Research Report

PETER HUDOMIET, MICHAEL D. HURD, SUSANN ROHWEDDER

Identifying Early Predictors of Cognitive Impairment and Dementia in a Large Nationally Representative U.S. Sample

For more information on this publication, visit **www.rand.org/t/RRA3207-1**.

Published by the RAND Corporation, Santa Monica, Calif.

Cover: imtmphoto / AdobeStock

About This Report

Dementia affects a large and growing number of older adults in the United States and worldwide. Although many risk factors for dementia have been identified in prior research, an improved ability to predict possible dementia risk years before onset would have several important benefits, such as helping older adults prepare for the risk of developing this condition. Dementia prediction would also enable health-care providers and the government to identify which parts of the population are at highest risk and could permit better targeting of efforts and resources to delay the onset of dementia or mitigate its effects. Furthermore, improved prediction models would sharpen dementia trend forecasts, which would assist in planning for future care needs.

The primary goal of our analyses in this report is to identify the early predictors of dementia and cognitive impairment, as measured in a large survey, to inform the scientific community, health-care providers, clinicians, and other stakeholders about who might be at elevated risk of developing dementia several years before onset.

This report is part of a series on dementia care, which includes the following publications thus far::

- On the value to individuals of knowing about their dementia risk, even when there are no effective treatment options: Michael D. Hurd, Peter Hudomiet, and Susann Rohwedder, *Benefits of Seeking Early Detection of Cognitive Decline*, RAND Corporation, RR-A3207-3, 2024.
- On patient demand for screening, diagnosis, and treatment: Susann Rohwedder, Peter Hudomiet, and Michael D. Hurd, *Individuals' Interest in Cognitive Screening, Dementia Diagnosis, and Treatment: New Estimates from a Population-Representative Sample*, RAND Corporation, RR-A2643-2, 2024.
- On the ability of the health-care system to deliver testing and treatment to the population in the face of capacity constraints: Jodi L. Liu, Lawrence Baker, Annie Chen, Jessie Wang, and Federico Girosi, *Modeling Early Detection and Geographic Variation in Health System Capacity for Alzheimer's Disease–Modifying Therapies*, RAND Corporation, RR-A2643-1, 2024; and Lawrence Baker, Annie Chen, Jessie Wang, Federico Girosi, and Jodi L. Liu, *The Future of Alzheimer's Care in America: How Patient Demand and Health Care System Capacity Could Affect the Delivery of Alzheimer's Disease–Modifying Treatments*, RAND Corporation, TL-A2643-1, 2024.

This study builds on an extensive body of work by the authors that was funded by the National Institute on Aging. Funding for this study was provided by Genentech, Inc.

Social and Behavioral Policy Program

RAND Social and Economic Well-Being is a division of RAND that seeks to actively improve the health and social and economic well-being of populations and communities throughout the world. This research was conducted in the Social and Behavioral Policy Program within RAND Social and Economic Well-Being. The program focuses on such topics as risk factors and prevention programs, social safety net programs and other social supports, poverty, aging, disability, child and youth health and well-being, and quality of life, as well as other policy concerns that are influenced by social and behavioral actions and systems that affect well-being. For more information, email sbp@rand.org.

Acknowledgments

We thank Kimberly Jinnett, Liz Mearns, Karina Raimundo, and Diana Slowiejko at Genentech for their feedback and support. Stephen Lee and Joanna Carroll provided excellent programming support for the project. We also appreciate the peer reviews by an anonymous external reviewer and Roland Sturm at RAND.

Summary

The number of older adults in the United States and worldwide is growing, and because age is the most important predictor of dementia, the number of persons living with this condition is also expected to grow. Detecting elevated risk for dementia years before its onset would help older adults prepare for the risk of developing this condition, permit health-care providers and the government to more efficiently target resources to delay the onset or mitigate the effects of the condition, and guide policymakers to invest in infrastructure and human capital to meet care demands. In this report, we aim to identify predictors of dementia and cognitive impairment for individuals in the United States up to 20 years in advance using the cognition and dementia measures from the Health and Retirement Study (HRS).

Approach

In this report, we evaluate the predictive power of 181 potential risk factors for dementia using a validated probabilistic measure of dementia and cognitive impairment that was developed in prior research. This measure is available in the HRS—a large, nationally representative, longitudinal survey—which allows us to evaluate the predictive power of many potential dementia risk factors, such as demographics, socioeconomic status (SES), labor-market measures, lifestyle and health behaviors (such as exercising and smoking), subjectively reported and objectively measured health, genes, parental health, cognitive abilities, and psychosocial factors (such as personality traits, social activities, and loneliness). We estimate how these factors predict cognitive impairment and dementia of individuals two, four, and twenty years after age 60.

Key Findings

This study yielded the following findings:

- We predicted quantitatively meaningful and statistically significant variation in dementia prevalence among persons approximately age 80 according to their observed characteristics when they were about age 60.
- In terms of explained variation, an individual's baseline cognitive abilities, health, and functional limitations were the strongest predictors of dementia, whereas parental health, family size, marital history, and demographics were the weakest ones.
- We found that having poor physical health, a stroke, lower cognitive abilities, functional limitations, and particular genes strongly predict future incidence and prevalence of cognitive impairment and dementia, which is in-line with prior literature.
- Individuals born in the Southern United States face higher chances of developing cognitive impairment and dementia, even when we controlled for an expanded set of factors.

- Other factors associated with a higher chance of developing cognitive impairment or dementia are not having a private health insurance plan at age 60, never having worked or having worked only a few years, having diabetes or a body mass index of 35 or more at age 60, never drinking alcohol or drinking excessively, never exercising, scoring low on various physical tests (such as breathing, grip strength, walking speed, and balance), being less conscientious, and having low engagement in hobbies and novel information activities.
- Black and Hispanic individuals face statistically significant higher chances of cognitive impairment and dementia, but these differentials shrink or disappear when accounting for observable differences, such as SES.

Recommendations

The results of this study can be used to improve dementia prediction and prevention efforts. Increased precision in the prediction of dementia prevalence in the U.S. adult population would help to plan for the evolution of the very high monetary and caregiving costs associated with dementia. At the individual level, identifying people at elevated risk for dementia would permit the channeling of resources to them that could encourage them to engage in advance planning and to pursue a lifestyle that promotes brain health. Regarding prevention, our findings indicate which interventions and behavioral changes might be the most promising to evaluate in future research. Although our observational data and methods do not quantify the causal effects of the risk factors, our study indicates which interventions are most likely beneficial. For example, older individuals striving to maintain high cognitive function for a longer time might benefit from early lifestyle modifications, such as performing physical exercise, working additional years, engaging in hobbies and novel information activities after retirement, and maintaining good physical health. Policymakers and health-care providers should consider ways of promoting these healthy behaviors in the adult population and strengthening individuals' access to quality health care.

Contents

About This Report ... iii
Summary ... v
Figures and Tables ... ix

CHAPTER 1
Introduction and Approach ... 1
Overview of Approach ... 1
Conceptual Framework of Prediction Models ... 2

CHAPTER 2
Data and Measurement of Main Outcome and Predictor Variables ... 5
Outcome Variables ... 5
Predictor Variables ... 7
Sample and Descriptive Statistics ... 8

CHAPTER 3
Prediction Models ... 13
Basic Models ... 13
Multivariate Models ... 13
Missing Values ... 14

CHAPTER 4
Results ... 15
Basic Model Results ... 15
Multivariate Model Results ... 25

CHAPTER 5
Limitations and Conclusions ... 35
Limitations ... 35
Conclusions ... 35

Abbreviations ... 37
References ... 39

Available at www.rand.org/t/RRA3207-1

ANNEX
Identifying Early Predictors of Cognitive Impairment and Dementia in a Large Nationally Representative U.S. Sample: Supplemental Annex of Figures and Tables

Figures and Tables

Figures

1.1. Conceptual Framework of the Relationship Between Dementia Risk Factors and Cognitive Impairment and Dementia ... 2
4.1. Age-Adjusted Four-Year Dementia Incidence, by Selected Demographic, Socioeconomic Status, and Labor Predictors ... 18
4.2. Age-Adjusted Four-Year Dementia Incidence, by Selected Lifestyle and Health Behavior Predictors ... 20
4.3. Age-Adjusted Four-Year Dementia Incidence, by Selected Psychosocial Predictors ... 22
4.4. Age-Adjusted Four-Year Dementia Incidence, by Selected Self-Reported Health and Functional Limitations Predictors ... 23
4.5. Age-Adjusted Four-Year Dementia Incidence, by Selected Physical Health Predictors ... 24
4.6. Age-Adjusted Four-Year Dementia Incidence, by Selected Cognitive Predictors ... 26

Tables

2.1. Measures Used to Predict Cognitive Impairment and Dementia ... 7
2.2. Descriptive Statistics About the Three Samples ... 9
4.1. Strongest Predictors of Dementia in Each Category in the Basic Models ... 15
4.2. Significant Predictors in the Multivariate Models of Dementia ... 27
4.3. Significant Predictors in the Multivariate Models of Cognitive Impairment ... 29

CHAPTER 1

Introduction and Approach

Dementia affects a large and growing number of older adults in the United States and worldwide, and—in terms of total costs—dementia is the most expensive medical condition primarily because of the high costs of formal and informal care (Alzheimer's Association, 2023; Wimo et al., 2023). Although many risk factors for dementia have been identified in prior research, accurately predicting whether someone will develop dementia remains challenging (Javeed et al., 2023). The ability to identify dementia risk factors several years before its onset would help older adults and their families prepare for the risk of developing this condition and permit the health-care providers and agencies that are involved in providing or paying for long-term care to identify which parts of the population are at highest risk and, thus, better target resources for delaying the onset or mitigating the effects of dementia.

In this report, we evaluate the predictive power of 181 potential dementia risk factors. Using data from the Health and Retirement Study (HRS)—a large, nationally representative, longitudinal survey that has measures of dementia and information on a large set of covariates—we investigate to what extent these factors predict cognitive impairment and dementia two, four, and twenty years after age 60. We developed the dementia and cognitive impairment measures in prior research, which we calibrated to a clinical dementia diagnosis in a smaller sample (N = 856) (for the prior study, see Hudomiet, Hurd, and Rohwedder, 2022). The large set of potential dementia risk factors that we evaluate include demographics, socioeconomic status (SES), labor-market measures, lifestyle and health behaviors (such as exercising and smoking), subjectively reported and objectively measured health, genes, parental health, cognitive abilities, and psychosocial factors (such as personality traits, social activities, and loneliness).

The primary goal of this study is to identify the strongest predictors of cognitive impairment and dementia, which could be used to evaluate older adults' risks of developing this condition. The secondary goal is to identify modifiable risk factors that might be used for interventions to slow cognitive decline. However, we note that although our findings might suggest certain causal channels between risk factors and outcomes, the observational methods used in this study cannot establish causal mechanisms, which is left for future research.

Overview of Approach

This study uses data from the HRS, a biennial, longitudinal survey of the U.S. population over age 50. In prior work, we developed cognition and dementia measures for all study participants ages 65 and older from the HRS waves in 1992 to 2016 (Hudomiet, Hurd, and Rohwedder, 2022). We discuss the key features of these measures in more detail in Chapter 2.

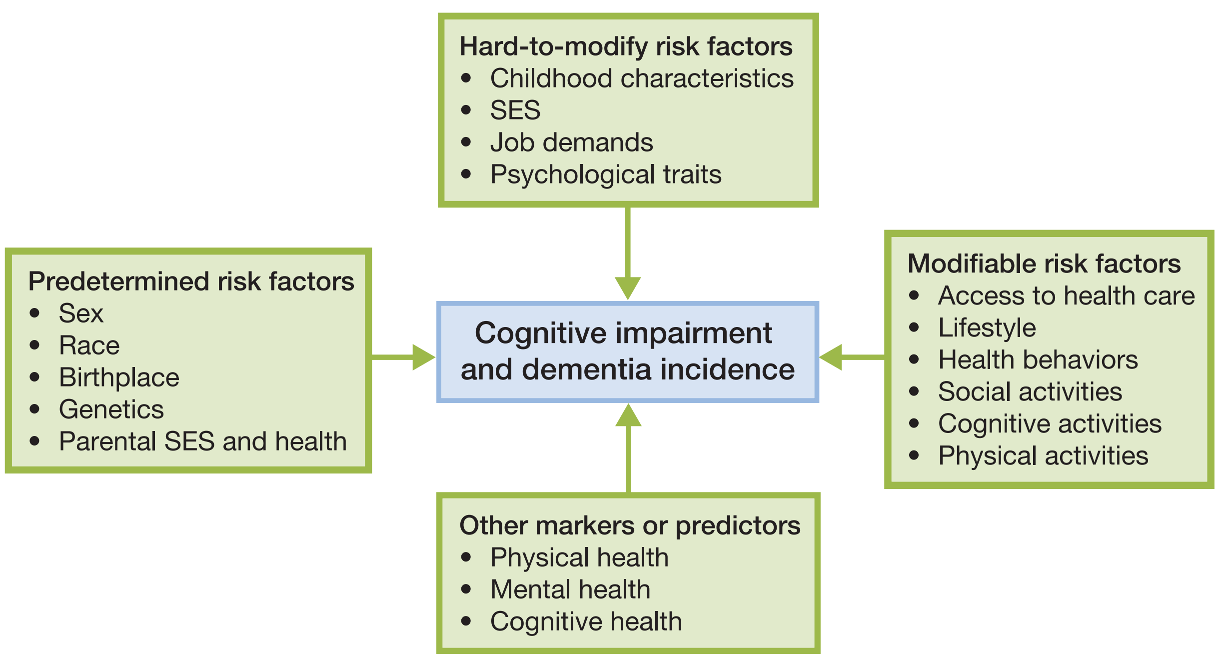
For the purpose of studying candidate predictors of dementia several years before onset, we use regression models to estimate the relationship between the risk factors and cognitive outcomes. We study two types of outcomes. The first is the two- and four-year incidence of dementia and of cognitive impairment, not dementia (CIND). The second is a long-term prediction model for dementia and CIND prevalence at age 80 using individuals' characteristics at age 60. The details of these prediction models are discussed in Chapter 3. We discuss the results of simple models and multivariate models, which test whether the predictor variables remain strongly associated with the outcomes when accounting for other factors, in Chapter 4, and in the final chapter, we summarize the limitations of these models and provide our overall conclusions.

