



Published in final edited form as:

*Alzheimer Dis Assoc Disord.* 2020 ; 34(4): 325–332. doi:10.1097/WAD.0000000000000400.

## PREDICTORS OF DEMENTIA IN THE OLDEST OLD: A NOVEL MACHINE LEARNING APPROACH

Yichen Jia, MS<sup>1</sup>, Chung-Chou H. Chang, PhD<sup>1,2</sup>, Tiffany F. Hughes, PhD, MPH<sup>3</sup>, Erin Jacobsen, MS<sup>4</sup>, Shu Wang, PhD<sup>5</sup>, Sarah Berman, MD, PhD<sup>6</sup>, M. Ilyas Kamboh, PhD<sup>4,7,8</sup>, Mary Ganguli, MD, MPH<sup>4,6,8</sup>

<sup>1</sup>Department of Biostatistics, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA, USA

<sup>2</sup>Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

<sup>3</sup>Department of Sociology, Anthropology, and Gerontology, Youngstown State University, Youngstown, OH, USA

<sup>4</sup>Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

<sup>5</sup>Department of Biostatistics, University of Florida, Gainesville, FL, USA

<sup>6</sup>Department of Neurology, University of Pittsburgh School of Medicine Pittsburgh, PA, USA

<sup>7</sup>Department of Human Genetics, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA, USA

<sup>8</sup>Department of Epidemiology, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA, USA

### Abstract

**Background**—Incidence of dementia increases exponentially with age; little is known about its risk factors in the ninth and tenth decades of life. We identified predictors of dementia with onset after age 85y in a longitudinal population-based cohort.

**Methods**—Based on annual assessments, incident cases of dementia were defined as those newly receiving Clinical Dementia Rating (CDR®) 1. We used a machine learning method, Markov modeling with HyDaP clustering, to identify variables associated with subsequent incident dementia.

**Results**—Of 1,439 participants, 641 reached age 85y during ten years of follow-up and 45 of these became incident dementia cases. Using HyDaP, among those aged 85+y, probability of incident dementia was associated with worse self-rated health, more prescription drugs, subjective memory complaints, heart disease, cardiac arrhythmia, thyroid disease, arthritis, reported hypertension, higher systolic and diastolic blood pressure, and hearing impairment. In the

---

**Corresponding Author:** Chung-Chou H. Chang, PhD, 200 Meyran Ave., Suite 200, Pittsburgh, PA 15213, *Phone:* 412-692-4868, *changj@pitt.edu.*

Conflict of Interest

Dr. Ganguli was a member of Biogen “Patient Journey Working Group” from 2016–2017. The remaining authors declare no conflicts of interest.

subgroup aged 85–89y, risk of dementia was also associated with depression symptoms, not currently smoking, and lacking confidantes.

**Conclusion**—An atheoretical machine learning method revealed several factors associated with increased probability of dementia after age 85y in a population-based cohort. If independently validated in other cohorts, these findings could help identify the oldest-old at the highest risk of dementia.

### Keywords

85+; 90+; Epidemiology; Risk factors; Community study

---

## Introduction

As life expectancy increases, particularly in the high-income countries, the public health burden of dementia will increase in parallel. Identifying predictors of dementia can potentially allow early interventions to be targeted to those at greatest risk of dementia. Two facts are now established about the incidence of dementia, i.e., the rate at which new cases of dementia occur in the population: first, that the incidence rate increases with age<sup>1–4</sup>, and second, that among the oldest-old, defined as 85+ y or 90+ y in different studies, few if any independent risk factors have been found other than age<sup>5–7</sup>.

We previously reported, after 5 years of follow-up of an ongoing longitudinal, population-based cohort study, that incidence of dementia increased exponentially with age, with a median age at onset of 87 years. That is, half the incident dementia cases in the community experienced symptom onset at or after age 87. Using joint modeling to account for attrition, several well-recognized risk and protective factors showed significant effects in cases with onset below age 87y. However, no measured factor showed effects after that age<sup>4</sup>. Now, after 10 years of follow-up, we employed a novel machine learning approach to determine whether any variables in our database might be associated with future incident dementia in the oldest old. We emphasize that our search was for potential predictors of dementia, and not for biomarkers of underlying dementia-causing pathologies. We included predictor variables which could represent independent risk factors as well as others which could be early subclinical features or epiphenomena of the dementing process.

