



Nationwide Nuances: Dementia Predictions Across Diverse Demographics
Engineering Dynamics

POC: Dave Kaniss
Dave@engrdynamics.com
240-925-3470

Executive Summary

Alzheimer's disease/dementia/cognitive decline is a controversial and complex set of diagnoses. Attempts to study prevalence patterns have been wildly distorted by different systems of measurement, including medical diagnosis, self-report, and testing by survey-takers. However, whatever the system of measurement, analysts agree that dementia rates in developed countries have been plummeting for the last few decades (Hudomiet et al. 2022). There is also little question that the two most important demographic factors in dementia risk are age and education (Armstrong 2019). Evidence strongly suggests that increasing education has contributed to declines in dementia, and analysts note that there are implications for gender in this, as women have now become better educated than men (Hudomiet et al. 2022). Inasmuch as socioeconomic factors such as race appear to matter, they mostly appear significant related to known factors that influence risk for dementia, such as marriage rates, social contact, diabetes, and blood pressure (Hersi et al. 2017). Effects of gender/biological sex are controversial within this literature, but the increasing consensus appears to be that it is insignificant (although the patterns for women are quite different from those for men) (Hudomiet et al. 2022). Other than age and education, only two demographic factors appear to be robust against broader individual controls: foreign-born immigration status (Hudomiet et al. 2022), and LGBT identification (Hsieh et al. 2022). Although much of the LGBT differences are due to depression, large effects remain even after controlling for depression and numerous other factors (Hsieh et al. 2022). Many individual physical risk factors are significant, including history of traumatic brain injury, prolonged exposure to certain metals such as aluminum, and malnutrition (Armstrong 2019). Individual behavioral factors also matter, such as current smoking (increase) and current mild alcohol consumption (decrease) (Hersi et al. 2017; Hsieh et al. 2022; Jeon et al. 2023).

For our analysis, we will be using multiple waves of the National Health Interview Survey (NHIS), which is a free and publicly available nationally representative cross-sectional survey of the non-institutionalized U.S. adult population up to age 85 (<https://nhis.ipums.org/>). The average annual sample is around 100,000 people in 45,000 households. Key dependent variables are DEMENTIAEV ("have you ever been told by a health professional that you have dementia, including alzheimer's") from 2019-22; LAMEMORCON ("have you ever had difficulty remembering, concentrating, or both?"), which can be focused solely on respondents who had difficulty remembering, from 2010-22; LAMEMCONDIF, which asks for level of difficulty remembering for the same years; LAMEMDIFOFT, which asks for frequency of difficulty remembering for the same years. We believe using these broader measures of cognitive decline will help capture people all along the dementia spectrum. One of the common problems in previous studies of dementia has been attempting to study this condition as a black-and-white diagnosis when it is clearly a spectrum disorder.

The NHIS contains a vast array of potential independent variables pertaining to both demographics and health that can be used to look for predictive factors associated with cognitive decline, from cardiovascular issues to mental health and even to the largely under-researched variable of sexual orientation. With previous analyses suggesting that factors associated with dementia have also shifted across time, having more than a decade of data will allow us to analyze the potential effects of time, alone and in combination with many other key predictor variables.

Data Description

Basic Information

The National Health Interview Survey (NHIS) is a free and publicly available data set sponsored by the Centers for Disease Control (CDC) and conducted by the United States Census Bureau annually since 1957. Relevant questions for the purposes of studying cognitive decline have been asked every year since 2010, resulting in 12 years of data for key dependent variables. The survey is conducted throughout the year to minimize seasonal bias, with computer assisted face-to-face interviewing occurring whenever possible (although at least a third of respondents are usually contacted by phone). As a major Census Bureau project, it is a gold standard of representation for the U.S. non-institutionalized population up to age 85 (excluding current prisoners, wholly military families, and most importantly for these purposes, those in residential homes). It utilizes a complex stratification sampling system (which has been simplified since 2019), requiring the use of sample weights to achieve representative estimates of the population. Prior to 2019, information was collected from one household adult on all members of the household; after 2019, it solely focuses on the queried adult and one child in the household if present (although children are irrelevant for this analysis).

IPUMS (the Integrated Public Use Microdata Series) has “harmonized” the data across years, making it possible to easily download manageable data sets and compare questions and answers across time. Sample sizes vary greatly across survey waves. During the years for which good data on cognitive decline are available, sample size varies greatly from a low of 28,854 households/35,115 individuals in 2022 to a high of 60,347 households/112,053 individuals in 2014 for the total sample. However, from 2010-2018, only a quarter subsample was asked the relevant questions about cognitive decline, and specific sampling details for those questions are described below.