An annex to this report, with supplemental figures and tables, is available at www.rand.org/t/RRA3207-1.

Conceptual Framework of Prediction Models

The conceptual framework underlying the prediction models is shown in Figure 1.1. The primary risk factor for CIND and dementia is age. The prevalence of dementia below age 65 is negligible but increases with age exponentially, reaching 50 percent at advanced ages. There is substantial heterogeneity in the U.S. adult population in the risk of developing these conditions. Some risk factors are *predetermined* (i.e., they cannot be modified or mitigated). Such risk factors include sex, race and ethnicity, birthplace, parental characteristics, and genetic markers, such as the Apoe4 gene (Pires and Rego, 2023). Prior literature found variation in dementia risk along these dimensions. For example, compared with men, women have a slightly elevated chance of living with dementia at advanced ages (Ferretti et al., 2018; Kim et al., 2018). Additionally, individuals of racial and ethnic minority backgrounds also face higher risks than non-Hispanic White individuals (Gianattasio, Ciarleglio, and Power, 2020).

FIGURE 1.1
Conceptual Framework of the Relationship Between Dementia Risk Factors and Cognitive Impairment and Dementia



Modifiable risk factors cover individual characteristics that can be changed to possibly reduce an individual's chances of developing cognitive impairment and dementia.[1] Some of these factors, such as childhood characteristics, SES, job characteristics, and psychological traits, can be modified in the population only over many years. Other factors—such as lifestyle; health behaviors (e.g., smoking, drinking); social, cognitive, and physical activities; and high-quality health care—are more amenable to immediate intervention. Prior literature found strong associations between these risk factors and dementia. For example, the chance of having dementia is substantially lower among individuals who have higher levels of education (Prince et al., 2014), those who have worked in cognitively demanding jobs (Andel, Finkel, and Pedersen, 2016; Then et al., 2014), and those who regularly exercise (Ahlskog et al., 2011).

Our prediction models also include factors that are not strictly considered risk factors but potential markers of early cognitive decline. For example, individuals who report being in poor health or who score low on cognition tests might have an elevated chance of dementia incidence in the future because these factors indicate the presence of prediagnosed cognitive problems that might eventually lead to dementia. We also use such factors in our prediction models but analyze them separately from the more-traditional risk factors.

[1] In this report, we use *cognitive impairment* to refer to the general concept of this condition and *CIND* to refer to the particular way of measuring this condition.

CHAPTER 2

Data and Measurement of Main Outcome and Predictor Variables

The data for this study come from the HRS, a nationally representative, biennial longitudinal survey of adults over age 50.[1] The first wave was conducted in 1992 with a target population of the cohorts born in 1931 through 1941 (Juster and Suzman, 1995). Older cohorts and a younger cohort (born in 1942 through 1947) were added in 1998, so that the 1998 study represented the population born in 1947 or earlier. Since 1998, refresher cohorts of 51- to 56-year-olds have been added to the HRS every six years to maintain a population representation of adults ages 51 and older. The sample includes about 20,000 individuals per wave; more than 45,000 individuals have participated in the HRS since its inception. The HRS oversamples Black and Hispanic individuals to ensure adequate sample size for studies focused on racial and ethnic disparities. Sample weights are available to adjust the demographic distribution of the sample to the general U.S. population.

The HRS is the largest survey of the U.S. elderly population and has detailed information about the individuals' cognitive abilities, dementia status, and risk factors. The survey has detailed information about demographics, SES, health-care use, spending, income and wealth, physical and mental health, and several measures of cognitive function. The exceptionally long duration of the HRS is of particular importance to this study; even though we restricted our sample to individuals interviewed at least once between 2000 and 2016, some respondents have been interviewed over a period of 24 years, which permits the long-term tracking of cognition. Our analyses focus on individuals ages 65 and older because the HRS's cognition and dementia measures are only elicited for this age group. There are 97,629 person-year observations for the U.S. population ages 65 and older in the HRS waves from 2000 to 2016.

Whenever possible, we use constructed variables from the RAND HRS Longitudinal File, a processed and widely used version that includes the most-important measures from the core HRS data.[2]

Outcome Variables

The HRS collects many cognition and memory measures, including items from the Telephone Interview for Cognitive Status, which was adapted from the widely used Mini-Mental State Examination cognition screening tool. These measures can be combined to obtain an index of cognitive functioning. The measures include self-rated memory (from excellent to poor), immediate and delayed word recall tests, backward counting and serial sevens subtraction tests, date and object naming tests, naming the president and vice president tests, and a vocabulary test (Fisher et al., 2015). To use this information to assign a diagnosis of CIND or demen-

[1] The HRS is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and conducted by the University of Michigan (Survey Research Center, Institute for Social Research, University of Michigan, 2023).

[2] The RAND HRS Longitudinal File is a user-friendly dataset based on the HRS core data (Bugliari et al., 2023). This file was developed at RAND with funding from the National Institute on Aging and the Social Security Administration.

tia requires calibration in a way that accounts for differential item functioning by individual characteristics, for interview mode, and for several other factors. In prior work, we combined the HRS cognition measures from the core survey with a clinical dementia assessment in the Aging, Demographics, and Memory Study (ADAMS) subsample of the HRS (Heeringa et al., 2009). Using this subsample, we developed an algorithm that maps the cognition measures from the core survey to dementia and CIND status, which we then applied to the entire HRS sample. The algorithm is based on estimating a longitudinal latent variable model of cognition, dementia, and mortality. The details of the approach are discussed in our prior research (see Hudomiet, Hurd, and Rohwedder, 2022), and the properties of the resulting calibrated cognition measures are provided in its accompanying documentation (see Hudomiet, Hurd, and Rohwedder, 2023). In the following paragraph, we provide a brief summary of this algorithm.

The model assumes that individuals are endowed with latent (or unobserved) cognitive ability, c_{it}, where i and t refer to individuals and survey waves, respectively. Cognitive ability is normalized in the following way: When c_{it} is less than 0, the person has dementia; when it is between 0 and 1, the person has CIND; and when it is greater than 1, the person has normal cognitive function. Those with higher cognitive ability perform better on the HRS cognitive functioning tests, but this relationship cannot be estimated on HRS data alone because cognitive ability is not observed. However, it can be estimated from the ADAMS subsample, which has information about dementia, CIND, or normal status and the HRS cognitive function measures. Therefore, by knowing this relationship, we can estimate the probabilities that HRS participants might have dementia as a function of their performance on the HRS cognitive tests and other data. For example, those who perform less well on the HRS tests have a higher probability of having dementia. The model includes many additional features, such as how cognitive ability affects mortality and the propensity to provide self- or proxy-interviews, how cognitive ability changes longitudinally, and how the relationship between cognitive ability and the HRS cognitive functioning tests vary by education, race and ethnicity, and age.

The derived dementia measures are not equivalent to a clinical dementia diagnosis. However, compared with earlier models of dementia that use HRS data, the derived measures have several key advantages. The model is fit to individual longitudinal data and accounts for noise, which can lead to a large rate of reverse transitions (i.e., transition from dementia to not dementia) if not addressed. Cross-sectional cognition models typically predict far more spurious movements (e.g., increases, large decreases) in cognition, likely because of noise that is amplified by wave-to-wave differencing. For example, when we classified individuals' dementia status according to the cross-sectional cutoff method in Crimmins et al. (2011), we found a 16.2-percent rate of reverse transition. Such a classification error will lead to biased estimates of dementia incidence. Our model accounts for observation error, and the rate of reverse transitions is only 0.2 percent.[3] By using the full sequence of an individual's longitudinal data, our model produces smaller standard errors of prevalence than frequently estimated cross-sectional models, which permits the analyses of subpopulations. We directly model dementia prevalence by race and ethnicity to eliminate the bias that other commonly used methods produce along this dimension.

We used this model to predict the following measures:

- **probability of dementia (PrDem)**, which is the probability that an individual in a survey wave has dementia. This variable takes values between 0 and 1 (i.e., between a 0-percent and 100-percent chance of having dementia).

[3] An alternative approach to mitigate the problem of frequent reverse transitions from dementia is to require dementia identification in two subsequent survey waves (Zissimopoulos et al., 2018). However, that method does not solve the problem of classification error.

- **probability of CIND (PrCIND)**, which is defined analogously to PrDem. CIND was the terminology used in ADAMS to mark cognitive impairment and is similar to mild cognitive impairment. This variable also takes values between 0 and 1.
- **expected latent cognition status (ECog)**, which is the expected value of an individual's latent cognition level and is an unbounded continuous variable.

Predictor Variables

Table 2.1 lists the 181 predictor variables we used by type and source of information. The first set of measures includes demographic predictors, such as sex, age, race and ethnicity, current residence (by census division and urbanicity), birth residence (by census division), and veteran and marital status.

The second set includes socioeconomic and labor-market measures, such as education, household income, wealth, earnings, labor-market status, occupation of the longest-held job, basic job characteristics (whether the job required physical effort, lifting, kneeling, or good eyesight and whether the job involved stress), the

TABLE 2.1

Measures Used to Predict Cognitive Impairment and Dementia

Predictor Type	Measures	Source	Number of Individual Predictors
Demographic	Sex, age, race and ethnicity, current residence and urbanicity, birth residence, veteran status, and marital status	Core HRS	8
SES and labor	Education, income, wealth, earnings, labor-market status, occupations, job characteristics, total years worked, private or public health insurance plans, life insurance, and long-term care insurance	Core HRS	34
Lifestyle and health behaviors	Smoking, alcohol use, exercising, cholesterol, flu shots, breast check, mammogram, pap smear, and prostate exam	Core HRS	14
Parental health	Parents alive, age at death, activities of daily living (ADLs), illnesses, and nursing home status	Core HRS	16
Family size and marital history	Household size; number of children, brothers, sisters, and marriages; and length of current or longest marriage	Core HRS	11
Psychosocial	Personality traits, positive or negative affect, life satisfaction, ongoing stressors, lifetime traumas, loneliness, social support, and activities (e.g., social, physical, cognitive)	HRS psychosocial supplement	32
Self-reported health and functional limitations	General health, chronic conditions, ADLs, instrumental activities of daily living (IADLs), depression, body mass index (BMI), back problems, chronic pain, restless sleep, survival expectations, and health-care utilization	Core HRS	42
Physical health measures	Pulse, breathing test, grip strength, balance test, walk test, BMI, blood pressure, and polygenic score of Alzheimer's disease	Enhanced face-to-face interviews	9
Cognitive abilities	Self-rated memory, word recall, counting, date and object naming, naming presidents, and vocabulary tests	Core HRS	15

SOURCE: The information in this table comes from the authors' review and selection of candidate measures in the HRS waves from 1992 to 2016.

total number of years worked, and whether the individual has public or private health insurance coverage, life insurance, or long-term care insurance.

The third set includes measures of lifestyle and health behaviors, such as currently or ever smoking; exercising; checking cholesterol levels; taking flu shots; and having breast checks, mammograms, and pap smears among women or prostate exams among men.

The fourth set includes information about the health status of individuals' parents, such as vital status, their age at death, whether they have or had a limitation in one or more activities of daily living (ADLs) or other major illnesses, and whether they resided in a nursing home.

The fifth category covers family composition and marital status history, including information about the number of children, brothers, and sisters; household size; prior marriages; and the length of the individual's current and longest marriage.

All information discussed so far was available in the core HRS data. Most measures were available in the total sample, but several were not available in some survey waves or some subpopulations because of logical skips, such as occupations being asked of only those who ever worked.

