047]
2008	-0·064***	[-0·105; -0·023]	-0·016	[-0·053; 0·021]
2010	-0·070***	[-0·111; -0·029]	-0·004	[-0·041; 0·033]
2012	-0·061***	[-0·102; -0·02]	-0·006	[-0·043; 0·031]
Constant	0·162***	[0·086; 0·238]	0·270***	[0·18; 0·36]

Sample: HRS, 1998-2014, age 70+, dies before 2014. N=8,733. Statistics are weighted. ***, **, and * denote statistical significance at 1%, 5%, and 10% respectively. The significance stars and the confidence intervals reflect sampling variation and the uncertainty of the dementia calibration models. Lifetime earnings are estimated using a procedure detailed in the Appendix.

Appendix

A1. Missing data in the HRS

As any survey data, the HRS has missing values, though non-response rates are low.[1,2] The missing cases were carefully investigated and imputed using logical and simple single imputation methods by RAND[2] or by us. Implementing a more sophisticated multiple imputation framework was not feasible in this study, due to the complexity of our two-step procedure and the bootstrapped standard errors.[1] Because the purpose of this study was to predict the lifetime risk of dementia, imputation biases in predictor variables would only bias the results if they were systematically different in the ADAMS sample and the rest of the HRS. We are not aware of any reasons why this would be the case.

Now we discuss some of the decisions that were made about missing values. When the year of death was not available in the NDI or the survey (83 cases), we imputed the year after the last HRS interview. When the month of death was missing (94 cases) we either impute the month after the last interview (when the year of death was known to be the same as the last interview) or June (when the year of death was missing). The handful of missing values in demographic the variables were replaced by their modes, such as being white or being married. The next section discusses how we dealt with missing lifetime earnings.

A2. Estimating lifetime earnings

The Social Security's Primary Insurance Amount (PIA) is our preferred measures of socioeconomic status. The PIA is the social security benefit a person would receive if he started collecting benefits at the normal retirement age.[2] The PIA is a monotonic function of individuals' total lifetime earnings, and therefore it is an informative summary measure of individuals' earnings histories. The PIA can be estimated for each individual within a family, but we used a household level measure by taking the larger PIA values of husbands and wives.

The PIA can be constructed from data on Social Security Retirement (SSR) income and the claiming age, which is the age at which the individual first received SSR income. Workers who claim SSR income at a later age receive "delayed retirement credit" and ultimately higher retirement benefits, and therefore their SSR income needs to be adjusted to reflect their PIA. For example, if a person retires after the normal retirement age, his SSR income is higher than his PIA, and it needs to be reduced accordingly. There are some complications, however. First, the adjustment factors may be different for widows and disabled individuals. Second, individuals' reported SSR income, disability and marital statuses may change over time in the HRS. Third, the HRS data contain some missing values, and outliers that are inconsistent with SSA rules. This appendix explains how we dealt with these problems.

A2.1. The standard case

We first discuss how we estimated the PIA for individuals who are not disabled, not widows, and whose SSR income and claiming ages are reported in the HRS. The social security website contains the adjustment factors we used to adjust their SSR income reports.[3] The HRS asks about the claiming age once, but the SSR income is asked in each survey wave. We adjusted each valid HRS income reports by the SSA adjustment factors and by the consumer price index (CPI)[4] using the formula

$$PIA_{wave} = \frac{Reported\ SS\ Income}{Adjustment} \times \frac{CPI_{2014}}{CPI_{wave\ year}}.$$

[1] The current version of the model ran for about 5 days on a linux server at RAND.
[2] See the explanation at https://www.ssa.gov/agency/glossary/.
[3] We used Table 2.A20 at https://www.ssa.gov/policy/docs/statcomps/supplement/2017/2a20-2a28.html.
[4] We used the June-to-June CPI-W values from https://www.ssa.gov/oact/STATS/cpiw.html.

If this calculation was higher than the maximum individual Social Security value, we replaced it with the maximum ($31,704).[5]

The next step was to choose the wave that best captured the respondent's socioeconomic status. Our algorithm aimed at choosing the PIA from the wave in which the respondent's age was as close to 70 as possible, but at the same time avoiding imputed and zero SSR income values.[6] We used the following algorithm:

1. Use the latest non-imputed and non-zero observation between ages 65 and 70.
2. If not available, then use values between ages 60 and 75 in the following order: 71 64 72 63 73 62 74 61 75 60.
3. If still not available, then use ages 76 and above (using youngest observed age first).
4. If still not available, then repeat steps 1-3 allowing for imputed SSR values.
5. If still not available, then use the median of the non-zero PIA values below age 60.

Finally, the respondent level PIAs of married couples were replaced by the larger PIA values of husbands and wives.

A2.2. Disabled respondents

Normally, those who start collecting SSR income earlier than the normal retirement age receive lower than normal benefits. However, there is no such adjustment for disabled individuals. Therefore, the standard-case adjustment factor was replaced by 1 (i.e. unadjusted) for disabled respondents who claimed early.[7]

Disabled respondents are also allowed to claim earlier than the minimum claim age, which is 62 in standard cases and 60 for widows. Before age 62, benefits are categorized as Social Security Disability Insurance (SSDI). At age 62, these benefits are transferred to regular Social Security Retirement benefits, but this switch was not always properly reported by the respondent. Therefore, we sometimes used SSDI values instead of SSR values for disabled respondents.[8] The algorithm to select the wave the PIA is based on followed the standard case. However, SSDI income was considered imputed, and so it was not used in the first three steps of the algorithm.

A2.3. Widowed respondents

The adjustment factor for widows may differ from the standard case as well. Because the HRS does not record if a particular SSR income is a widow benefit or a regular one, we approximated it. We assumed individuals received widow benefits and used the adjustment factors accordingly[9] in the following instances:

1. Female respondents who reported to be widows prior to first claiming SSR benefits.
2. Non-disabled male widows who first claimed SSR income between ages 60 and 62.[10]

---

[5] In 2014, the maximum monthly benefit for an individual retiring at the normal retirement age was $2,642 (https://www.ssa.gov/oact/cola/examplemax.html). The yearly value of $31,704 equals 12 times this monthly maximum. This cap was applied after CPI-adjusting all values in the panel to 2014 dollars. Overall, less than 1% of the sample was influenced by the cap in any given wave.

[6] Missing SSR income is regularly imputed by the RAND-HRS team as documented in Sections 2 and 3 of the RAND HRS Detailed Imputations 2014 (V2) Codebook at https://www.rand.org/content/dam/rand/www/external/labor/aging/dataprod/randhrsimp1992_2014v2.pdf.

[7] If a disabled respondent claimed social security after normal retirement age, the adjustment factor remained the standard case value. We defined disabled respondents as those who were flagged as receiving SSDI in the sample or who listed "disabled" as their labor force status in any wave.

[8] Zero SSR income was replaced by SSDI unless the person also reported receiving Supplemental Security Income from SSA. In the latter case the accuracy of the SSDI reports may be lower, because individuals receive Social Security income from multiple sources.[2]

[9] See Table 2.A22 in the Annual Statistical Supplement, 2017: https://www.ssa.gov/policy/docs/statcomps/supplement/2017/2a20-2a28.htm

[10] The only valid reasons for claiming before 62 are widowhood or disability. When a widowed male was disabled, we assumed he claimed his own value.

For widowed men who were age 62 or older when they first claimed benefits (or disabled male widows of any age), we assumed they receive their own benefits. We made this distinction due to the observed differences in earnings and work history for men and women in the age cohorts of our sample.