For more detailed information on sampling, accessibility, and survey design, see:

https://www.cdc.gov/nchs/nhis/about_nhis.htm

https://www.cdc.gov/nchs/nhis/about_nhis.htm#procedures

https://nhis.ipums.org/nhis/userNotes_sampledesign.shtml

<https://nhis.ipums.org/nhis-action/faq#ques0>

https://nhis.ipums.org/nhis/userNotes_size.shtml

Utility & Rigor

The NHIS contains 4 key dependent variables measuring cognitive decline in slightly different ways:

DEMENTIAEV (“have you ever been told by a health professional that you have dementia, including alzheimer’s”) from 2019-22 and 2007

https://nhis.ipums.org/nhis-action/variables/DEMENTIAEV#survey_text_section

Sampling universe was about 30,000 cases annually from 2019-22, with around 350 people answering “yes” each year.

LAMEMCONDIF (“Do you have difficulty remembering or concentrating?”--no, some, a lot, or cannot do at all) from 2010-22. From 2010-17, this question was asked to a quarter subsample, and from 2018-2022 was asked of the full sample.

https://nhis.ipums.org/nhis-action/variables/LAMEMCONDIF#survey_text_section

The universes for the variable vary greatly by year, but the average annual number of cases for “a lot” is more than 700 in 2018-22 with the full sample, and more than 300 in 2010-2017 with

the quarter-sample. Less than 20 people report annually they cannot remember or concentrate at all.

LAMEMORCON (“have you ever had difficulty remembering, concentrating, or both?”), which can be focused solely on respondents who had difficulty remembering, from 2010-22.

This variable sorts out responses from the previous variable to clarify whether respondents specifically have difficulty remembering, *or* concentrating, *or* both.

https://nhis.ipums.org/nhis-action/variables/LAMEMORCON#description_section

The universes for this variable vary greatly by year, but the average number of cases reporting “difficulty remembering only” is well over 1,000 annually, and more than 1,600 for “difficulty remembering and concentrating.”

LAMEMDIFOFT (“how often would you say you have difficulty remembering?”--sometimes, often, or all the time) which focuses on respondents who reported difficulty remembering, from 2010-22.

https://nhis.ipums.org/nhis-action/variables/LAMEMDIFOFT#survey_text_section

The universes for this variable vary greatly by year, with about 1,300 people annually reporting “frequently” plus “all the time” in 2018-22, and more than 600 annually for the same categories in 2010-17.

In terms of reliability and validity, pooling the samples for the years 2019-22 (when the question on dementia diagnosis was asked), among those who say they have never been diagnosed with dementia, 84% report no difficulty remembering, 12% say they have difficulty sometimes, 3% frequently, and 1% all the time. Among those who say they have been diagnosed with dementia, 16% report no difficulty remembering, 27% say they have difficulty sometimes, 26% frequently, and 31% all the time. (Note that sample sizes here are sufficiently large that the 1% of people who say they have never been diagnosed with dementia but have difficulty remembering “all the time” still constitutes 1399 cases). Of course, people may have difficulty remembering for a variety of causes (and people with dementia may, by definition, be unreliable reporters of their own experience of memory). Regardless, concordance between these variables appears to be about 84% in both directions, and it appears that using both measures separately and together should maximize the predictive power of models.

In terms of potential independent variables for technology to scour within this vast data set, there are classic socioeconomic indicators such as race, Hispanic ethnicity, whether the respondent is foreign-born; age, sex, sexual orientation, current marital and cohabiting status; whether the respondent is a veteran; educational attainment; whether the respondent ever worked or the last time they worked for pay; whether the family owns or rents their home; poverty line ratio, as well as measures of food insecurity; overall health status and BMI; medical history of angina, asthma, any cancer, coronary heart disease, high cholesterol, diabetes, hypertension, stroke, and major breathing obstructions; alcohol consumption, smoking, and physical activity; functional limitations, such as difficulty washing, dressing, walking, seeing, or hearing; fatigue and chronic pain; access to medical care and insurance; receipt of vaccinations; and simple measures for anxiety, depression, and panic disorder.