The sixth category covers psychosocial information from the HRS psychosocial supplement, which has been available every other wave (i.e., once every four years) since 2006. We used information on the Big Five personality traits (neuroticism, conscientiousness, extraversion, agreeableness, and openness to experience), positive or negative affect, life satisfaction, ongoing stressors, lifetime traumas, loneliness, social support, and various activities (e.g., social, physical, cognitive). More information about these measures is provided in Smith et al. (2023).

We also used a detailed set of self-reported health and functional limitation measures that include general health (whether excellent, very good, good, fair, or poor); nine ever-had chronic conditions (such as hypertension, diabetes, cancer); ADLs and limitations in the instrumental activities of daily living (IADLs), including both the total number of limitations and individual IADLs; the Center for Epidemiologic Studies Depression Scale (CES-D); body mass index (BMI); back problems; chronic pain; restless sleep; survival expectations (the subjective probability of living to age 75 and other target ages); and health-care utilization (e.g., doctor visits, hospital nights, dental visits).

Additionally, we included objectively measured physical health measures that have been collected in enhanced face-to-face interviews every other wave (i.e., once every four years) since 2006, including pulse rate, a breathing test of lung capacity, a grip strength test, balance tests, a walking time test, BMI, blood pressure, and a derived polygenic score of Alzheimer's disease. More information about these measures can be found in Crimmins et al. (2008) and Ware et al. (2021).

Furthermore, we used cognition measures from the core HRS data on self-rated memory (from excellent to poor), immediate and delayed word recall tests, backward counting and serial sevens subtraction tests, date and object naming tests, naming the president and vice president tests, and a vocabulary test (Fisher et al., 2015).

Sample and Descriptive Statistics

Table 2.2 shows descriptive statistics for the three samples we used in the prediction models. The first one, the two-year incidence, corresponds to person-year observations observed in two consecutive HRS waves (about two years apart) with the person not having dementia in the first wave (i.e., the baseline wave), which is measured by whether their ECog value was greater than 0. There are 69,494 person-year observations in this sample. The four-year incidence sample is defined similarly but measures the incidence of dementia between the baseline wave and about four years (or two HRS waves) later. There are 54,816 person-year observations in that sample. Both samples were used to model the incidence of dementia. For the models of the incidence

TABLE 2.2
Descriptive Statistics About the Three Samples

	Two-Year Incidence			Four-Year Incidence			Long-Term Prediction		
Measures	*N*	Mean	SD	*N*	Mean	SD	*N*	Mean	SD
Age									
Baseline	69,496	74.10	6.70	54,816	73.60	6.50	5,873	60.70	1.00
Outcome							5,873	78.40	2.10
Sex									
Male	69,496	0.42	0.49	54,816	0.42	0.49	5,873	0.45	0.50
Female	69,496	0.58	0.49	54,816	0.58	0.49	5,873	0.55	0.50
Race									
Non-Hispanic White	69,494	0.77	0.42	54,815	0.78	0.42	5,873	0.74	0.44
Non-Hispanic Black	69,494	0.13	0.33	54,815	0.13	0.33	5,873	0.15	0.36
Non-Hispanic other race	69,494	0.02	0.14	54,815	0.02	0.13	5,873	0.02	0.14
Hispanic	69,494	0.08	0.27	54,815	0.08	0.27	5,873	0.09	0.29
Marital status									
Married or partnered	69,457	0.62	0.49	54,789	0.63	0.48	5,873	0.78	0.42
Divorced or separated	69,457	0.09	0.29	54,789	0.09	0.29	5,873	0.11	0.31
Widowed	69,457	0.27	0.44	54,789	0.26	0.44	5,873	0.08	0.28
Never married	69,457	0.03	0.16	54,789	0.02	0.16	5,873	0.03	0.17
Birthplace									
New England	69,275	0.05	0.21	54,653	0.05	0.21	5,856	0.04	0.20
Mid-Atlantic	69,275	0.15	0.35	54,653	0.14	0.35	5,856	0.13	0.34
East North Central	69,275	0.18	0.38	54,653	0.18	0.38	5,856	0.16	0.36
West North Central	69,275	0.11	0.32	54,653	0.12	0.32	5,856	0.11	0.31
South Atlantic	69,275	0.15	0.35	54,653	0.14	0.35	5,856	0.16	0.37
East South Central	69,275	0.08	0.28	54,653	0.08	0.28	5,856	0.10	0.30
West South Central	69,275	0.10	0.30	54,653	0.10	0.30	5,856	0.10	0.30
Mountain	69,275	0.03	0.18	54,653	0.03	0.18	5,856	0.04	0.19
Pacific	69,275	0.05	0.22	54,653	0.05	0.22	5,856	0.05	0.21
Not in the United States	69,275	0.10	0.30	54,653	0.10	0.30	5,856	0.11	0.31
Education									
Less than high school	69,484	0.24	0.42	54,807	0.24	0.43	5,872	0.24	0.43
GED	69,484	0.04	0.21	54,807	0.04	0.21	5,872	0.05	0.21
High school	69,484	0.33	0.47	54,807	0.33	0.47	5,872	0.33	0.47
Some college	69,484	0.20	0.40	54,807	0.20	0.40	5,872	0.19	0.39
College or more	69,484	0.19	0.39	54,807	0.19	0.39	5,872	0.18	0.39

Table 2.2—Continued

	Two-Year Incidence			Four-Year Incidence			Long-Term Prediction		
Measures	*N*	Mean	SD	*N*	Mean	SD	*N*	Mean	SD
Health									
Excellent	69,495	0.09	0.29	54,816	0.10	0.30	5,873	0.17	0.37
Very good	69,495	0.30	0.46	54,816	0.31	0.46	5,873	0.30	0.46
Good	69,495	0.34	0.47	54,816	0.34	0.47	5,873	0.32	0.47
Fair	69,495	0.20	0.40	54,816	0.19	0.40	5,873	0.15	0.36
Poor	69,495	0.06	0.24	54,816	0.05	0.23	5,873	0.06	0.23
Don't know or refused to answer	69,495	0.00	0.03	54,816	0.00	0.02	5,873	0.00	0.02
Dementia									
Incidence (two-year)	69,496	0.04	0.20	53,924	0.03	0.17			
Incidence (four-year)	53,924	0.07	0.26	54,816	0.07	0.26			
Prevalence							5,873	0.13	0.29
CIND									
Incidence (two-year)	54,944	0.10	0.30	44,265	0.08	0.28			
Incidence (four-year)	44,893	0.18	0.39	44,893	0.18	0.39			
Prevalence							5,873	0.30	0.32

SOURCE: The data in this table come from the authors' calculations using information from the HRS waves from 1992 to 2016.

NOTE: GED = General Educational Development. All three samples have a longitudinal design, in which predictor variables are measured in the baseline wave and restricted to individuals not having dementia, and the dementia outcomes are measured in a later wave—one wave (or two years) later for two-year incidence; two waves (or four years later) for four-year incidence, and about ten waves (or about 20 years) later for the long-term predictions. The long-term predictions sample includes individuals observed at least once in the baseline window (ages 55–64) and in the outcome window (ages 75–84).

of CIND, the samples were further restricted to observations with ECog in the normal cognition range of greater than 1 in the baseline wave. Table 2.2 shows the mean and standard deviation (SD) of the predictor variables in the baseline wave and their dementia outcomes either two, four, or twenty years after age 60.

Table A.1 in the annex shows similar statistics on other predictor variables in the three samples.

Our long-term prediction models predict dementia and CIND prevalence at approximately age 80 using predictor variables measured around age 60, which we operationalized as follows. The sample consists of individuals observed at least once in the 75–84 age range (outcome window) and at least once in the 55–64 age range (baseline window). There are 5,873 individuals in this sample. Their outcomes were taken from the survey wave in which they were closest to age 80, and their predictor variables came from waves in which they had a valid nonmissing value and their age was closest to 60. The latter step was carried out separately for each predictor variable because the variables might have different missing value patterns, so the valid values might come from different waves in the baseline window. Table 2.2 shows the mean values of age and dementia prevalence in the outcome window and the rest of the variables from the baseline window.

The average age in the incidence samples is about 74 years. In the long-term prediction sample, the average baseline age is 61 years, and the average age in the outcome window is 78 years. The latter is less than 80 years because of mortality. A little more than one-half of the sample are women because they live longer than men on average and are more likely to be observed at advanced ages. Most individuals are either married or widowed, but about 10 percent are divorced, and 3 percent were never married. The proportion of widows is substantially higher in the incidence samples because of age; the long-term prediction sample measures marital status in the baseline window (ages 55–64), and the incidence sample measures it later, at an average age of 74.

All three samples have good geographic representation; about 10 percent of the persons were born abroad, and between 3 percent and 18 percent were born in one of the nine U.S. census divisions. Approximately 24 percent of the sample have less than a high school education, 4–5 percent have a GED, about 33 percent have a high school education, about 20 percent have dropped out of college, and about 19 percent have at least a bachelor's degree. All samples are heterogeneous in health, and the incidence samples have worse health on average because of the older age of the individuals.

The two-year incidence of dementia is 3–4 percent, the four-year incidence is 7 percent, and 13 percent of the long-term prediction sample have dementia in the outcome window. The two-year incidence of CIND (i.e., the percentage of individuals whose cognitive status is CIND or dementia two years after they were observed in cognitively normal status) is 8–10 percent, the four-year incidence is 18 percent, and 30 percent of the long-prediction sample have CIND status in the outcome window.

CHAPTER 3

Prediction Models

Basic Models

The basic dementia incidence models take the following form:

$$\Pr\left(ECog_{i,w+d} < 0 \middle| ECog_{i,w} \geq 0\right) = \beta_0 + \beta_1 a_{i,w} + \beta_2 a_{i,w}^2 + \beta_3 a_{i,w}^3 + y x_{i,w}. \tag{3.1}$$

We model the probability that an individual's cognition falls from the non-dementia range in wave w to the dementia range in wave $w + d$ as a cubic function of age ($a_{i,w}$) and a predictor variable ($x_{i,w}$) measured in wave w. Categorical predictor variables are included as a set of dummies, and continuous variables are added either linearly or as a cubic function if we found evidence for a nonlinear relationship.

We ran separate regression models for each predictor variable and then ranked them by the partial R-squared value on $x_{i,w}$, which is the fraction of the variance of the outcome variable explained by the predictor variable. The partial R-squared takes values between 0 and 1, and higher values correspond to more important predictors. In general, the partial R-squared value is higher if the predictor variable has a stronger relationship with the outcome and the variance of the predictor variable is higher. For example, a factor that very strongly predicts dementia but is very rare in the population might not have a high partial R-squared value.

The basic long-term prediction model takes the following form:

$$\Pr Dem_{i,80} = \beta_0 + \beta_1 a_{i,80} + y x_{i,60}. \tag{3.2}$$

Thus, this model only includes an individual's age in the outcome window and the predictor variable measured in the baseline wave. Again, categorical predictor variables are included as a set of dummies, and continuous variables are added either linearly or as a cubic function. We also estimated separate models for each predictor variable and ranked them by the partial R-squared value.

Multivariate Models

The multivariate prediction models are similar to the basic models in Equations 3.1 and 3.2 but include more than one predictor variable. We estimated models with different sets of predictors and prioritized those identified as strong by the basic models, those that are available in larger samples, and those that have fewer missing values. The precise specifications are discussed in Chapter 4.

Missing Values

Most variables have relatively few missing values in the HRS. Values might be missing because the HRS did not ask those questions of either some individuals (skip patterns) or not at all in a particular wave, or because a person did not provide a valid answer, such as "don't know or refused to answer" (DK/RF). We model active missing values, in which the question was asked but not answered, as a separate response category. For example, we used the following categories for the self-reported health question:

- excellent
- very good
- good
- fair
- poor
- DK/RF.

Active missing values in continuous variables were imputed as the sample mean of that variable, and a flag variable of the missing values was added to the regression models. When a question was not asked of the person, the values remained missing.

We used cleaned and preprocessed data from the RAND HRS Longitudinal File. Missing values in some variables were already replaced by imputations, such as in the cognitive measures and income and wealth. These imputed values were treated as valid answers in our prediction models.

CHAPTER 4

Results

Basic Model Results

Table 4.1 shows the strongest predictors identified by the basic models using the following procedure:

1. We estimated the partial R-squared values of each predictor variable.
2. We retained those with at least a value of 0.005 (or one-half of 1 percent) in the incidence models or a value of 0.01 (1 percent) in the long-term prediction models. We used a higher threshold in the latter because the partial R-squared values tended to be higher in those models.
3. If more than five predictor variables passed these thresholds in a predictor category, we selected the top five with the highest partial R-squared values.
4. We implemented this logic in all three models (two-year incidence, four-year incidence, and long-term prediction) and retained the predictors in the top five in at least one of the models.

Among the predictors in Table 4.1, individuals' cognitive ability and health-related predictors had the highest partial R-squared values, whereas parental health, family size, marital history, and demographics had the lowest values.