When a respondent had Social Security values observed before and after widowhood, we attempted to choose the non-widowed observation for the respondent-level PIA when possible, to minimize the risk of using the wrong adjustment factors. In practice it meant that in the first three steps of the standard case algorithm we used non-imputed and non-widowed values.

We also attempted to avoid widowed values when choosing the maximum household PIA. If we only found a widowed PIA value for one respondent, but the other spouse was observed to have claimed Social Security before death, then we used the value of the non-widowed respondent.[11]

A2.4. Missing claiming age

The adjustment factor is based off birth year and claim age. Birth years are observed for the entire sample. The age (in months) they first claimed Social Security benefits, however, are sometimes missing and/or inconsistent with SSA rules. We used the following algorithm to deal with these problems:

1. The claim age of non-widow, and non-disabled respondents, who reported a value less than 62 was replaced by 62, the youngest valid claiming age. For widows, we use the corresponding minimum age of 60. Disabled individuals were allowed to have any claiming ages.
2. If respondents reported first receiving benefits more than two years before the HRS interviews without providing an exact age (which was an answer option in some HRS waves), we replaced the claiming age with 62 (60 for widows) if they were first observed in the sample at age 64 or younger.
3. If the claim age was missing, but we observed a transition in the HRS from not-receiving SSR income in one wave to receiving it in the next, the claiming age was set to the age at the mid-point date between the two HRS interviews. If that value was below 62 (60 for widows), it was replaced with 62 (60).
4. Finally, if the claiming age was still missing, we set it to 63 (62 years and 2 months for widows), which are the median claiming ages for respondents who claim between ages 62 and 70 (60 and 70 for widows).

A2.5. Missing SSR income values

The PIA of respondents who did not have any reported SSR income in the HRS were imputed. If the respondent reported to have started receiving Social Security Retirement income, but never reported a positive amount, the PIA was set to 0.[12] A few respondents in our 65 years and older sample had not yet started claiming before their latest HRS wave. The PIAs of these individuals were imputed using a regression based model.[13] The imputation donors were individuals who first claimed social security after the age of 65. We used the following predictor variables:

- Gender
- Race (non-Hispanic white, non-Hispanic black, Hispanic, other)
- Education level (less than high school, high school, some college, college or more)
- Years worked
- Indicator if years worked was zero or missing.
- Birthyear cohorts (1900-1919, 1920-1929, 1930-1939, 1940-1949)
- Average log earnings (using all HRS waves with non-zero values)
- Average log income (using all HRS waves with non-zero values)
- Average log wealth (using all HRS waves with positive values)

[11] There were two exceptions, when we did not give a preference for the non-widowed values: (1) when the to-be-deceased spouse was disabled but the (future) widow was not; and (2) when the to-be-deceased spouse was female. In these cases, we took the maximum of the respondent and spouse values.
[12] There are a few valid reasons a respondent may receive zero benefits. For instance, those who deduct their Medicare Part B premiums from their SS benefits, may lose all SS benefits if the initial values were low enough.
[13] Of the 21,462 individuals in the sample, 713 respondents (3·3%) are imputed. Of these cases, however, only 503 (2·3%) are picked as the household level PIA.

- Indicators if the earning or wealth averages were non-positive.

The output of the regression model is in Table A1. The imputed values were then the sum of the predicted values of this regression and mean-zero normally distributed residuals with the estimated variance.

A2.6. Multiple observed spouses

A small number of respondents have multiple observed spouses in the sample. For these individuals, we choose the maximum household PIA from the wave closest to when the respondent was age 70.

A2.7. Summary

Table A2 summarizes the number of observations in our sample by type. About 72% of the sample got the PIA from a standard (not widow, not disabled, not imputed) individual (15,483 out of 21,462 cases); a quarter were from widows or disabled individuals, and only about 2% were imputed. However, about a third of the claiming ages needed to be modified (7,855 cases), typically because the claiming ages were not reported in the HRS.

Table A3 shows summary statistics of our final PIA measure in the total sample and in interesting subgroups. The mean of PIA is about $16,000, and the distribution is fairly symmetric. The 5th percentile is about $6,000, and the 95th percentile is about $32,000. There is a relatively minor gender difference, perhaps because our preferred PIA is defined at the household level. At the sample time, there are large differences by education (the higher educated have higher PIA), race (non-Hispanic whites have higher PIA than minorities), and disability as well as marital statuses (non-disabled and married individuals have higher PIA).

A3. Estimating dementia probabilities in the HRS (Step 1 of our procedure)

The HRS has a large number of cognition measures[3] such as TICS scores (in self-interviews), Jorm IQCODE (in proxy interviews) and self-rated memory (both in self and in proxy interviews). These variables and the ADAMS diagnosis of dementia were used to estimate probabilities that individuals were demented in any HRS waves since 1998.

We used a modified version of the methodology developed by Hurd et al. (2013)[4]. They ran ordered probit models of the ADAMS diagnosis (demented, CIND, normal) as functions of HRS variables prior to ADAMS. They ran different models for proxy and non-proxy interviews due to the difference in the available variables. Then they used the predicted values of these regressions to assign dementia probabilities to all individuals at all HRS waves, even when ADAMS assessment was not available. The weighted average of these dementia probabilities is an unbiased estimate of dementia prevalence in the population. The average of the probabilities in subgroups (such as by age, gender, education) is an unbiased estimate of dementia prevalence in the subgroups *as long as* the subgroup identifiers are included in the ordered probit models. If the subgroup identifiers are not included, the group differences are biased toward zero.[5] Therefore, Hurd et al. also included a set of demographic variables in the ordered probit models of dementia.

The goal in this paper was to estimate dementia prevalence prior to death. The model of Hurd et al. did not include any measures of mortality status in the HRS-ADAMS model of dementia. In our current sample, however, it is necessary to include mortality information, because otherwise the difference in the dementia probabilities between the surviving and the dying samples were biased toward zero. This would likely bias the dementia probabilities down in our sample because dementia and mortality are positively related. Therefore, we also include measures of health and mortality in the ordered probit models. We use self-reported health status, prior diagnosis of hypertension, diabetes, stroke, and psychiatric problems (that are known dementia risk factors[6]), and functions of time-to-death, measures to the precision of 0.1 years. Out of the 856 ADAMS participants 665 (78%) have died by 2014, and their time-to-death is observed. To increase precision and sample size we also include the rest of the 191 individuals in the models: their time-to-death is imputed to 12 years (the last HRS interview plus one year), but we

also include an indicator for survival status to pick up average imputation bias.[14] Age and time-to-death were measures from the ADAMS interviews, but all other predictor variables were measured in the prior HRS interviews.

The outputs of the estimation models are in Tables A4 and A5. Both tables show two specifications: one with and one without adding time-to-death information to the model. Time-to-death improves model fit, but it is only statistically significant in proxy interviews. The point estimates indicate that those who are further away from death are less likely to be demented even when all the other HRS measures (including cognition) are controlled. Thus, mortality adjustment (by adding the timing of death) is important to reliably measure dementia prevalence prior to death. The coefficients on the other predictors also make sense: the various memory measures and ever having a stroke are strong predictors of dementia in self interviews; and self-reported memory, interviewer reported cognitive limitations and Jorm IQCODE are strong predictors of dementia in proxy interviews.

Then we used the predicted values from these regressions to predict dementia probabilities for all HRS respondents who died by 2014 after the age of 65. (We used the 65 rather than 70 age cutoff to add an extra 5 year window to our final sample.) For this sample, time-to-death is observed, so any imputation bias from the estimation plays no role. The dementia probabilities are estimated for all HRS waves between 1998 and 2014. Age and time-to-death are calibrated to one year after the HRS interviews, because the ADAMS diagnosis was also about one year after the HRS interviews. The predicted values, thus, show the probabilities of dementia one year after HRS interviews among those who survive to that date.