Innovation

Whereas so many data sets try to predict whether a person “has dementia” based on much-debated criteria about what “counts” as dementia, this data set allows for much more nuanced calculations based on scaled experiences of difficulty remembering, in addition to a

direct measure of whether the respondent was diagnosed with dementia. Using machine learning to specifically target people who say they are currently experiencing difficulty with memory “sometimes” (as opposed to those who say they experience difficulties “frequently” or “all the time”) might have great potential for looking for the predictors of early stages of cognitive decline.

Many studies and reviews have attempted to look at a wide variety of factors in predicting who will get dementia, but very few data sets have the vast sample size and variable set of the NHIS. With such large sample sizes, the experiences of merely 1% of people become possible to mathematically predict, especially with pooled data. Other analyses that have tried to simultaneously control for the effects of multiple predictors of cognitive decline have usually been constrained by sample size, forcing most previous research to depend heavily on the uncertainty of meta-analyses. Most multivariate models require vast sample sizes to produce stable estimates when many variables are added, but unlike the NHIS, few studies possess the large samples to support those mathematical processes. In short, the large sample size, accompanied by with so many socioeconomic variables, as well as physical and mental health variables, makes it possible to ask and answer probability questions with the NHIS that are extremely difficult without such a large data set.

Furthermore, dementia rates have plummeted across the last two decades in America, and these data allow for exploration of trends across time in a way that few data sets can. However, initial data exploration with the NHIS does not support previous findings showing a decrease in dementia from 2010-2022; indeed, by many different measures and with no controls, very tentatively, memory difficulties appear to have increased somewhat across this time frame, especially among those 75 and older (although we should keep in mind these numbers only apply to individuals who are not institutionalized, and these numbers could be showing that people with dementia are more likely to be living at home now than they were 10 years ago).

Moreover, since 2013, the NHIS contains data on sexual orientation, which is a category that has been largely unexplored in analyses of dementia to date, despite speculative hints in the literature that bisexuals might be at particular risk. Initial data exploration tentatively suggests that bisexuality is indeed strongly associated with memory difficulties.

Sample Characteristics and Representation

The NHIS is a nationally representative data set which until 2018 intentionally oversampled some racial and ethnic minorities, necessitating the use of sampling weights to produce population estimates for data analysis. However, prior research suggests that most effects of race and ethnicity on dementia can be controlled for with several factors that the NHIS measures: marriage, diabetes, and blood pressure (unfortunately, the NHIS does not have many measures for social contact for more than a scant handful of years, which is another relevant variable here). The NHIS also measures foreign-born status, which previous research suggests is significant, as well as age and education, which are the two most significant demographic factors.

Participants in the NHIS are drawn from all over the US, and the Census Bureau goes to great lengths to try to ensure high levels of participation and representation. Whether participants are compensated in this government-sponsored survey is not stated in their main user information.

Unweighted, the raw pooled sample for 2010-2022 that has valid responses for the central question LAMEMCONDIF is distributed thus along major independent variables:

Education: Not completed high school--28,791 (11.4%), High school graduate--56,930 (22.3%), Some college--82,501 (32.3%), College--53,063 (20.8%), Graduate or professional degree--32,699 (12.8%)

Nativity: Born in US--209,510 (82.1%), born outside US--41,896 (16.4%), unknown--3,536 (1.4%)

Hispanic ethnicity: Not Hispanic--218,778 (85.7%), Hispanic--36,375 (14.3%)

Race: White--195,185 (76.5%), Black--31,340 (12.3%), Asian--14,592 (5.7%), Indigenous only--2,476 (1.0%), Other and multiple races--5,616 (2.2%), Unknown--5,838 (2.3%)

Sex: Male--115,761 (45.4%), Female--139,382 (54.6%)

Age: 18-30--47,216 (18.5%), 31-40--41,267 (16.2%), 41-50--38,521 (15.1%), 51-60--43,823 (17.2%), 61-70--43,429 (17.0%), 71-80--27,036 (10.6%), 81-85--13,565 (5.3%)

With such a large sample size, it is possible to use pooled NHIS data to analyze extremely small and understudied subgroups, such as immigrants and Indigenous people. With over 13,000 people aged 81-85, it is even possible to simply focus on trends in this older age group, or on the 27,000 respondents aged 71-80, who presumably are the group to target for earlier detection of AD/ADRD.

Usability

NHIS data are freely and publicly government available across years through IPUMS and should remain so indefinitely. All identifying information for respondents has been removed from the public use files, and analytic use of this data should pose no confidentiality risks to respondents.