TABLE 4.1
Strongest Predictors of Dementia in Each Category in the Basic Models

	Partial R-Squared Value		
Category and Item	Two-Year Incidence	Four-Year Incidence	Long-Term Prediction
Demographic			
Birth census division	0.003	0.006	0.018
Race	0.002	0.003	0.011
SES and labor			
Years of education	0.005	0.010	0.033
Total number of years worked	0.003	0.006	0.016
Labor-market status	0.002	0.004	0.025
Occupation of longest-held job	0.001	0.003	0.012
Has private health insurance	0.002	0.003	0.023
Lifestyle and health behaviors for men			
Light physical activities	0.011	0.012	N/A
Moderate physical activities	0.008	0.011	N/A
Vigorous physical activities	0.003	0.005	N/A

Table 4.1—Continued

Category and Item	Partial R-Squared Value		
	Two-Year Incidence	Four-Year Incidence	Long-Term Prediction
Ever drinking alcohol	0.003	0.006	0.007
Number of days drinking alcohol	0.003	0.005	0.012
Number of drinks per day	0.003	0.005	0.012
Lifestyle and health behaviors for women			
Light physical activities	0.016	0.022	N/A
Moderate physical activities	0.008	0.011	N/A
Vigorous physical activities	0.002	0.004	N/A
Ever drinking alcohol	0.003	0.005	0.005
Number of days drinking alcohol	0.002	0.003	0.011
Number of drinks per day	0.002	0.003	0.012
Parental health	—	—	—
Family size and marital history	—	—	—
Psychosocial			
Hobby activities	0.011	0.008	N/A
Positive affect	0.010	0.008	N/A
Negative affect	0.010	0.011	N/A
Novel information activities	0.009	0.009	N/A
Satisfaction with health	0.009	0.015	N/A
Big Five personality trait: neuroticism	0.006	0.010	N/A
Big Five personality trait: conscientiousness	0.007	0.009	N/A
Self-reported health and functional limitations			
Number of IADLs	0.034	0.039	0.037
Number of ADLs	0.017	0.021	0.027
Self-reported health	0.016	0.026	0.056
Number of nursing home nights	0.009	0.006	0.004
CES-D score	0.007	0.013	0.016
Health limits work	0.005	0.008	0.025
Physical health measures			
Walk test time	0.017	0.026	N/A
Breathing test	0.012	0.013	N/A
Semi-tandem balance test time	0.012	0.014	N/A

Table 4.1—Continued

Category and Item	Partial R-Squared Value		
	Two-Year Incidence	Four-Year Incidence	Long-Term Prediction
Cognitive abilities			
Delayed word recall	0.042	0.057	0.052
Immediate word recall	0.041	0.054	0.049
Serial sevens test	0.020	0.031	0.060
Knows the year	0.015	0.017	0.004
Knows the day of the week	0.013	0.013	0.001
Names the vice president	0.012	0.017	0.019
Vocabulary test	N/A	N/A	0.039
Self-reported memory	0.010	0.016	0.026

SOURCE: The data in this table come from the authors' calculations using information from the HRS waves from 1992 to 2016.

NOTE: N/A refers to measures that are not available in the particular samples, and a dash (—) indicates categories with no strong predictors.

Demographic Predictors

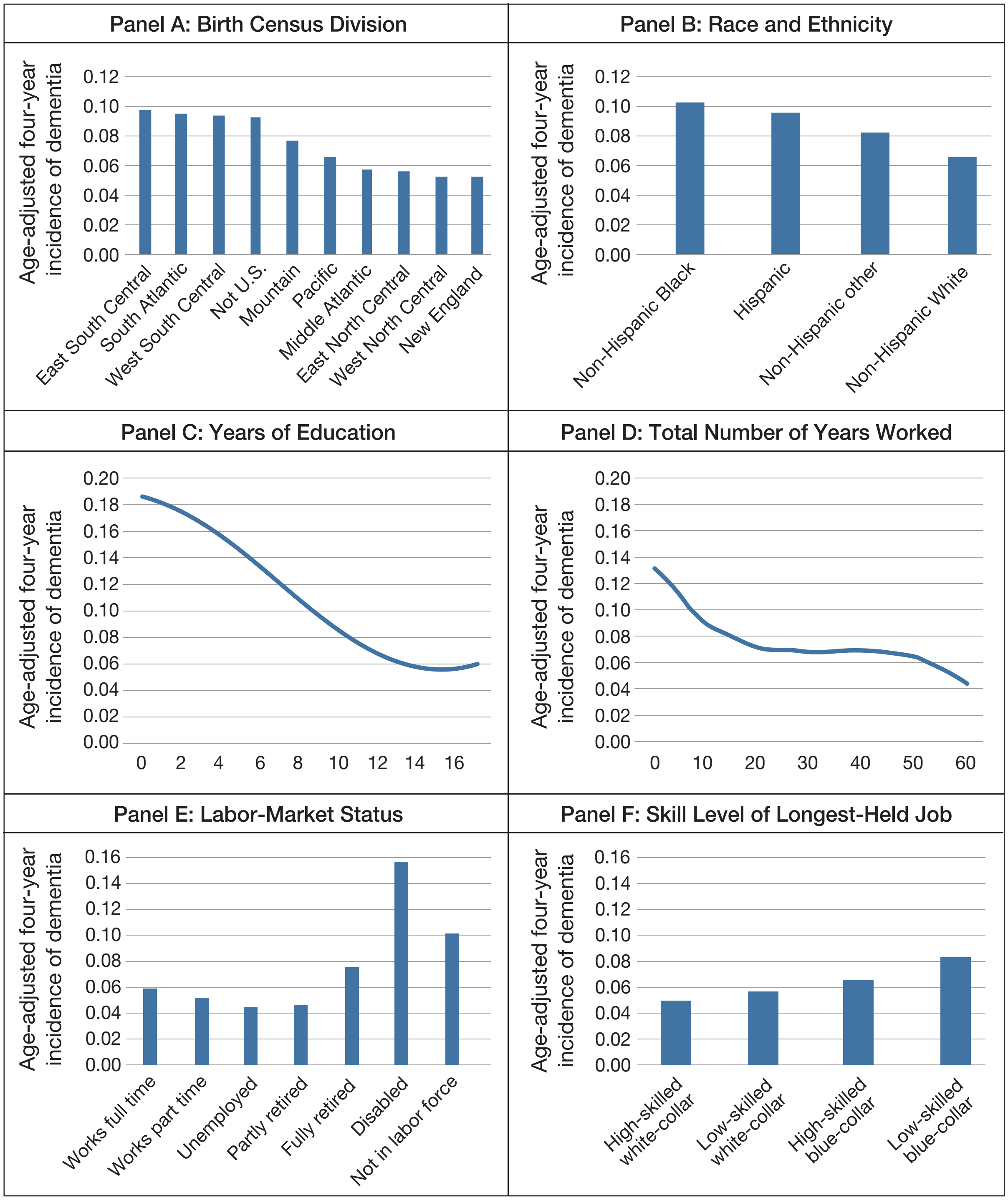
Among the demographic predictors, only birth census division and race and ethnicity passed the importance threshold. Panels A and B in Figure 4.1 show the relationship between these variables and age-adjusted four-year incidence. Age adjustment re-evaluates the model prediction after replacing each sample member's observed age with the sample mean. The corresponding two-year and long-term prediction relationships are provided in the annex. The incidence of dementia is substantially higher among those who were born in the Southern states or abroad than in the rest of the country. The four-year dementia incidence is slightly less than 10 percent among those who were born in one of the three southern census divisions, a little more than 9 percent among those who were born abroad, and only slightly more than 5 percent among those who were born in New England or the Midwest (i.e., the East North Central and West North Central census divisions). The incidence and prevalence of dementia are also substantially higher among non-Hispanic Black and Hispanic individuals than among non-Hispanic White individuals.

Socioeconomic Status and Labor Predictors

We identified five SES and labor-market measures as strong predictors of dementia incidence and prevalence: years of education, total number of years worked, labor-market status, skill level of their longest-held job, and having a private health insurance plan. The relationships between these predictors and four-year incidence are shown in Panels C, D, E, and F of Figure 4.1 and in Panel F of Figure 4.4. Our prior research found a strong education gradient in dementia risk, but the main difference was between those with and without a high school degree, and the gradient above a high school degree was not large (Hudomiet, Hurd, and Rohwedder, 2022). We found the same result in this study, but we also found a very strong and monotonic gradient by education below a high school degree. The four-year incidence is highest among those with no formal education (at 18.7 percent), declines almost linearly to 6.7 percent among those with 12 years of education, but it is practically flat for greater education levels. Of the sample, 74 percent have a high school education or greater, which means that the education gradient is mostly a contrast between those lacking a high school diploma and those who graduated high school. Nevertheless, it is worth emphasizing that among those lacking a high school diploma, incidence decreases strongly with education.

FIGURE 4.1

Age-Adjusted Four-Year Dementia Incidence, by Selected Demographic, Socioeconomic Status, and Labor Predictors



SOURCE: The data in this figure come from the authors' calculations using information from the HRS waves from 1992 to 2016 for individuals ages 65 and older who were dementia-free at wave *t*.

NOTE: Predictions are based on regression models of dementia status at wave *t* + 2 (approximately four years later) as a cubic function of age and the predictor variables. Each panel corresponds to a predictor variable in the baseline wave: census division at birth, race and ethnicity, years of education, total number of years worked, labor-market status, and skill level of the longest-held job.

Total years worked is also a strong predictor of dementia, and the main difference is between those who never worked versus those who worked at least a few years. At the same time, the differential at the top of the distribution is small. The four-year age-adjusted incidence is 13.0 percent among those who never worked, 8.9 percent among those who worked ten years, and 6.9 percent among those who worked 40 years.

Labor-market status is another strong predictor. The four-year incidence is, by far, the largest among those who reported a disability (at 15.7 percent). At the same time, the differences in the other labor-market status groups (i.e., working full time, working part time, unemployed, partly retired, fully retired, or not in the labor force) are small. Some of these relationships might be because of reverse causality: Those with a severe cognitive limitation but not yet dementia might report having a disability in the survey. However, we also found a similarly substantial differential in the long-term prediction model. Those who had a disability at age 60 had a 32-percent chance of having dementia at age 80, whereas dementia prevalence was around 10 percent in the other groups.

Regarding skill level of the longest-held job, the four-year incidence of dementia is highest among those who worked in low-skilled blue-collar jobs and lowest among those in high-skilled white-collar jobs. Another mostly work-related factor is having a private health insurance plan at age 60, which is associated with lower dementia risk; the age-adjusted four-year incidence was 6.1 percent among those with private health insurance (versus 8.7 percent for those without private health insurance). The annex shows a similarly substantial differential in the long-term prediction models. Individuals who had private health insurance coverage at age 60 had a 10.8-percent chance of having dementia 20 years later versus a 22.1-percent chance for those who did not.

Lifestyle and Health Behaviors

Factors in the lifestyle and health behavior category were analyzed separately by sex because many of these factors were only available in one or the other group. However, we found that the sex-dependent factors were not strong predictors of dementia, and the strongest factors were similar by sex. We found that the lack of exercising (as measured by light and moderate physical activity) is the strongest predictor of dementia incidence in both sexes. Figure 4.2 shows large differentials in four-year dementia incidence between those who never exercised and those who exercised at least sometimes, whereas the differential between those who occasionally exercised and those who regularly exercised is negligible.

Alcohol consumption is also predictive of dementia incidence; the incidence and prevalence of dementia are the lowest among those with moderate alcohol consumption, and the risk is elevated among those who never drink or who drink excessively. We found the same patterns in both sex groups and in all three dementia models that we considered, as shown in the annex. Other lifestyle and health behavior factors are not strong predictors of dementia, such as smoking, checking cholesterol levels, having flu shots, and getting mammograms, pap smears, and breast or prostate exams.