A4: Fitting a dynamic model on dementia probabilities (Step 2 of our procedure)

The next step was to estimate dementia probabilities at death and several discrete time points prior to death from the dementia probabilities at the HRS interviews (acquired in the previous step). We used a dynamic latent variable model of cognitive ability. We assumed that cognitive ability followed a deterministic function of demographics, age and time-to-death, and a stochastic Brownian motion to allow individuals to flexibly deviate from the deterministic trend. Formally, we assumed that latent cognitive ability, c_{it}^{*} , evolved as follows:

$$c_{it}^{*} = \beta' x_{it} + f\left(a_{it}, d_{it}\right) + \alpha_i + \eta_{it}, \tag{0.1}$$

$$\eta_{it} = \eta_{i,t-1} + \sqrt{a_{it} - a_{i,t-1}}\,\delta_{it}, \tag{0.2}$$

where x_{it} are basic covariates (gender, education, race) $f\left(a_{it}, d_{it}\right)$ is the deterministic function of age (a_{it}) and time-to-death (d_{it}). Subscript i refers to individuals and t refers to HRS waves. In our preferred model we used a cubic of age, a fifth order polynomial of time-to-death and interactions between age, time-to-death and time-to-death squared. α_i represents heterogeneity in cognitive ability at age 65, which is assumed to follow a standard normal distribution:

$$\alpha_i \sim N\left(0,1\right) \tag{0.3}$$

The restrictions on the mean and the variance of α_i are normalizations. Finally, η_{it} represents the stochastic evolution of cognitive ability in a Brownian motion fashion. The variance of the innovation term ($\sqrt{a_{it} - a_{i,t-1}}\delta_{it}$) between two HRS interviews is then $\left(a_{it} - a_{i,t-1}\right)\sigma_{\delta}^{2}$. An individual is demented if his cognitive ability falls below zero, $c_{it}^{*} < 0$.

The dementia probabilities estimated in the prior step are optimal predictions of dementia status given a set of predictors (TICS, IQCODE, etc.) from the current HRS wave. For simplicity (and without loss of generality) in this

[14] The ADAMS sample only included 199 individuals with proxy interviews in the latest HRS wave. This is a relatively small sample size, and we decided against further shrinking it by excluding the survivors to 2014 wave.

model we assume, instead, that the dementia probabilities in step 1 (p_{it}) were optimal predictions given a single noisy signal variable c_{it} :

$$c_{it} = c_{it}^{*} + \varepsilon_{it} , \tag{0.4}$$

$$p_{it} = \Pr\left(c_{it}^{*} < 0 \mid c_{it}\right). \tag{0.5}$$

Intuitively, one can think of the noisy signal c_{it} as the index of all the HRS variables used in the ordered probit model of dementia. The noise term is assumed to be normally distributed:

$$\varepsilon_{it} \sim N\left(0, \sigma_{\varepsilon}^{2}\right) \tag{0.6}$$

The data points used for estimation of the dynamic model (in this step 2) are the step 1 dementia probabilities, p_{it} -s, and the parameters to be estimated are the coefficients in the deterministic trend, and the variances of the random components.

This is a fully specified model and it can be estimated by Maximum Likelihood. Once estimated, we can predict dementia probabilities at any values of age and time-to-death, as functions of the model parameters and the HRS dementia probabilities.

$$p_{i}^{adj}\left(a,d\right) = \Pr\left(c_{i}^{*}\left(a,d\right) < 0 \mid p_{i1},...,p_{iT}\right) \tag{0.7}$$

The output of the model is in Table A6. Most coefficients are highly significant. We experimented with more flexible specifications, but the predicted values were very similar, and we sometimes had convergence issues in some bootstrap samples, so we decided to go with this simpler specification.

Panel A of Figure A1 shows the predicted values of the ordered probit models, and Panel B shows the predicted values of the dynamic model on the sample of those who died after age 65. Panels C and D compare wave-to-wave (approx. 2 year) changes in dementia probabilities. The sample includes those who are observed to die in the HRS after 1998. The prevalence of dementia in this group is higher than in the general population, around 25 percent. Mean prevalence and mean changes in prevalence are almost the same in the ordered probit and the dynamic models. This was expected, because the dynamic model was not supposed to change mean prevalence. The standard deviations, however, are different. The adjusted probabilities (Panel B) have a higher standard deviation, because the model predicts more 0%-s and 100% dementia probabilities. That is, the dynamic model produces more certain predictions. This is not surprising, because these probabilities are based on more information: they use all waves rather than just one. Wave-to-wave changes in the adjusted probabilities (Panel D) have a smaller standard deviation, because the model predicts fewer spurious within-person changes. In fact, the unadjusted probabilities (Panel C) have far more negative values. We expect the probabilities of dementia to increase with age, and thus, negative changes likely reflect noise. The dynamic model predicted far fewer negative values, even though changes were not forced to be non-negative.

The two panels of Figure A2 shows the histograms of the dementia probabilities five years and half a year before death in our final sample: those who died after age 70. Both histograms have bimodal distributions. The mean dementia probability five years before death is 18·7% and the mean six months before death is 43·2%. Most observations have probabilities very close to zero or one, which implies that our procedure can distinguish demented and non-demented cases most of the time. However, a non-trivial fraction of the sample has less certain dementia probabilities.

A5. Testing the statistical procedure on synthetic data

We tested our statistical methodology on synthetic data with known and realistic parameter values. First, we describe how the dataset was simulated. Second, we compare the estimated lifetime risk of dementia to its (assumed) true value.

The model follows the setup described in Section E4. Cognitive ability, however, only depends on age and time-to-death linearly:

$$c_{it}^{*} = \beta_0 + \beta_1 a_{it} + \beta_2 d_{it} + \alpha_i + \eta_{it}, \tag{0.8}$$

Survival probabilities are modeled in a Gompertz framework with hazard rate

$$h_i\left(a_{it}\right) = \lambda_i^0 \lambda^1 \exp\left(\lambda^1 a_{it}\right) \tag{0.9}$$

$$\ln \lambda_i^0 = \theta_0 + \theta_1 \alpha_i \tag{0.10}$$

The scale parameter (λ^1) is assumed to be homogeneous in the population, but we allow cognitive ability to predict the shape parameter (λ_i^0).

Next, we simulated data for HRS waves from 1998 to 2014. Given that we only model the 65+ year old population, we started with the 1900-1949 birth cohorts. We assumed birth cohorts are symmetric and share the same parameter values of cognition and survival. Then we simulated their cognitive abilities and survival status to the 1998-2014 HRS waves. We kept all alive person-year observations above age 65.

Finally, we modeled the timing of ADAMS interviews. We assumed ADAMS interviews followed the 2000 wave. The interviews were scheduled T_i months after the 2000 HRS interview, where

$$\ln T_i = \tau_0 + \tau_1 \alpha_i + \upsilon_i \tag{0.11}$$

$$\upsilon_i \sim N\left(0, \sigma_\upsilon^2\right) \tag{0.12}$$

The timing of ADAMS interviews, thus, was allowed to correlate with cognitive ability. We chose values so that the interviews are approximately one year after the HRS interviews, on average, just like in the real HRS-ADAMS data. If an individual survived to his ADAMS interview, we observed his dementia status (whether c_{it}^{*} was above or below zero).

The chosen parameter values are in Table A7. We simulated 100,000 individuals, but only 63,304 survived to age 65. Altogether we had 321,411 person year observations.