Team Introduction

The Engineering Dynamics team is an interdisciplinary team of engineers, AI specialists, and Human Sciences experts united by the belief that "Engineering is a People Problem." By blending human and applied sciences, we create effective engineering solutions that understand and respond to the people they serve. Our Human Sciences team provides valuable insights into the human factors that influence the adoption and use of technology. While algorithms and processes are well-suited to solving technical challenges, understanding the human component is essential to addressing any complex engineering problem.

The team also has considerable expertise in AI and ML, with extensive backgrounds in applying these technologies to various real-world challenges. Their experience includes developing state-of-the-art algorithms for data analysis, natural language processing, and prediction modeling, ensuring that their solutions benefit from the latest advances in the field. In previous projects, we've employed similar interdisciplinary approaches to address social disparities in health, public safety, and technology domains. This collective experience positions us to develop a solution that succeeds in predicting early predictors of dementia. Our team is well-established and has consistently proven successful in a wide range of previous challenges by identifying structural factors that contribute to overall success. This cohesive dynamic ensures efficient problem-solving, fostering an environment where innovation thrives. We believe in fostering trust and engagement with underserved communities by ensuring their perspective and experiences inform our solutions. Our commitment to individual communities is unwavering and propels us to deliver a solution centered on their unique healthcare needs.

Julie Fennell, Ph. D., project leader, and sociologist, brings many years of research in sexual minority communities. With a strong background in public health, she guides our efforts emphasizing community engagement and structural factors, as well as highlighting Diversity, Equity, Inclusion, and Accessibility (DEIA) in our approach.

Bethel Quick, psychologist, and mechanical engineer, provides operational oversight, spearheads the development of strategic partnerships with health providers, and leads our insight into behavioral science. Her expertise ensures our solutions are nuanced, understanding, and adaptive to diverse needs and experiences.

Dave Kaniss, as our Systems Engineer, ensures that our solutions integrate seamlessly into existing healthcare structures, assessing and adjusting the blueprint to accommodate evolving healthcare trends and technologies, promising the solution's long-term feasibility and sustainability.

Tim Kaniss, our lead data scientist focusing on C/C++ programming, constructs and refines our model, transforming complex data into actionable insights to improve healthcare services..

References

- Armstrong, Richard A. 2019. "Risk Factors for Alzheimer's Disease." *Folia Neuropathologica* 57 (2): 87–105. <https://doi.org/10.5114/fn.2019.85929>.
- Hersi, Mona, Brittany Irvine, Pallavi Gupta, James Gomes, Nicholas Birkett, and Daniel Krewski. 2017. "Risk Factors Associated with the Onset and Progression of Alzheimer's Disease: A Systematic Review of the Evidence." *NeuroToxicology, Determinants of Neurological Disease*, 61 (July): 143–87. <https://doi.org/10.1016/j.neuro.2017.03.006>.
- Hsieh, Ning, Hui Liu, and Wen-Hua Lai. 2020a. "Elevated Risk of Cognitive Impairment Among Older Sexual Minorities: Do Health Conditions, Health Behaviors, and Social Connections Matter?" *The Gerontologist* 61 (3): 352–62. <https://doi.org/10.1093/geront/gnaa136>.
- Hudomiet, Péter, Michael D. Hurd, and Susann Rohwedder. 2022. "Trends in Inequalities in the Prevalence of Dementia in the United States." *Proceedings of the National Academy of Sciences* 119 (46): e2212205119. <https://doi.org/10.1073/pnas.2212205119>.
- Jeon, Keun Hye, Kyungdo Han, Su-Min Jeong, Junhee Park, Jung Eun Yoo, Juhwan Yoo, Jinkook Lee, SangYun Kim, and Dong Wook Shin. 2023. "Changes in Alcohol Consumption and Risk of Dementia in a Nationwide Cohort in South Korea." *JAMA Network Open* 6 (2): e2254771. <https://doi.org/10.1001/jamanetworkopen.2022.54771>.
- Manly, Jennifer J., Richard N. Jones, Kenneth M. Langa, Lindsay H. Ryan, Deborah A. Levine, Ryan McCammon, Steven G. Heeringa, and David Weir. 2022. "Estimating the Prevalence of Dementia and Mild Cognitive Impairment in the US: The 2016 Health and Retirement Study Harmonized Cognitive Assessment Protocol Project." *JAMA Neurology* 79 (12): 1242–49. <https://doi.org/10.1001/jamaneurol.2022.3543>.