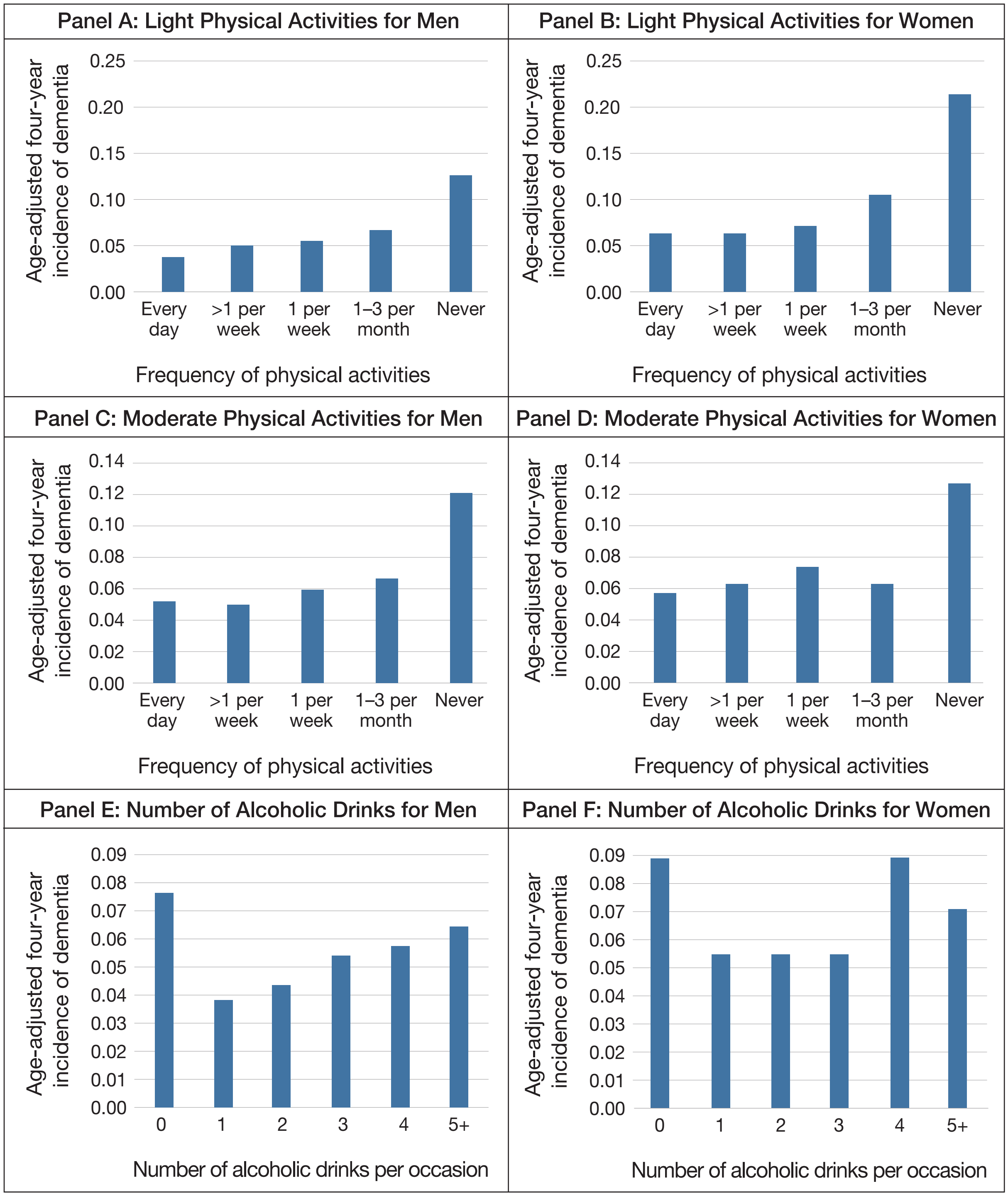
Parental Health, Family Size, and Marital History

None of the parental health measures (i.e., parents' vital status, age at death if applicable, ADLs, severe sickness, dementia, and nursing home status) were found to predict the dementia outcomes strongly. These findings suggest that to characterize an individual's risk of developing dementia, considering the individual's own characteristics is a lot more informative than information on their parents.

We found similarly weak associations between dementia and measures of family size and marital history, such as household size, number of children or siblings, and number of marriages.

FIGURE 4.2

Age-Adjusted Four-Year Dementia Incidence, by Selected Lifestyle and Health Behavior Predictors



SOURCE: The data in this figure come from the authors' calculations using information from the HRS waves from 1992 to 2016 for individuals ages 65 and older who were dementia-free at wave *t*.

NOTE: Predictions are based on regression models of dementia status at wave $t + 2$ (approximately four years later) as a cubic function of age and the predictor variables. Each panel corresponds to a predictor variable in the baseline wave: frequency of light physical activities, frequency of moderate physical activities, and number of alcoholic drinks per occasion.

Psychosocial Predictors

Our analyses identified seven strong psychosocial predictors of dementia incidence. Among the strongest predictors are two activity measures (hobby and novel information activities), positive and negative affect, two of the Big Five personality traits (neuroticism and conscientiousness), and individuals' satisfaction with their health. The hobby activities capture how often individuals engage in (1) doing word games; (2) playing card or board games, such as chess; (3) home and car maintenance or gardening; (4) making clothes, knitting, or embroidering; and (5) engaging in a hobby or a project. The novel information activity measure captures how often individuals (1) participate in education or training, (2) write letters or stories, or (3) use the computer. The long-term prediction models did not analyze these measures because they have only been available since 2006.

Figure 4.3 gives a graphical representation of these findings. Panels E and F of Figure 4.3 show that those who engage more in novel information and hobby activities have substantially lower chances of dementia incidence, and the primary difference is at the bottom of the distribution among those who never or sometimes participate in these activities. We also investigated the individual items but found that the aggregated item predicted dementia more strongly than the individual items, perhaps because doing any of these activities is protective. However, we note that our results might also reflect reverse causality: Individuals who experience cognitive problems might stop doing these activities. We could not run the long-term prediction models on these measures because of data limitations, but it would be worthwhile to investigate these issues in future research.

Figure 4.3 also shows that those with more conscientious and less neurotic personalities, those who are more satisfied with their health, and those who score lower on negative affect all have substantially smaller chances of developing dementia than individuals in the other categories.

Self-Reported Health and Functional Limitations

Several health and functional limitation measures strongly predict dementia incidence and prevalence, as evidenced by their high partial R-squared values (see Table 4.1) and the graphical relationships shown in Figure 4.4. The strongest predictors are self-reported health and the numbers of ADL and IADL limitations: Those with worse self-reported health and more ADL and IADL limitations have a substantially higher chance of developing dementia than other individuals, and the differentials are—once again—largest at the bottom of the distributions. Even though some of these relationships might be a result of reverse causality, we found similarly strong associations in the long-term prediction models, which are less prone to reverse causality. In other words, individuals' health and functional limitations at age 60 are among the strongest predictors of having dementia at age 80.

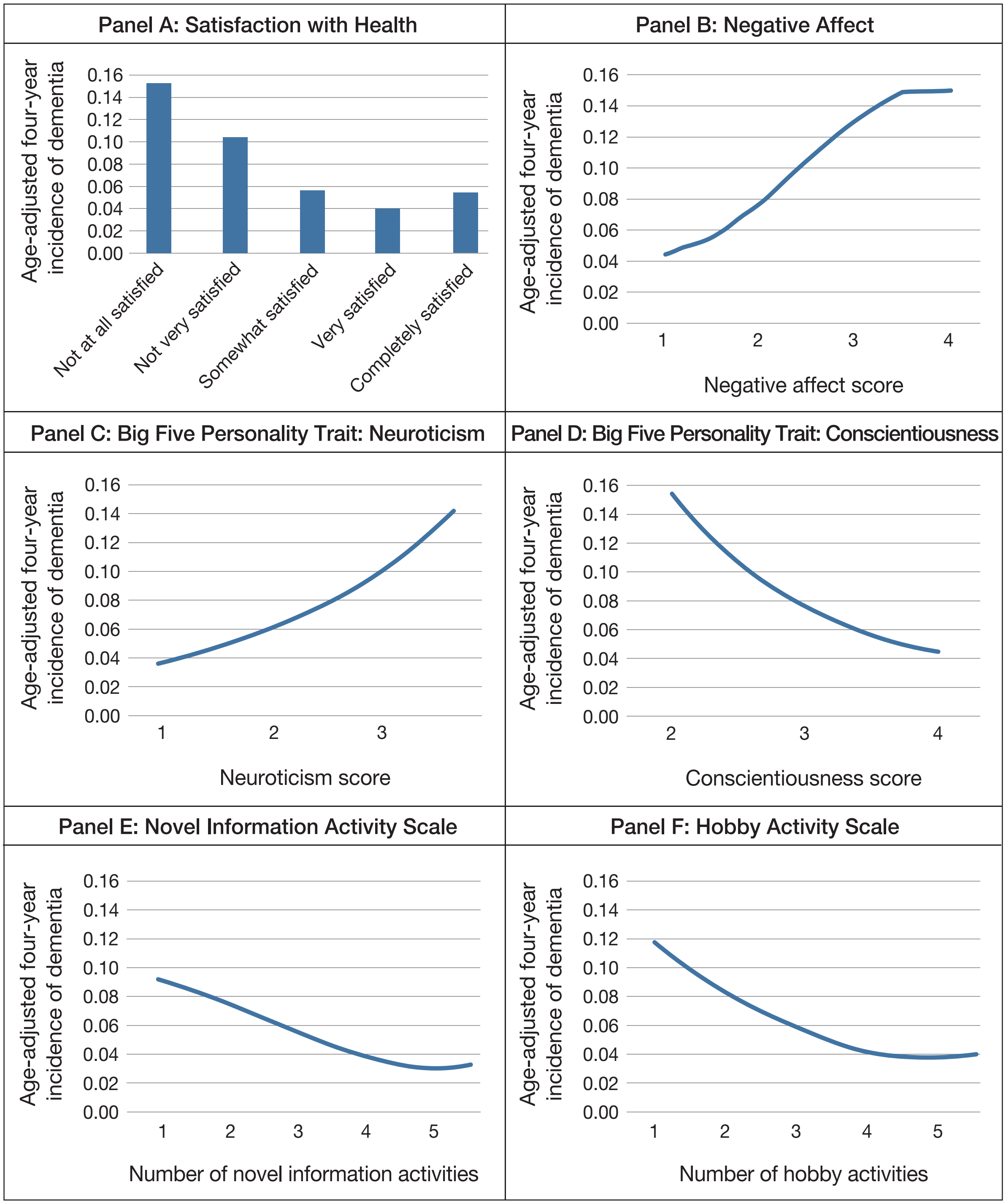
Physical Health Predictors

We analyzed objective physical health measures in the incidence models.[1] Three of them passed our inclusion criteria in the list of strong predictors: the walk time test, the breathing test, and the semi-tandem balance test (see Figure 4.5). The walk time test measures the time that individuals take to walk 100 inches at a normal pace; the breathing test measures the amount of air that individuals can forcefully breathe out of their lungs; and the semi-tandem balance test measures the amount of time that individuals can hold a standing position with one foot in front of the other, up to ten seconds. These measures have been shown to predict and

[1] The physical measures were not collected in the earlier HRS waves, which means that the longitudinal follow-up was too short and prevented us from running the long-term prediction models for these variables.

FIGURE 4.3

Age-Adjusted Four-Year Dementia Incidence, by Selected Psychosocial Predictors

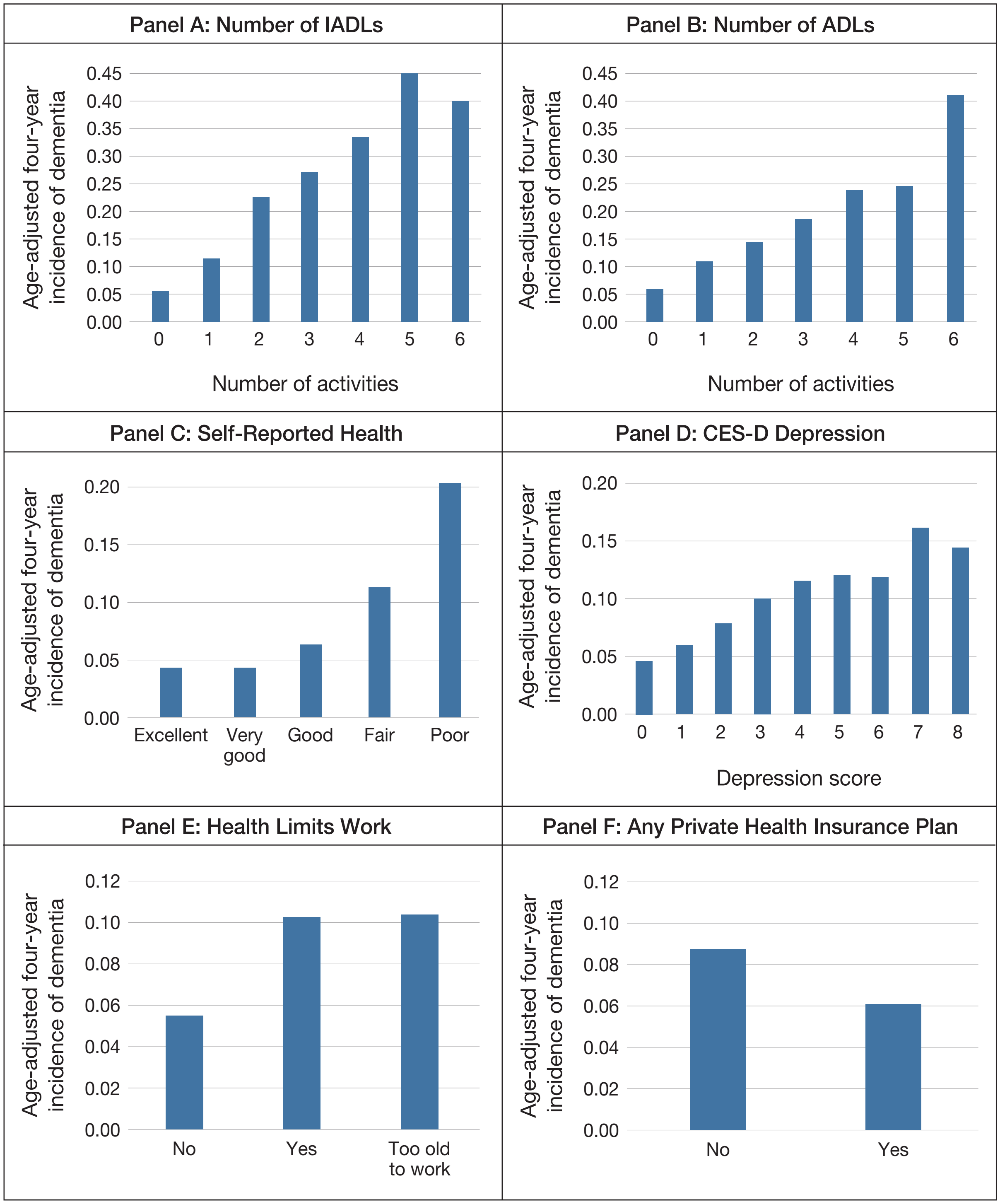


SOURCE: The data in this figure come from the authors' calculations using information from the HRS waves from 1992 to 2016 for individuals ages 65 and older who were dementia-free at wave *t*.