The first three rows of Table A8 show dementia probabilities in the entire sample, irrespective of survival status to the last 2014 wave. (Our preferred sample, however, is restricted to those who are observed to die before 2014.) We used the noisy dementia signal (c_{it}) and a cubic polynomial of age in a probit model of dementia status in ADAMS to predict dementia probabilities. The estimated prevalence of dementia is 16·27% among those who survived at least one year after their HRS interviews, which is very close to the true 16·23% value. The estimated prevalence is also very close when we do not restrict the sample to those survive one more year (16·93% vs. 16·95%). Note, however, that these latter estimates are somewhat unusual, because they include dementia statuses of people after their deaths, which can only be "observed" in a synthetic dataset.

The second three rows show analogous results in the sample that died before the last HRS interview in 2014 (our sample of interest). To predict dementia probabilities, this model includes a cubic function of time-to death and a dummy indicating survival to 2014, as well as c_{it} and a cubic of age. Time-to-death was imputed to 12 years among those who survived to 2014, just like in the real HRS data. The estimated dementia probabilities are very close to the true values. In the sample that survived at least one more year after their HRS interviews, we estimated a prevalence of 23·26%, which is very close to the true 23·01%. When we do not restrict the sample to those who survived at least one more year, the prevalence estimate is even closer to the truth (24·03% vs. 24·06%).

The second step of our procedure is to fit a dynamic model of cognitive status on the dementia probabilities from Step 1. The results in Table A9 confirm that the procedure works well. The estimated prevalence of dementia at death and six months prior to death (30·66% and 29·71%) are very close to the true values (30·45% and 29·46%).

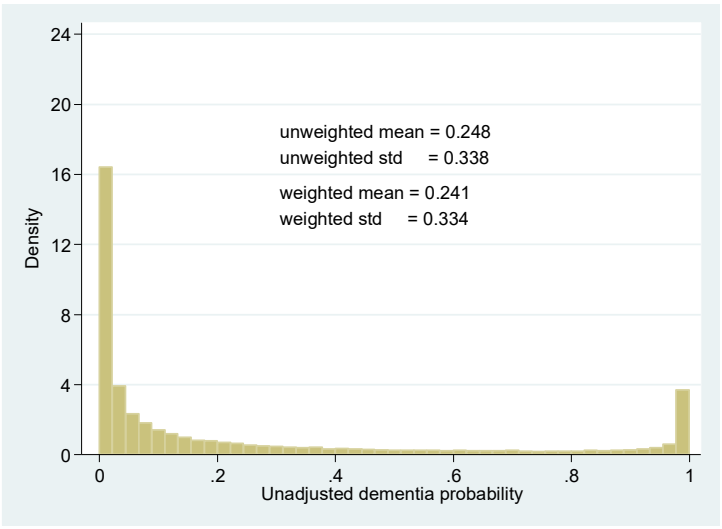
A6. Standard errors

Under the assumptions of the model, this procedure produces consistent estimates of dementia prevalence prior to death. The standard errors, however, need to be adjusted to take into account the fact that dementia probabilities are estimated rather than known. We developed a bootstrap procedure to provide valid standard errors. The procedure builds on Hurd et al. (2013). We first used a parametric bootstrap model[7] to create 200 bootstrap estimates of the dementia probabilities (from step 1). Then, we fitted our dynamic (step 2) model on each of these bootstrap estimates to get 200 dementia probabilities at death (and various times prior to death). Finally, we used a non-parametric bootstrap method to take the average of these probabilities. We draw 10 samples with replacement from each of the 200 parametric bootstrap estimates ($200 \times 10 = 2000$ draws in total)[15]. We ran a detailed Monte Carlo Simulation on synthetic data and we found that our procedure accurately recovered the 95% coverage rate.

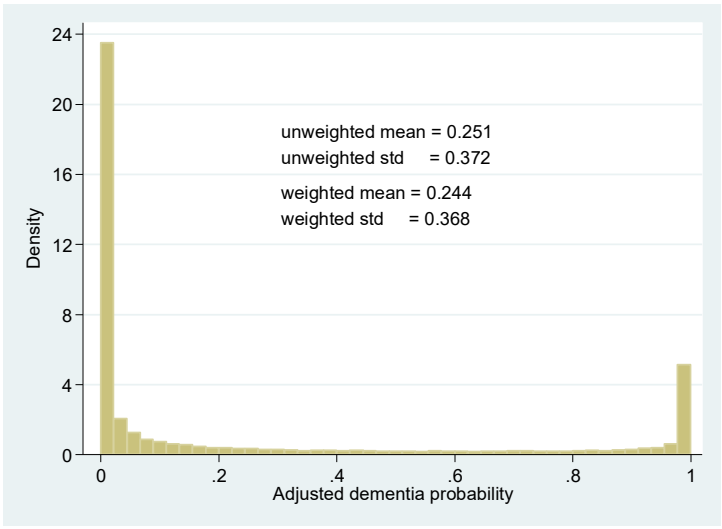
[15] The non-parametric bootstrap in the last step is necessary. The parametric bootstrap resamples residuals, but it does not resample predictors, such as TICS scores. Thus, taking the average of the 200 bootstrap estimates provides too small standard errors and too narrow confidence intervals. The non-parametric bootstrap resamples the entire dataset, and thus, it accurately accounts for the variation in the predictors.

A7: Additional figures and tables

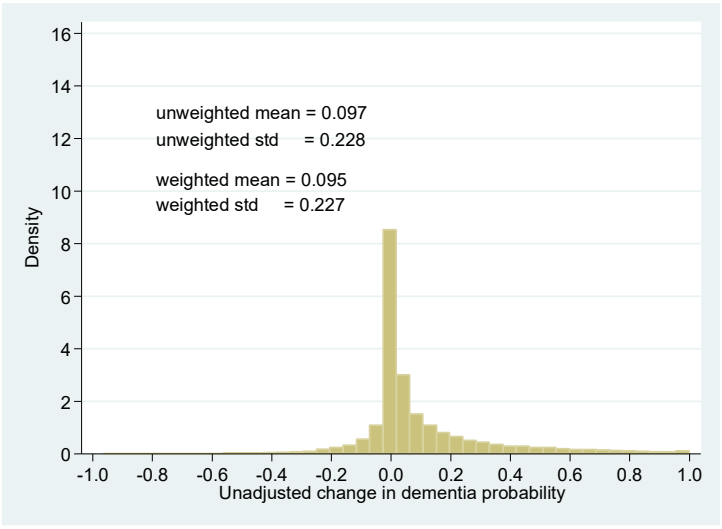
Figure A1. Histograms of dementia probabilities and changes in dementia probabilities



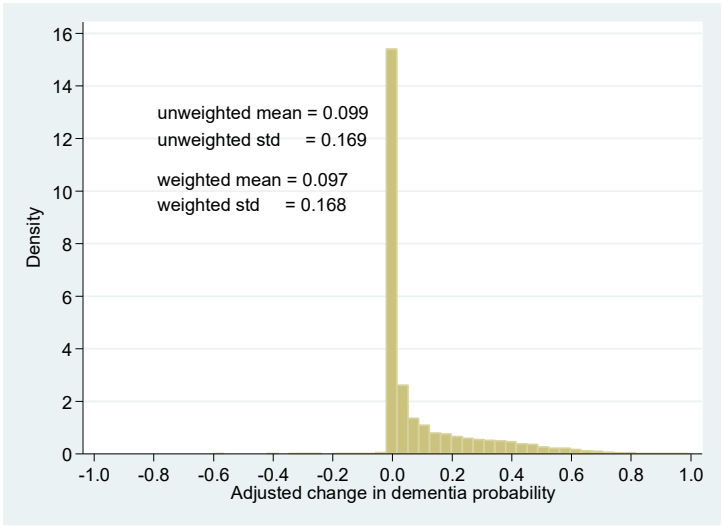
Panel A: Unadjusted dementia probabilities



Panel B: Adjusted dementia probabilities



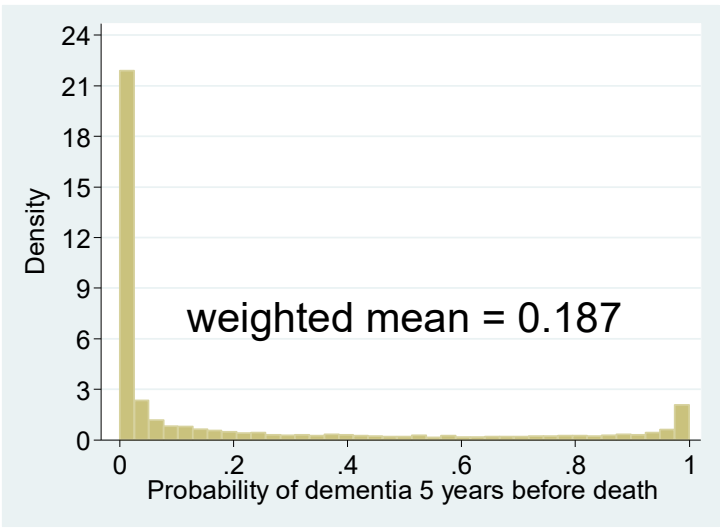
Panel C: Wave-to-wave changes in unadjusted dementia probabilities



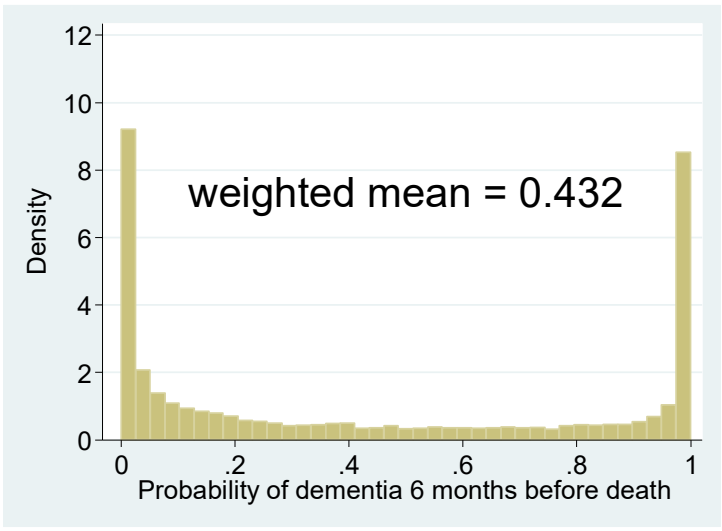
Panel D: Wave-to-wave changes in adjusted dementia probabilities

*Sample: Individuals who die after age 65 (N=9,479), in all waves after age 65 (total number of person year observations is 35,252). Panels C & D are further restricted to observations observed who are observed in two consecutive waves (total number is 25,773). The unadjusted dementia probabilities show the predicted values of the ordered probit models estimated from the ADAMS-HRS data, and they are based on information from a single HRS wave. The adjusted probabilities show the predicted values from a dynamic model of cognition fitted on the dementia probabilities. These probabilities use information from all waves.

Figure A2. Histogram of the lifetime risk of living with dementia for six months or five years



Panel A: Five years before death



Panel B: Six months before death

Table A1. Imputation model of primary insurance amount

	Coef.	95% CI
Birthyear: 1901-1919	ref.	ref.
Birthyear: 1920-1929	-608·0	[-1895·5; 679·5]
Birthyear: 1930-1939	-975·6	[-2354; 402·8]
Birthyear: 1940-1949	1744·9**	[320; 3169·8]
Female	-2327·3***	[-2905·5; -1749·1]
White non-Hispanic	ref.	ref.
Black non-Hispanic	-1085·5***	[-1902·2; -268·8]
Other non-Hispanic	-2262·2***	[-3703·6; -820·8]
Hispanic	-1569·5***	[-2512·6; -626·4]
LT HS	ref.	ref.
GED/HS Dip.	259·7	[-540·2; 1059·6]
Some College	389·8	[-499·8; 1279·4]
College	476·7	[-428; 1381·4]
Years worked	71·1***	[50·7; 91·5]
Years worked zero/missing	1377·7**	[4·4; 2751]
Average log earnings	1558·2***	[1229·5; 1886·9]
Zero/missing earnings	15012·9***	[11678·4; 18347·4]
Average log income	837·9***	[335·8; 1340]
Average log wealth	350·9***	[152·9; 548·9]
Zero/Negative/missing wealth	744·4	[-2017·2; 3506]
Constant	-16131·0***	[-20873·9; -11388·1]
R-squared	0·403	
Number of Donors	1790	
Number of Recipients	713	
Total sample size	21462	

Notes: Sample: HRS, 1998-2014, Age 65+. *, **, and *** indicate statistical significance at 10, 5 and 1% levels respectively.

Table A2. The distribution of case types when defining PIA, HRS, 1998-2014, Age 65+

	Was the claiming age report altered?		
Cases	No	Yes	Total
PIA from a standard respondent	9,963	5,520	15,483
PIA from a disabled respondent	2,184	38	2,222
PIA from a widowed respondent	687	2,056	2,743
PIA from a widowed & disabled respondent	301	210	511
Imputed PIA (hasn't claimed benefits yet)	472	31	503
Total	13,607	7,855	21,462

Table A3. Summary statistics of PIA by subgroups, HRS, 1998-2014, Age 65+

	N	mean	sd	p5	median	p95
Total	21,462	16,417	6,084	6,017	16,750	31,704
Male	9,327	17,140	5,978	6,691	17,576	31,704
Female	12,135	15,862	6,106	5,642	16,137	31,704
LT HS	6,222	13,864	5,504	5,186	13,922	28,660
High school	7,457	16,876	5,662	7,083	17,407	31,704
Some College	4,092	17,407	5,943	6,900	17,866	31,704
College	3,685	18,703	6,526	6,166	19,499	31,704
White non-Hispanic	16,122	17,343	5,768	7,167	17,736	31,704
Black non-Hispanic	3,078	13,927	6,176	4,369	13,585	30,293
Other non-Hispanic	449	14,628	7,003	4,402	14,311	31,704
Hispanic	1,806	12,858	5,829	4,109	12,112	27,962
Not disabled	18,461	16,876	5,964	6,564	17,237	31,704
Disabled	3,001	13,594	6,056	4,337	13,444	28,417
Married/partnered	14,786	17,410	5,782	7,254	17,843	31,704
Widowed	3,670	14,760	5,906	5,236	14,754	31,704
Single, not widowed	3,005	13,560	6,398	2,830	13,417	30,336

Table A4. Ordered probit model of cognitive status (1. demented, 2. CIND, 3. normal), HRS-ADAMS, self-interviews