NOTE: Predictions are based on regression models of dementia status at wave *t* + 2 (approximately four years later) as a cubic function of age and the predictors. Each panel corresponds to a predictor variable in the baseline wave: satisfaction with health, scores on affect and personality traits, and number of activities.

FIGURE 4.4

Age-Adjusted Four-Year Dementia Incidence, by Selected Self-Reported Health and Functional Limitations Predictors

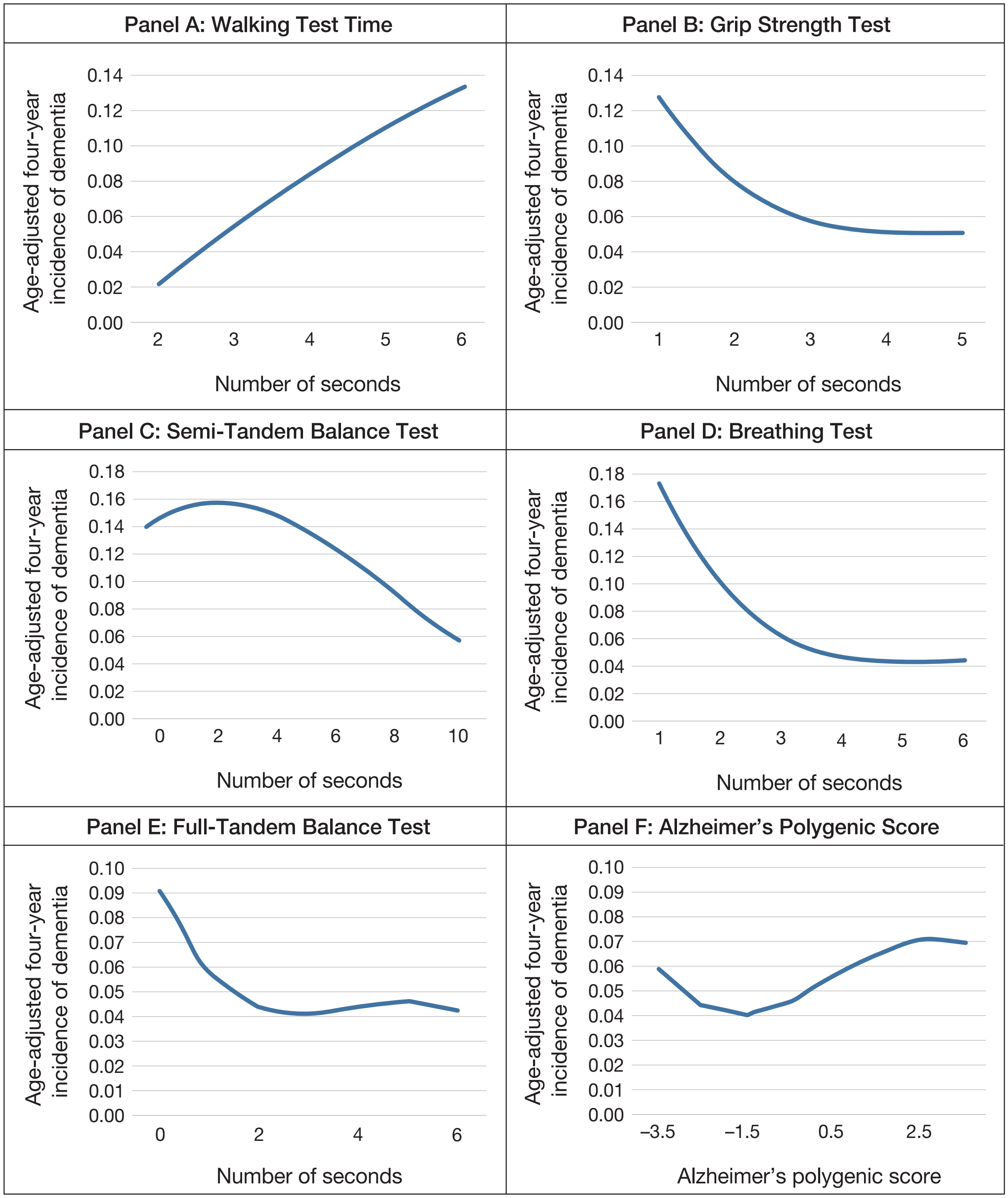


SOURCE: The data in this figure come from the authors' calculations using information from the HRS waves from 1992 to 2016 for individuals ages 65 and older who were dementia-free at wave *t*.

NOTE: Predictions are based on regression models of dementia status at wave *t* + 2 (approximately four years later) as a cubic function of age and the predictors. Each panel corresponds to a predictor variable in the baseline wave: number of IADLs, number of ADLs, self-reported health, CES-D score, whether health limits work, and any private health insurance plan.

FIGURE 4.5

Age-Adjusted Four-Year Dementia Incidence, by Selected Physical Health Predictors



SOURCE: Features data from the authors' calculations using information from the HRS waves from 1992 to 2016 for individuals ages 65 and older who were dementia-free at wave *t*.

NOTE: Predictions are based on regression models of dementia status at wave *t* + 2 (approximately four years later) as a cubic function of age and the predictors. Each panel corresponds to a predictor variable in the baseline wave: walking test time, grip strength test, breathing test, semi- and full-tandem balance tests, and the polygenic score for Alzheimer's disease.

mark underlying health problems (Crimmins et al., 2008). We found a very strong gradient in dementia risk by these measures. Those who could walk 100 inches in two seconds have a 2.1-percent chance of dementia incidence in four years, whereas this likelihood is 13.3 percent among those who took six seconds to walk as far. Individuals who could breathe out six liters of air from their lungs have a 4.3-percent chance of dementia incidence compared with a 17.1-percent chance among those who could breathe out only one liter. The semi-tandem balance test is also a strong predictor, especially at the bottom of the distribution. Figure 4.5 includes three additional measures that barely missed the inclusion criteria for being considered strong predictors: the full-tandem balance test, the grip strength test, and a polygenic Alzheimer's score. These measures also show steep—though less strong—gradients in four-year dementia incidence.

Cognitive Abilities

The cognitive ability measures, which were assessed at the respective baseline of each model, are the strongest predictors of subsequent dementia in all model specifications (see annex for supplemental figures and tables). Our models identified eight strong cognitive predictors; the immediate word recall test, the delayed word recall test, and the serial sevens subtraction tests are the strongest among them. The immediate word recall test counts how many words individuals remember after hearing a list of ten words. The delayed word recall test asks individuals to repeat the exact words a few minutes later. The serial sevens test instructs individuals to subtract 7 from 100 five times in a row.

Figure 4.6 shows a steep gradient in four-year dementia incidence by these measures, and there are larger differentials at the bottom of the distribution once again. Importantly, we found strong relationships in all three models, including the long-term prediction models. In other words, lower cognitive capacity at age 60 is a powerful predictor of individuals' dementia status 20 years later.

Multivariate Model Results

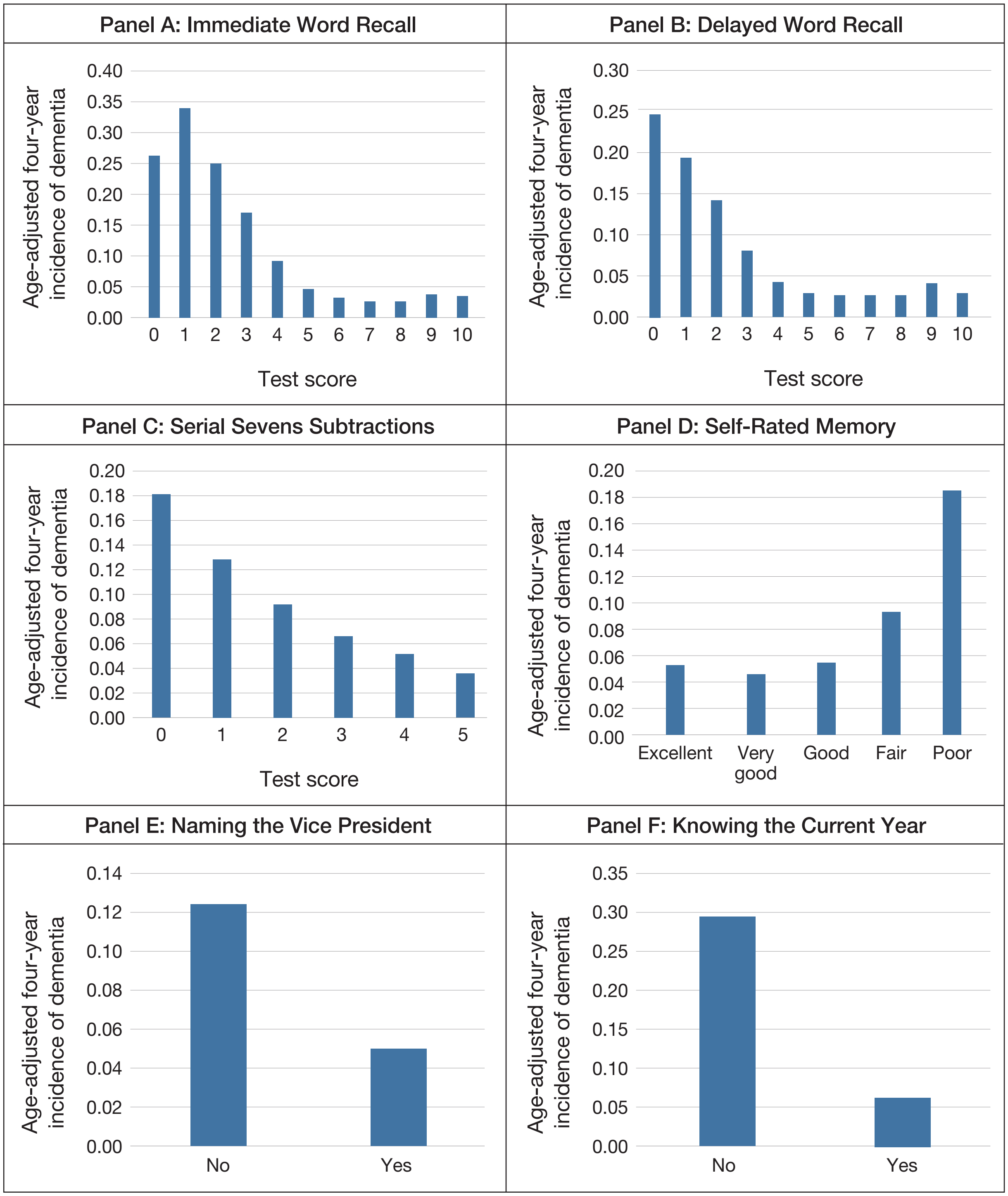
The multivariate models test whether the predictor variables remain strongly associated with the outcomes when accounting for other factors. We estimated 16 sets of regression models that vary in the outcomes (dementia or CIND), the modeling framework (two-year incidence, four-year incidence, and long-term prediction), the availability of the predictors (predictors with few missing values versus those with many missing values), and whether the model was stratified by sex or estimated for the full sample. Each set of regressions included different sets of controls added iteratively. The detailed estimation results are provided in the annex.