	No mortality		With mortality	
	Coef.	95% CI	Coef.	95% CI
Female	-0·788	[-3·212; 1·636]	-1·254	[-3·827; 1·319]
Non-white	0.335**	[0·041; 0·629]	0.275*	[-0·031; 0·581]
Hispanic	0.491***	[0·128; 0·854]	0.471**	[0·103; 0·839]
Widowed	-0·005	[-0·385; 0·375]	0·051	[-0·343; 0·445]
Widowed X Female	0·199	[-0·291; 0·689]	0·129	[-0·377; 0·635]
(Age-65)/10	-1·034	[-4·701; 2·633]	-2·191	[-5·97; 1·588]
(Age-65)/10 squared	0·133	[-1·937; 2·203]	0·829	[-1·305; 2·963]
(Age-65)/10 cube	-0·01	[-0·369; 0·349]	-0·115	[-0·485; 0·255]
(Age-65)/10 X Female	0·621	[-3·785; 5·027]	1·992	[-2·551; 6·535]
(Age-65)/10 squared X Female	-0·192	[-2·616; 2·232]	-0·854	[-3·355; 1·647]
(Age-65)/10 cube X Female	0·022	[-0·39; 0·434]	0·116	[-0·307; 0·539]
HS dropouts	ref.	ref.	ref.	#VALUE!
High school	-0.331**	[-0·623; -0·039]	-0.344**	[-0·646; -0·042]
Some college	-0·079	[-0·422; 0·264]	-0·099	[-0·456; 0·258]
College	-0·305	[-0·705; 0·095]	-0.361*	[-0·778; 0·056]
Memory excellent, very good or good	ref.	ref.	ref.	#VALUE!
Memory fair	-0·102	[-0·337; 0·133]	-0·104	[-0·347; 0·139]
Memory poor	-0.325*	[-0·705; 0·055]	-0·291	[-0·679; 0·097]
Health excellent, very good or good	ref.	ref.	ref.	#VALUE!
Health fair	-0.308**	[-0·565; -0·051]	-0·18	[-0·447; 0·087]
Health poor	-0·035	[-0·396; 0·326]	0·108	[-0·26; 0·476]
IADL	-0·031	[-0·153; 0·091]	-0·008	[-0·133; 0·117]
ADL	0·037	[-0·083; 0·157]	0·056	[-0·066; 0·178]
Ever had hypertension	-0·038	[-0·256; 0·18]	-0·008	[-0·233; 0·217]
Ever had diabetes	-0·182	[-0·439; 0·075]	-0·065	[-0·332; 0·202]
Ever had stroke	-0.677***	[-0·987; -0·367]	-0.684***	[-1·003; -0·365]
Ever had psychiatric problems	-0·205	[-0·507; 0·097]	-0·225	[-0·533; 0·083]
CESD depression scale	0·009	[-0·046; 0·064]	0·013	[-0·044; 0·07]
Time-to-death in decades			0·679	[-1·524; 2·882]
Time-to-death squared			0·501	[-1·482; 2·484]
Survived to 2014			-0·172	[-1·085; 0·741]
Time-to-death in decades X Female			-0·701	[-3·574; 2·172]
Time-to-death squared X Female			-0·405	[-2·947; 2·137]
Survived to 2014 X Female			1.200**	[0·051; 2·349]
Immediate word recall	0.156***	[0·066; 0·246]	0.150***	[0·056; 0·244]
Delayed word recall	0.211***	[0·131; 0·291]	0.196***	[0·112; 0·28]
Serial 7	0.177***	[0·104; 0·25]	0.178***	[0·102; 0·254]
Backward counting	0·07	[-0·089; 0·229]	0·131	[-0·036; 0·298]
Remembering dates	0.346***	[0·226; 0·466]	0.338***	[0·215; 0·461]
Knowing scissor	-0·205	[-0·942; 0·532]	-0·173	[-0·933; 0·587]
Knowing cactus	0.252*	[-0·028; 0·532]	0.286*	[0; 0·572]
Knowing the president	0.344**	[0·023; 0·665]	0.408**	[0·077; 0·739]
Immediate word recall missing	0·206	[-0·394; 0·806]	0·144	[-0·475; 0·763]
Delayed word recall missing	0.365*	[-0·025; 0·755]	0·324	[-0·078; 0·726]
Serial 7 missing	-0·004	[-0·394; 0·386]	0·001	[-0·399; 0·401]
Backward counting missing	0·251	[-0·511; 1·013]	0·427	[-0·355; 1·209]
1st cut point	0·069	[-2·058; 2·196]	0·446	[-1·777; 2·669]
2nd cut point	1·614	[-0·518; 3·746]	2.147*	[-0·083; 4·377]
Pseudo R-squared	0·364		0·409	
N	657		657	

Table A5. Ordered probit model of cognitive status (demented, CIND, normal), HRS-ADAMS, proxy interviews

	No mortality		With mortality	
	Coef.	95% CI	Coef.	95% CI
Female	2·367	[-1·674; 6·408]	1·011	[-3·65; 5·672]
Non-white	-0·301	[-1·224; 0·622]	-0·152	[-1·148; 0·844]
Hispanic	0·165	[-0·672; 1·002]	-0·129	[-1·052; 0·794]
Widowed	-0.974*	[-2·034; 0·086]	-1.189*	[-2·437; 0·059]
Widowed X Female	1·005	[-0·641; 2·651]	1.828*	[-0·118; 3·774]
(Age-65)/10	-0·068	[-2·112; 1·976]	-0·077	[-2·323; 2·169]
(Age-65)/10 squared	-0·01	[-0·492; 0·472]	0·053	[-0·474; 0·58]
(Age-65)/10 X Female	-3·263	[-7·289; 0·763]	-2·559	[-7·094; 1·976]
(Age-65)/10 squared X Female	0·747	[-0·249; 1·743]	0·61	[-0·523; 1·743]
HS dropouts	ref.	ref.	ref.	ref.
High school	0·016	[-0·831; 0·863]	0·28	[-0·655; 1·215]
Some college	-0·207	[-1·765; 1·351]	-0·525	[-2·25; 1·2]
College	-0·167	[-1·545; 1·211]	-0·129	[-1·938; 1·68]
Memory excellent, very good or good	ref.	ref.	ref.	ref.
Memory fair	-0·26	[-0·942; 0·422]	-0·151	[-0·884; 0·582]
Memory poor	-1.215**	[-2·168; -0·262]	-1.041**	[-2·06; -0·022]
Health excellent, very good or good	ref.	ref.	ref.	ref.
Health fair	-0·369	[-1·116; 0·378]	-0·112	[-0·943; 0·719]
Health poor	-0.668*	[-1·425; 0·089]	-0·375	[-1·222; 0·472]
IADL	-0.204*	[-0·433; 0·025]	-0·173	[-0·422; 0·076]
ADL	0·025	[-0·216; 0·266]	0·177	[-0·093; 0·447]
Ever had stroke	-0·175	[-0·879; 0·529]	-0·538	[-1·33; 0·254]
Time-to-death in decades			5.927***	[1·584; 10·27]
Time-to-death squared			-3.961*	[-8·055; 0·133]
Survived to 2014			2.009*	[-0·304; 4·322]
No cognitive limitation	ref.	ref.	ref.	ref.
Some cognitive limitation	-0·624	[-1·416; 0·168]	-0·629	[-1·523; 0·265]
Cognitive limitations	-1.041**	[-1·915; -0·167]	-1.177**	[-2·214; -0·14]
Jorm IQCODE	-0.659*	[-1·343; 0·025]	-0.774**	[-1·464; -0·084]
Any imputation	-0·312	[-1·296; 0·672]	-0·783	[-1·945; 0·379]
1st cut point	-4.522***	[-7·566; -1·478]	-2.882*	[-6·185; 0·421]
2nd cut point	-2.721*	[-5·704; 0·262]	-0·527	[-3·812; 2·758]
Pseudo R-squared	0·524		0·605	
N	199		199	

Table A6. Output of the dynamic model of dementia probabilities, HRS, 1998-2014, Age 65+ and dies before 2014

	Coef.	95% CI
Coefficients in mean cognition		
Constant	1.866***	[1·752; 1·98]
(Age-65)/10 (A)	-0.713***	[-0·868; -0·558]
A squared	-0.402***	[-0·504; -0·3]
A cube	0.107***	[0·085; 0·129]
Time-to-death in decades (T)	3.189***	[2·871; 3·507]
Time-to-death squared	-9.228***	[-10·641; -7·815]
Time-to-death cube	15.371***	[12·417; 18·325]
Time-to-death fourth power	-11.060***	[-13·594; -8·526]
Time-to-death fifth power	2.902***	[2·147; 3·657]
(Age-65)/10 X Time-to-death	1.262***	[1·14; 1·384]
(Age-65)/10 X (Time-to-death squared)	-0.582***	[-0·674; -0·49]
Female	-0.188***	[-0·284; -0·092]
High school dropout	-0.358***	[-0·472; -0·244]
High school graduate	ref.	ref.
Some college	0.141*	[-0·01; 0·292]
College graduate	0.468***	[0·301; 0·635]
White	ref.	ref.
Black	-0.426***	[-0·551; -0·301]
Other race	-0.233*	[-0·509; 0·043]
Hispanic	-0.077	[-0·271; 0·117]
T X Female	-0.541***	[-0·633; -0·449]
T X High school dropout	-0.158***	[-0·266; -0·05]
T X Some college	-0.015	[-0·152; 0·122]
T X College graduate	0.049	[-0·106; 0·204]
T X Black	-0.039	[-0·162; 0·084]
T X Other race	-0.128	[-0·406; 0·15]
T X Hispanic	-0.083	[-0·261; 0·095]
A X Female	0.026	[-0·035; 0·087]
A X High school dropout	0.059	[-0·014; 0·132]
A X Some college	0.012	[-0·084; 0·108]
A X College graduate	-0.151***	[-0·259; -0·043]
A X Black	0.004	[-0·084; 0·092]
A X Other race	0.023	[-0·177; 0·223]
A X Hispanic	-0.010	[-0·143; 0·123]
Coefficients in log std of the random walk innovation		
Female	0.239***	[0·176; 0·302]
High school dropout	-0.074*	[-0·148; 0]
High school graduate	ref.	ref.
Some college	0.083*	[-0·003; 0·169]
College graduate	0.161***	[0·067; 0·255]
White	ref.	ref.
Black	-0.026	[-0·116; 0·064]
Other race	-0.050	[-0·244; 0·144]
Hispanic	0.065	[-0·051; 0·181]
Constant	0.002	[-0·067; 0·071]
Coefficients in log std of the noise term		
Female	0.031	[-0·008; 0·07]
High school dropout	-0.161***	[-0·208; -0·114]
High school graduate	ref.	ref.
Some college	-0.152***	[-0·209; -0·095]
College graduate	-0.018	[-0·081; 0·045]

White	ref.	ref.
Black	-0.085***	[-0·138; -0·032]
Other race	-0.133**	[-0·247; -0·019]
Hispanic	-0.125***	[-0·198; -0·052]
Constant	0.146***	[0·105; 0·187]
N	9479	

* Age and time-to-death are measured one year ahead of HRS interviews.

Table A7. Used parameter values for the simulation of the synthetic dataset

β_0	9.5
β_1	-1
β_2	0.1
σ_δ	0.5
σ_ε	1
λ^1	$\exp(-0.1)$
θ_0	-7.95
θ_1	-0.1
τ_0	0.55
τ_1	0.1
σ_υ	$\exp(-2)$

Table A8. Estimated dementia probabilities at HRS interviews in a synthetic dataset, results from the first step of our procedure

	N	True value (fraction demented)	Estimated probability of dementia
Dementia at HRS interviews, all person-year observations	321,411	15.55%	
Dementia one year after HRS interviews, all person-year observations	321,411	16.95%	16.93%
Dementia one year after HRS interviews, survived at least one year	307,111	16.28%	16.27%
Dementia at HRS interviews, all person-year observations who die within the sample	115,223	22.31%	
Dementia one year after HRS interviews, all person-year observations who die within the sample	115,223	24.06%	24.03%
Dementia one year after HRS interviews, survived at least one year but dies within the sample	100,923	23.01%	23.26%

Table A9. Estimated dementia probabilities at death and six months prior to death in a synthetic dataset, results from the second step of our procedure

	N	True value (fraction demented)	Estimated probability of dementia
Dementia, 6 months before death	27,491	29.46%	29.71%
Dementia at death	27,491	30.45%	30.66%

Table A10. Probabilities of living with dementia for six months, one, two, or five years by demographic groups

	5 years		2 years		1 year		6 months	
	mean	95% CI	mean	95% CI	mean	95% CI	mean	95% CI
Total 70+	0·187	[0·124; 0·25]	0·317	[0·267; 0·367]	0·386	[0·335; 0·437]	0·432	[0·374; 0·489]
Males	0·125	[0·063; 0·186]	0·259	[0·201; 0·318]	0·338	[0·274; 0·402]	0·391	[0·318; 0·465]
Females	0·237	[0·165; 0·31]	0·364	[0·308; 0·419]	0·425	[0·366; 0·484]	0·464	[0·396; 0·533]
Less than HS	0·233	[0·171; 0·296]	0·376	[0·327; 0·425]	0·451	[0·4; 0·502]	0·499	[0·441; 0·557]
High school	0·171	[0·104; 0·239]	0·301	[0·241; 0·36]	0·371	[0·308; 0·433]	0·417	[0·348; 0·486]
Some college	0·146	[0·068; 0·223]	0·255	[0·182; 0·328]	0·312	[0·24; 0·384]	0·353	[0·275; 0·431]
College or more	0·132	[0·04; 0·223]	0·250	[0·16; 0·34]	0·312	[0·223; 0·402]	0·354	[0·259; 0·449]
White	0·178	[0·114; 0·242]	0·306	[0·254; 0·358]	0·374	[0·321; 0·427]	0·420	[0·359; 0·48]
Black	0·268	[0·191; 0·346]	0·418	[0·352; 0·485]	0·495	[0·429; 0·561]	0·541	[0·473; 0·61]
Other race	0·192	[0·102; 0·283]	0·318	[0·23; 0·406]	0·389	[0·303; 0·475]	0·436	[0·347; 0·525]
Non-Hispanic	0·186	[0·123; 0·249]	0·316	[0·265; 0·366]	0·385	[0·333; 0·436]	0·430	[0·372; 0·488]
Hispanic	0·211	[0·13; 0·292]	0·345	[0·269; 0·422]	0·414	[0·339; 0·49]	0·460	[0·38; 0·539]
Lowest quintile	0·251	[0·19; 0·312]	0·388	[0·341; 0·435]	0·458	[0·409; 0·507]	0·503	[0·448; 0·558]
2nd quintile	0·210	[0·145; 0·276]	0·347	[0·295; 0·398]	0·418	[0·367; 0·469]	0·464	[0·407; 0·522]
3rd quintile	0·174	[0·108; 0·239]	0·308	[0·253; 0·362]	0·379	[0·323; 0·435]	0·426	[0·364; 0·488]
4th quintile	0·154	[0·089; 0·219]	0·275	[0·218; 0·332]	0·343	[0·285; 0·4]	0·388	[0·324; 0·453]
Highest quintile	0·151	[0·085; 0·217]	0·275	[0·216; 0·334]	0·339	[0·279; 0·4]	0·383	[0·315; 0·45]
Died at age 70-75	0·073	[0·003; 0·143]	0·129	[0·065; 0·193]	0·164	[0·096; 0·232]	0·190	[0·114; 0·265]
Died at age 75-80	0·129	[0·056; 0·202]	0·220	[0·153; 0·286]	0·270	[0·201; 0·338]	0·305	[0·228; 0·382]
Died at age 80-85	0·187	[0·119; 0·255]	0·308	[0·247; 0·37]	0·376	[0·313; 0·439]	0·422	[0·352; 0·492]
Died at age 85-90	0·216	[0·147; 0·286]	0·375	[0·318; 0·432]	0·457	[0·399; 0·515]	0·510	[0·446; 0·575]
Died at age 90-95	0·264	[0·198; 0·331]	0·456	[0·403; 0·509]	0·553	[0·497; 0·61]	0·613	[0·549; 0·676]
Died at age 95+	0·325	[0·246; 0·403]	0·524	[0·427; 0·621]	0·618	[0·51; 0·727]	0·676	[0·563; 0·789]
last HRS wave was 1998	0·236	[0·165; 0·306]	0·342	[0·292; 0·392]	0·400	[0·351; 0·448]	0·439	[0·384; 0·494]
2000	0·215	[0·144; 0·287]	0·330	[0·279; 0·382]	0·395	[0·336; 0·453]	0·438	[0·373; 0·502]
2002	0·192	[0·129; 0·254]	0·314	[0·257; 0·371]	0·379	[0·319; 0·44]	0·423	[0·357; 0·488]
2004	0·184	[0·12; 0·248]	0·326	[0·268; 0·384]	0·397	[0·34; 0·454]	0·444	[0·379; 0·508]
2006	0·193	[0·13; 0·255]	0·332	[0·276; 0·388]	0·402	[0·348; 0·455]	0·448	[0·387; 0·508]
2008	0·160	[0·095; 0·225]	0·292	[0·234; 0·35]	0·365	[0·304; 0·427]	0·415	[0·343; 0·487]
2010	0·153	[0·084; 0·223]	0·304	[0·246; 0·361]	0·381	[0·321; 0·441]	0·430	[0·368; 0·492]
2012	0·155	[0·087; 0·223]	0·295	[0·234; 0·357]	0·369	[0·31; 0·428]	0·419	[0·355; 0·483]