Tables 4.2 and 4.3 summarize the most important insights from these models. Table 4.2 focuses on the dementia models, and Table 4.3 on the CIND models. We categorized the predictor variables in the following way:

- A *consistent risk factor*, which is indicated by two hyphens (--), is a predictor that increases the chances of dementia in a statistically significant way, in both the narrowest and the broadest models that we considered.
- An *explainable risk factor*, which is indicated by one hyphen (-), statistically significantly increases the chances of dementia in the narrowest but not the broadest model, suggesting that the addition of other control variables account for some or most of the variation captured by the predictor variable of interest in the narrowest regression models.
- *Consistent* and *explainable protective factors* are defined analogously to the risk factors and are indicated by two plus signs (++) and one plus sign (+), respectively.
- *Nonpredictors*, which are indicated by 0, were not statistically significantly related to the outcome in the narrowest model.

FIGURE 4.6

Age-Adjusted Four-Year Dementia Incidence, by Selected Cognitive Predictors



SOURCE: Features data from the authors' calculations using information from the HRS waves from 1992 to 2016 for individuals ages 65 and older who were dementia-free at wave *t*.

NOTE: Predictions are based on regression models of dementia status at wave $t + 2$ (approximately four years later) as a cubic function of age and the predictor variables. Each panel corresponds to a predictor variable in the baseline wave: number of correct words in immediate and delayed word recall tests, number of correct responses in serial sevens subtractions, self-rated memory, naming the vice president, and knowing the current year.

TABLE 4.2
Significant Predictors in the Multivariate Models of Dementia

	Risk or Protective Factor		
Category and Item	**Two-Year Incidence**	**Four-Year Incidence**	**Long-Term Prediction**
Demographic			
Age	--	--	--
Non-Hispanic Black (vs. Non-Hispanic White)	-	-	-
Hispanic	-	-	-
Birthplace in South Atlantic (vs. New England)	--	-	--
Birthplace in East South Central	--	-	--
Birthplace in West South Central	--	-	--
SES and labor			
Years of education	+	+	++
Total years worked	+	+	++
Low-skilled white-collar job (vs. high-skilled job)	0	0	++
Has private health insurance	+	+	++
On Medicaid	-	-	--
Lifestyle and health behaviors			
Never doing moderate physical activities (vs. daily)	--	--	N/A
Never doing light physical activities (vs. daily)	--	--	N/A
1 alcoholic drink per occasion (vs. 0)	++	++	0
2 alcoholic drinks per occasion	+	+	++
4 alcoholic drinks per occasion	0	+	0
5 or more alcoholic drinks per occasion	0	+	0
BMI 25–30 (vs. 0–25)	++	++	0
BMI 30–35	++	++	0
BMI >35	++	++	--
Self-reported health			
Self-reported health good (vs. excellent)	0	--	--
Self-reported health fair	--	--	--
Self-reported health poor	--	--	--
Ever had diabetes	0	0	--
Ever had a stroke	--	--	--
Ever had psychiatric problems	--	--	0
Has work-limiting health problems	++	0	0
Nursing home nights	-	-	--
Visited dentist	++	++	N/A

Table 4.2—Continued

Category and Item	Risk or Protective Factor		
	Two-Year Incidence	Four-Year Incidence	Long-Term Prediction
Experiences mild pain (vs. no pain)	+	0	0
Received home care	-	0	0
CES-D score	--	--	0
Had restless sleep	--	--	0
Functional limitations			
Has difficulty walking across room	0	--	0
Has difficulty dressing	++	0	0
Has difficulty bathing	--	--	0
Has difficulty using the map	--	--	0
Has difficulty using the phone	--	--	--
Has difficulty with money	--	--	0
Has difficulty with medications	--	--	0
Has difficulty shopping	--	--	N/A
Has difficulty preparing meals	--	--	N/A
Cognitive abilities			
Self-rated memory very good (vs. excellent)	--	0	0
Self-rated memory good	--	0	0
Self-rated memory fair	--	--	0
Self-rated memory poor	--	--	--
Immediate word recall	++	++	0
Delayed word recall	++	++	++
Serial sevens	++	++	++
Knows the year	++	++	N/A
Names the president	++	++	N/A
Names the vice president	++	++	N/A
Psychosocial			
Hobby activities	++	0	N/A
Novel information activities	++	0	N/A
Completely satisfied with health (vs. not at all)	--	0	N/A
Big Five personality trait: neuroticism	0	--	N/A
Big Five personality trait: conscientiousness	0	++	N/A

Table 4.2—Continued

Category and Item	Risk or Protective Factor		
	Two-Year Incidence	Four-Year Incidence	Long-Term Prediction
Physical health measures and genes			
Breathing test	0	++	N/A
Semi-tandem balance test time	--	--	N/A
Grip strength test	++	++	N/A
Alzheimer's polygenic score	--	--	--

SOURCE: The data in this table come from the authors' calculations using information from the HRS waves from 1992 to 2016.

NOTE: Two hyphens (--) indicate a consistent risk factor that statistically significantly increases the chances of dementia in the narrowest and the broadest models that we considered. One hyphen (-) indicates an explainable risk factor, which statistically significantly increases the chances of dementia in the narrowest but not the broadest model. Two plus signs (++) and one plus sign (+) indicate consistent and explainable protective factors respectively, which are defined analogously to the risk factors. 0 indicates nonpredictors that are not statistically significantly related to the outcome in the narrowest models. The table shows only predictors that were risk or protective factors in at least one model.

TABLE 4.3

Significant Predictors in the Multivariate Models of Cognitive Impairment

Category and Item	Risk or Protective Factor		
	Two-Year Incidence	Four-Year Incidence	Long-Term
Demographic			
Age	--	--	--
Non-Hispanic Black (vs. non-Hispanic White)	-	-	-
Hispanic	0	0	-
Separated or divorced (vs. married)	-	-	-
Widowed	0	0	-
Birthplace in South Atlantic (vs. New England)	0	0	-
Birthplace in East South Central	0	0	-
SES and labor			
Years of education	+	+	++
Total years worked	++	++	++
High-skilled blue-collar job (vs. high-skilled white-collar job)	--	--	--
Low-skilled blue-collar job	--	--	--
Never worked	--	--	-
Has private health insurance	++	++	+
On Medicaid	-	-	-
Lifestyle and health behaviors			
Never doing moderate physical activities (vs. daily)	--	--	N/A
Never doing light physical activities (vs. daily)	--	--	N/A
1–3 times of light physical activities per month	--	--	N/A
1 alcoholic drink per occasion (vs. 0)	+	+	++

Table 4.3—Continued

Category and Item	Risk or Protective Factor: Two-Year Incidence	Four-Year Incidence	Long-Term
2 alcoholic drinks per occasion	+	0	0
BMI 25–30 (vs. 0–25)	++	++	0
BMI 30-35	++	++	-
BMI >35	0	0	--
Self-reported health			
Self-reported health very good (vs. excellent)	--	--	0
Self-reported health good	--	--	--
Self-reported health fair	--	--	--
Self-reported health poor	--	--	--
Ever had diabetes	--	--	--
Ever had lung disease	--	0	0
Ever had a stroke	0	0	--
Ever had psychiatric problems	--	--	--
Nursing home nights	-	-	0
Visited dentist	++	0	N/A
CES-D score	--	--	-
Had restless sleep	--	--	0
Functional limitations			
Has difficulty dressing	0	0	++
Has difficulty using the map	--	--	--
Has difficulty using the phone	--	--	0
Has difficulty with money	--	--	--
Has difficulty preparing meals	--	--	N/A
Cognitive abilities			
Self-rated memory good (vs. excellent)	--	--	0
Self-rated memory fair	--	--	--
Self-rated memory poor	--	--	--
Immediate word recall	++	++	++
Delayed word recall	++	++	++
Serial sevens	++	++	++
Knows the year	++	++	N/A
Names the president	0	++	N/A
Names the vice president	++	++	N/A

Table 4.3—Continued

Category and Item	Risk or Protective Factor		
	Two-Year Incidence	Four-Year Incidence	Long-Term
Psychosocial			
Hobby activities	++	++	N/A
Novel information activities	0	++	N/A
Somewhat satisfied with health (vs. not at all)	++	0	N/A
Very satisfied with health	++	0	N/A
Completely satisfied with health	++	0	N/A
Big Five personality trait: conscientiousness	0	++	N/A
Physical health measures and genes			
Breathing test	++	++	N/A
Timed walk test time	++	0	N/A
Semi-tandem balance test time	0	--	N/A
Grip strength test	++	++	N/A
Alzheimer's polygenic score	--	--	--

SOURCE: The data in this table come from the authors' calculations using information from the HRS waves from 1992 to 2016.

NOTE: Two hyphens (--) indicate a consistent risk factor that statistically significantly increases the chances of dementia in the narrowest and the broadest models that we considered. One hyphen (-) indicates an explainable risk factor, which statistically significantly increases the chances of dementia in the narrowest but not the broadest model. Two plus signs (++) and one plus sign (+(indicate consistent and explainable protective factors respectively, which are defined analogously to the risk factors. 0 indicates nonpredictors that are not statistically significantly related to the outcome in the narrowest models. The table shows only predictors that were risk or protective factors in at least one model.

We note that the standard errors and *p*-values in these regressions are not adjusted for the multiple testing problem (Ioannidis, 2005), and some of the statistically significant coefficients—especially the weakest ones—might be false discoveries. Our primary goal here was to use a simple and consistent algorithm to categorize the strength of the predictor variables.

Tables 4.2 and 4.3 list all the predictor variables that were consistent or explainable in at least one of the three model types (two-year incidence, four-year incidence, and long-term prediction).

We first focused on the predictor variables that had relatively few missing values and iteratively added them to the models: first, the demographic predictors, followed by SES and labor, lifestyle and health behaviors, self-reported health, functional limitations, and cognitive abilities. These models were restricted to a consistent sample with no missing values in any predictor variables. The *narrow* models correspond to the models in which the predictor variables were first added, and the *broad* models correspond to the models with all predictor variables except the cognitive ability measures. We excluded the cognitive measures because they strongly predicted the outcome, and we saw evidence that these models overcontrolled for cognitive function and even reversed the relationship between some important predictor variables.

The remaining predictor variables (psychosocial measures and physical health measures and genes) were analyzed separately because they had many missing values. We estimated models with and without these measures using all predictors except the cognition variables.

Tables 4.2 and 4.3 show that older individuals are at greater risk of CIND and dementia even when accounting for a large number of other predictors. Individuals from racial and ethnic minority groups face an elevated chance of dementia in all three models, but the differentials were no longer statistically significant after accounting for the other predictors, such as SES. The patterns were similar for predicting CIND, but the Hispanic versus non-Hispanic White differential was not statistically significant in the two incidence models. Being born in the South is a consistent risk factor in two out of three models (two-year incidence and long-term prediction) and is an explainable risk factor of four-year dementia incidence. The geographic differences in CIND are less pronounced.

Several SES and labor-market variables are consistent protective factors in the long-term prediction models and explainable protective factors in the incidence models. Having a greater number of years of education, a greater number of total years worked, and private health insurance at age 60 are protective against dementia, and these three factors all remained statistically significant in the long-term prediction models, even when all other predictors were included. However, these differentials were not statistically significant in the broad incidence models that accounted for the other predictors. The patterns are similar but slightly different in the CIND models: All three factors are protective, but the total number of years worked is the only one that is consistent in all three CIND models.

The lifestyle and health behavior predictors are noteworthy because they are modifiable risk factors and can be controlled by the individual. People who never exercise are more likely to develop CIND and dementia. These predictors were not available in the long-term prediction models. The patterns in alcohol consumption are in-line with the basic models, too: Moderate alcohol consumption is associated with reduced CIND and dementia. These differentials typically remain statistically significant, even when all other predictors are controlled. We found a notable contrast in BMI: High BMI is a consistent protective factor in the incidence models and a consistent risk factor in the long-term prediction models. Individuals whose BMI was 35 or more at age 60 had a statistically significantly higher chance of having CIND and dementia 20 years later, even in the broad models with many controls. However, high BMI at older ages appears to be protective against dementia as suggested by two-year and four-year incidence. This latter result could reflect reverse causality; individuals whose cognitive health deteriorates might experience weight loss.

Having fair or poor health is a consistent risk factor in all the dementia and CIND models. Having had a stroke is a consistent risk factor in the dementia models but not in the CIND models. Having diabetes at age 60 statistically significantly increases the chances of having CIND and dementia at age 80, even when all other controls are included. Diabetes is also a consistent risk factor in the CIND incidence models but not in the dementia incidence models. Several other health indicators were predictive of dementia incidence but not in the long-term prediction models, which could reflect reverse causality. These factors include having psychiatric problems, depression, or restless sleep.

Several functional limitations are consistent risk factors in the dementia incidence models but not in the long-term prediction models, which could reflect reverse causality again. The only functional limitation that remained statistically significant in the broad long-term prediction models is having difficulties using the phone at age 60.

The cognition measures are the strongest and most consistent predictors of dementia and CIND incidence and prevalence, which shows that having cognitive reserve helps delay the onset of CIND and dementia.

The estimated effects of the psychosocial measures vary widely across models, and we did not see a clear pattern. For example, engaging in hobbies and novel information activities is protective against dementia incidence in two years but not in four years. Having a conscientious personality is estimated to be protective against dementia incidence in four years but not in two years. Thus, these results are somewhat sensitive, possibly because of these models' substantially smaller sample sizes and the large number of control variables.

Finally, several objectively measured physical health measures and genes are consistent predictors of CIND and dementia. The estimated effects of the grip strength test and the Alzheimer's polygenic score indicate that those two items are consistent predictors in all models. The estimated effects of the breathing test, the walking time test, and the semi-tandem balance time test indicate that those three items are consistent predictors in some models.

The annex includes estimated regression models stratified by sex; those models show that the results for men and women are very similar.

CHAPTER 5

Limitations and Conclusions

Limitations

This study has several limitations. We used observational data and statistical methods to estimate the relationship between the dementia risk factors and the outcomes. The primary goal of these models is to find statistical associations that can be used to estimate individuals' risk of developing dementia as a function of their characteristics. Although the models might suggest causal channels between the risk factors and the outcomes, they cannot establish causality, which is our first limitation.

The second limitation is that the cognitive impairment and dementia outcome measures produced by the models were based on an algorithmic prediction model rather than a clinical assessment. Although the measures were calibrated to a clinical assessment in a subsample of the HRS, such measures cannot be treated as equivalent to gold-standard clinical assessments. The measures closely approximated dementia prevalence by sex, age, and race and ethnicity, but the measures were not calibrated along all predictor variables used in this study.

The third limitation is that our statistical models treated the outcome measures as data, even though the outcome measures were based on a statistical model. Treating validated measures as data even when they are estimates is a fairly standard approach in the literature. Nevertheless, it is still a limitation.

The fourth limitation is that the multivariate regression models presented in Chapter 4 included many predictor variables, but the standard errors and p-values were not adjusted for the multiple testing problem. Consequently, some of the statistically significant findings might not hold up in an independent sample.

The fifth limitation is that some of the predictor variables were only available in the most recent HRS waves, which reduced the sample size in the incidence models and prevented their use in the long-term prediction models at the time of this study. It would be worthwhile to re-estimate these models when data from additional HRS waves become available.

Conclusions

We evaluated the predictive power of 181 potential risk factors for dementia and cognitive impairment using the HRS—a large, nationally representative, longitudinal survey—and a validated probabilistic measure of cognitive impairment and dementia. We studied the predictive power of many potential risk factors for dementia, such as demographics, SES, labor-market measures, lifestyle and health behaviors, self-reported and objectively measured health, genes, parental health, cognitive abilities, and psychosocial factors. We estimated how these factors predicted CIND and dementia two, four, and twenty years later.

Some of our findings were in-line with prior literature, such as that physical health, having had a stroke, cognitive abilities, functional limitations, and particular genes strongly predict future incidence and prevalence of cognitive impairment and dementia. We also identified predictors that either received less attention in the literature or had mixed results: Individuals who were born in the South face statistically significantly higher chances of developing dementia, even when controlling for many other factors, possibly because the quality of education is lower in Southern states than in the rest of the country (Seblova et al., 2023). We found similarly elevated chances of CIND and dementia among those who did not have a private health insurance plan at age 60, who never worked or worked only a few years, who had diabetes or a BMI of 35 or more at age 60, who never drank alcohol or drank excessively, who never exercised, who scored low on various physical measure tests (i.e., breathing, grip strength, walking speed, and balance), who had less conscientious personality traits, and who engaged less in hobbies and novel information activities. We found that non-Hispanic Black and Hispanic individuals face statistically significantly higher chances of experiencing dementia incidence and prevalence, but these differentials shrink or disappear when we account for observable differences, such as SES.

We found strong associations between the outcomes and several modifiable risk factors. Therefore, our results suggest that there might be scope for slowing cognitive decline and dementia among at-risk people through behavioral changes and interventions. Such lifestyle modifications could be achieved by individuals taking the initiative to make such necessary changes, and public policy could also play an important role. Our results suggest that it might be beneficial for maintaining cognitive health to exercise at least sometimes, even if it is only light physical activity, such as walking. Consuming alcohol in moderation, working longer, and engaging in hobbies and novel information activities after retirement are also associated with a lower risk of developing dementia. Similarly, maintaining good physical health is associated with reduced dementia incidence, which suggests that adopting a healthy lifestyle might be beneficial not only for general health but also for brain health. Furthermore, we found in the long-term prediction models that individuals whose BMI index was 35 or more at age 60, those who had diabetes, and those who did not have private health insurance at age 60 have an elevated chance of developing dementia in the next 20 years, and these differentials remain large and statistically significant even when accounting for all other predictors. All these findings point toward the importance for policymakers and other stakeholders to promote healthy behaviors in the population and to strengthen individuals' access to quality health care.

Our results might find two main uses. The first is in prediction. A macro-level prediction of prevalence would help plan for the very high monetary and caregiving costs if prevalence were to increase. At the individual level, identifying subpopulations at elevated risk would permit the channeling of resources to them that encourage them to engage in advance planning. The second use is in prevention. Although our results do not quantify the effects of an intervention, they suggest where to concentrate the research that could aim to quantify these effects.

Abbreviations

ADAMS	Aging, Demographics, and Memory Study
ADL	activity of daily living
BMI	body mass index
CES-D	Center for Epidemiologic Studies Depression Scale
CIND	cognitive impairment, not dementia
DK/RF	don't know or refused to answer
ECog	expected value of latent cognition status
HRS	Health and Retirement Study
IADL	instrumental activity of daily living
PrCIND	probability of cognitive impairment, not dementia
PrDem	probability of dementia
SD	standard deviation
SES	socioeconomic status

References

Ahlskog, J. Eric, Yonas E. Geda, Neill R. Graff-Radford, and Ronald C. Petersen, "Physical Exercise as a Preventive or Disease-Modifying Treatment of Dementia and Brain Aging," *Mayo Clinic Proceedings*, Vol. 86, No. 9, September 2011.

Alzheimer's Association, "2023 Alzheimer's Disease Facts and Figures," *Alzheimer's and Dementia*, Vol. 19, No. 4, April 2023.

Andel, Ross, Deborah Finkel, and Nancy L. Pedersen, "Effects of Preretirement Work Complexity and Postretirement Leisure Activity on Cognitive Aging," *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, Vol. 71, No. 5, September 2016.

Bugliari, Delia, Joanna Carroll, Orla Hayden, Jessica Hayes, Michael D. Hurd, Stephen Lee, Regan Main, Colleen M. Mccullough, Erik Meijer, Philip Pantoja, and Susann Rohwedder, *RAND HRS Longitudinal File 2020 (V1)*, RAND Corporation, TL-A2097-1-v2, 2023. As of August 7, 2023:
https://www.rand.org/pubs/tools/TLA2097-1-v2.html

Crimmins, Eileen, Heidi Guyer, Kenneth Langa, Mary Beth Ofstedal, Robert Wallace, and David Weir, *Documentation of Physical Measures, Anthropometrics and Blood Pressure in the Health and Retirement Study*, Survey Research Center, University of Michigan, February 2008.

Crimmins, Eileen M., Jung Ki Kim, Kenneth M. Langa, and David R. Weir, "Assessment of Cognition Using Surveys and Neuropsychological Assessment: The Health and Retirement Study and the Aging, Demographics, and Memory Study," *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, Vol. 66, Suppl. 1, July 2011.

Ferretti, Maria Teresa, Maria Florencia Iulita, Enrica Cavedo, Patrizia Andrea Chiesa, Annemarie Schumacher Dimech, Antonella Santuccione Chadha, Francesca Baracchi, Hélène Girouard, Sabina Misoch, Ezio Giacobini, Herman Depypere, and Harald Hampel, "Sex Differences in Alzheimer Disease—The Gateway to Precision Medicine," *Nature Reviews Neurology*, Vol. 14, No. 8, August 2018.

Fisher, Gwenith G., Halimah Hassan, Jessica D. Faul, Willard L. Rodgers, and David R. Weir, *Health and Retirement Study Imputation of Cognitive Functioning Measures: 1992–2010*, Survey Research Center, University of Michigan, January 2015.

Gianattasio, Kan Z., Adam Ciarleglio, and Melinda C. Power, "Development of Algorithmic Dementia Ascertainment for Racial/Ethnic Disparities Research in the US Health and Retirement Study," *Epidemiology*, Vol. 31, No. 1, January 2020.

Heeringa, Steven G., Gwenith G. Fisher, Michael D. Hurd, Kenneth M. Langa, Mary Beth Ofstedal, Brenda L. Plassman, Rogers Willard, and David R. Weir, *Aging, Demographics and Memory Study (ADAMS): Sample Design, Weighting and Analysis for ADAMS*, Institute for Social Research, University of Michigan, June 2009.

Hudomiet, Péter, Michael D. Hurd, and Susann Rohwedder, "Trends in Inequalities in the Prevalence of Dementia in the U.S.," *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 119, No. 46, November 2022.

Hudomiet, Péter, Michael D. Hurd, and Susann Rohwedder, "Documentation Predicted Cognition and Dementia Measures Release 1," *Proceedings of the National Academy of Sciences of the United States of America*, January 24, 2023.

Ioannidis, John P. A., "Why Most Published Research Findings Are False," *PLoS Medicine*, Vol. 2, No. 8, August 2005.

Javeed, Ashir, Ana Luiza Dallora, Johan Sanmartin Berglund, Arif Ali, Liqata Ali, and Peter Anderberg, "Machine Learning for Dementia Prediction: A Systematic Review and Future Research Directions," *Journal of Medical Systems*, Vol. 47, No. 1, February 2023.

Juster, F. Thomas, and Richard Suzman, "An Overview of the Health and Retirement Study," *Journal of Human Resources*, Vol. 30, 1995.

Kim, Mi-Young, Kyeongjin Kim, Chang Hyung Hong, Sang Yoon Lee, and Yi-Sook Jung, "Sex Differences in Cardiovascular Risk Factors for Dementia," *Biomolecules and Therapeutics*, Vol. 26, No. 6, November 2018.

Pires, Mariana, and Ana Cristina Rego, "Apoe4 and Alzheimer's Disease Pathogenesis-Mitochondrial Deregulation and Targeted Therapeutic Strategies," *International Journal of Molecular Sciences*, Vol. 24, No. 1, January 2023.

Prince, Martin, Emiliano Albanese, Maëlenn Guerchet, and Matthew Prina, *World Alzheimer Report 2014: Dementia and Risk Reduction—An Analysis of Protective and Modifiable Factors*, Alzheimer's Disease International, 2014.

Seblova, Dominika, Chloe Eng, Justina F. Avila-Rieger, Jordan D. Dworkin, Kelly Peters, Susan Lapham, Laura B. Zahodne, Benjamin Chapman, Carol A. Prescott, Tara L. Gruenewald, Thalida E. Arpawong, Margaret Gatz, Rich J. Jones, Maria M. Glymour, and Jennifer J. Manly, "High School Quality Is Associated with Cognition 58 Years Later," *Alzheimer's and Dementia*, Vol. 15, No. 2, April–June 2023.

Smith, Jacqui, Lindsay Ryan, Marina Larkina, Amanda Sonnega, and David Weir, *Psychosocial and Lifestyle Questionnaire: 2006–2022*, Survey Research Center, Institute for Social Research, University of Michigan, February 2023.

Survey Research Center, Institute for Social Research, University of Michigan, "The Health and Retirement Study," database, March 2023. As of August 7, 2024:
https://hrs.isr.umich.edu/

Then, Francisca S., Tobias Luck, Melanie Luppa, Marleen Thinschmidt, Stefanie Deckert, Karen Nieuwenhuijsen, Andreas Seidler, and Steffi G. Riedel-Heller, "Systematic Review of the Effect of the Psychosocial Working Environment on Cognition and Dementia," *Occupational and Environmental Medicine*, Vol. 71, No. 5, May 2014.

Ware, Erin, Arianna Gard, Lauren Schmitz, and Jessica Faul, "HRS Polygenic Scores—Release 4.3: 2006–2012 Genetic Data," data file, Survey Research Center, University of Michigan, February 2021.

Wimo, Anders, Katrin Seeher, Rodrigo Cataldi, Eva Cyhlarova, Joseph L. Dielemann, Oskar Frisell, Maëlenn Guerchet, Linus Jönsson, Angeladine Kenne Malaha, Emma Nichols, Paola Pedroza, Martin Prince, Martin Knapp, and Tarun Dua, "The Worldwide Costs of Dementia in 2019," *Alzheimer's and Dementia*, Vol. 19, No. 7, July 2023.

Zissimopoulos, Julie M., Bryan C. Tysinger, Patricia A. St. Clair, and Eileen M. Crimmins, "The Impact of Changes in Population Health and Mortality on Future Prevalence of Alzheimer's Disease and Other Dementias in the United States," *Journals of Gerontology, Series B, Psychological Sciences and Social Sciences*, Vol. 73, Suppl. 1, April 2018.