Table A10. Alternative estimates of the lifetime risk of dementia from the last core HRS interview before death by demographic groups

		One year after last HRS interview			
		no mortality adjustment		with mortality adjustment	
Sample	N	Mean	95% CI	Mean	95% CI
Total 70+	8733	0·356	[0·353; 0·405]	0·420	[0·397; 0·486]
Males	3987	0·288	[0·271; 0·342]	0·377	[0·326; 0·472]
Females	4746	0·413	[0·406; 0·471]	0·455	[0·42; 0·532]
Less than HS	3744	0·407	[0·397; 0·465]	0·487	[0·459; 0·557]
High school	2534	0·352	[0·334; 0·418]	0·406	[0·373; 0·485]
Some college	1415	0·290	[0·261; 0·363]	0·342	[0·304; 0·416]
College or more	1040	0·289	[0·257; 0·363]	0·342	[0·29; 0·44]
White	7265	0·348	[0·342; 0·398]	0·410	[0·385; 0·476]
Black	1246	0·427	[0·407; 0·525]	0·504	[0·477; 0·605]
Other race	222	0·338	[0·292; 0·451]	0·414	[0·366; 0·533]
Non-Hispanic	8167	0·357	[0·353; 0·406]	0·419	[0·395; 0·485]
Hispanic	566	0·337	[0·312; 0·439]	0·431	[0·397; 0·531]
PIA, 1st quintile	1747	0·396	[0·405; 0·474]	0·464	[0·457; 0·551]
PIA, 2nd quintile	1747	0·387	[0·364; 0·429]	0·452	[0·414; 0·511]
PIA, 3rd quintile	1747	0·366	[0·345; 0·41]	0·425	[0·389; 0·489]
PIA, 4th quintile	1746	0·335	[0·311; 0·375]	0·402	[0·355; 0·453]
PIA, 5th quintile	1746	0·299	[0·312; 0·378]	0·360	[0·353; 0·453]
Single at death	5161	0·411	[0·403; 0·46]	0·459	[0·432; 0·525]
Married at death	3572	0·278	[0·265; 0·336]	0·364	[0·332; 0·44]
Died 70-75	1275	0·143	[0·111; 0·175]	0·221	[0·172; 0·269]
Died 75-80	1613	0·238	[0·205; 0·271]	0·323	[0·27; 0·375]
Died 80-85	1884	0·367	[0·333; 0·401]	0·433	[0·379; 0·488]
Died 85-90	1915	0·463	[0·424; 0·502]	0·516	[0·461; 0·572]
Died 90-95	1382	0·570	[0·521; 0·62]	0·606	[0·549; 0·663]
Died 95+	664	0·624	[0·521; 0·726]	0·663	[0·565; 0·761]
Wave 1998	1182	0·341	[0·326; 0·397]	0·410	[0·38; 0·48]
Wave 2000	1217	0·347	[0·334; 0·405]	0·418	[0·389; 0·488]
Wave 2002	1075	0·359	[0·34; 0·411]	0·427	[0·39; 0·49]
Wave 2004	1080	0·366	[0·346; 0·421]	0·439	[0·402; 0·508]
Wave 2006	1066	0·381	[0·374; 0·448]	0·432	[0·411; 0·513]
Wave 2008	1262	0·332	[0·321; 0·39]	0·380	[0·352; 0·448]
Wave 2010	914	0·374	[0·365; 0·446]	0·443	[0·417; 0·532]
Wave 2012	937	0·354	[0·344; 0·421]	0·420	[0·386; 0·505]

*Bootstrapped standard errors in brackets. The table shows alternative lifetime risk estimates: the average dementia probabilities (ordered probit estimates in step 1 of our procedure) in the last HRS wave before death. The ordered probit models predicted dementia status in ADAMS as a function of HRS variables in the prior wave. The average time difference between these waves was slightly over one year. Thus, these estimates may be interpreted as showing dementia probabilities one year after the last HRS interview. We show two sets of estimates. The first does not include mortality information (time-to-death) in the model and it is likely downward biased in this sample as discussed in Section E3. The second model includes time-to-death in the model.

A8: References

1. Sonnega A, Faul JD, Ofstedal MB, Langa KM, Phillips JW, Weir DR. Cohort profile: The Health and Retirement Study (HRS). *Int J Epidemiol*. 2014;43(2):576-585.
2. RAND HRS Data, Version P. Produced by the RAND Center for the study of aging, with funding from the National Institute on Aging and the Social Security Administration; 2016.
3. Crimmins EM, Kim JK, Langa, KM, Weir, DR. (2011). Assessment of cognition using surveys and neuropsychological assessment: The Health and Retirement Study and the Aging, Demographics, and Memory Study. *J Gerontol B Psychol Sci Soc Sci*. 2011;66(suppl 1);i162-i171.
4. 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.
5. Hyslop DR, Imbens GW. Bias from Classical and Other Forms of Measurement Error. J Bus Econ Stat, 2001;19(4):475-81.
6. Prince M, Albanese E, Guerchet M, Prina M. World Alzheimer Report 2014, Dementia and Risk Reduction, An analysis of protective and modifiable factors. Alzheimer's Disease International (ADI), London. 2014.
7. Horowitz JL. Chapter 52 The Bootstrap. Handbook of Econometrics, 2001;5:3159-3228.