## Methods

### Study Population

The Monongahela-Youghiogheny Healthy Aging Team (MYHAT), is a population-based cohort study with a primary focus on incidence and risk factors for mild cognitive impairment (MCI) and dementia. The MYHAT cohort was recruited between 2006 and 2008 by age-stratified random sampling from publicly available voter registration lists for targeted small-town communities in southwestern Pennsylvania. Details of sampling, recruitment, and cohort characteristics have been previously reported<sup>4,8</sup>. Inclusion criteria were age 65y or older, living independently (not residing in a long-term care facility) at study entry; exclusion criteria were sensory impairment sufficient to preclude neuropsychological testing, and decisional incapacity. Initial screening was performed on 2,036 participants, of

whom 54 were excluded from the full evaluation based on substantial baseline cognitive impairment (scores <21 on the age- and education-adjusted Mini-Mental State Examination (MMSE) rendering them unsuitable for a study of MCI<sup>9,10</sup>. The full evaluation was conducted on the remaining 1,982 participants who, at study entry, had a mean (SD) age of 77.6 (7.4) years; 61.1% were women and 94.8% were of mixed European descent; their median educational level was high school graduate. All procedures were approved by the University of Pittsburgh Institutional Review Board, and all participants provided written informed consent.

### Assessment Protocol

At baseline and up to ten annual follow-up visits, trained research interviewers performed evaluations including but not limited to: questionnaire-based items including demographics, marital status, living arrangements, everyday functional independence, subjective memory complaints, health history (self-report of receiving specific diagnoses from health care professionals), hearing and vision, depression, anxiety, lifestyle factors including physical activity, social engagement, cognitive engagement, current and past alcohol and tobacco use, current prescription medications and nonprescription supplements; and also directly observed variables: a brief physical examination including blood pressure (BP), focused neurological examination, *APOE* genotyping, and a neuropsychological assessment<sup>11</sup>. Thus, the assessment included both potential independent risk factors as well as potential early signs or epiphenomena of the dementing process.

### Clinical Dementia Rating® (CDR)

All interviewers received online training and certification in the use of the Clinical Dementia Rating ® Staging Instrument (<https://knightadrc.wustl.edu/cdr/cdr.htm>) modified for field use<sup>12,13</sup>. After completing each annual assessment, the interviewer generated a CDR based on the standard algorithm.

### Tracking, follow-up, attrition

Following each annual visit, interviewers maintain contact with participants via telephone calls at regular intervals, and mail participants annual birthday cards and quarterly newsletters to maximize retention. However, given the increasing age of the cohort, we have experienced increasing attrition due to mortality, illness, relocation, and other forms of dropout over the ten-year observation period.

### Identification of incident dementia cases

Dementia was defined as CDR = 1, thus including mild, moderate, and severe dementia but not mild cognitive impairment. Removing all baseline prevalent cases of dementia with CDR = 1, the remaining individuals, with CDR of 0 or 0.5 at baseline or any annual visit, were, by definition, at risk of becoming incident cases with CDR = 1 at a subsequent visit. The age at onset of dementia for a given participant is taken as the midpoint between the dates of the assessment when CDR was = 1 for the first time and the previous assessment when CDR was <1. We did not exclude individuals with CDR of 0.5 because almost every incident case of

CDR =1 had an earlier CDR=0.5; however, not all cases of CDR of 0.5 progressed to CDR =1, as we have previously reported<sup>14</sup>.

## Statistical Methods

We previously reported the use of a hypothesis-driven joint modeling approach, accounting for attrition, which failed to identify potential risk and protective factors for incident dementia in those aged 87+y at the 5-year follow-up point<sup>4</sup>. At the 10-year point, taking into consideration the still relatively low absolute number of incident cases and increasing attrition over the ten years, we decided to explore the same question using a non-traditional machine-learning Markov modeling method briefly described below and in more detail in Supplemental File S1. The advantages of this approach are that it does not restrict results to *a priori* hypotheses, but rather uses the entire database with multiple observations for each participant to provide the best opportunity to identify factors which increase the probability of transitioning to incident dementia.

**Hybrid density- and partition-based (HyDaP) algorithm**—HyDaP is a nonparametric unsupervised clustering algorithm designed for data consisting of mixed (both continuous and categorical) variables and data with or without natural clustering. In order to obtain robust clustering findings, HyDaP incorporates participant characteristics from all assessment cycles. The algorithm can also identify variables that are the most important for clustering<sup>15</sup>. The *subject* in this method was each participant at every assessment, so that participants who remained in the study longer provided more data points. We clustered all the *subjects* based on their characteristics into several subgroups (*aka* clusters) such that *subjects* within a given cluster are similar to one another but different from those in other clusters. Note that the clusters are composed of observations, not participants, and the model is focused on identifying homogenous groups of observations, regardless of which participants they came from.

Where a participant had an intermittently missing covariate e.g., smoking status was available at Cycle 2 and Cycle 4 but missing at Cycle 3, we imputed the missing value by carrying the last observation forward. In this example, we impute this participant's smoking status at Cycle 3 using his/her smoking status at Cycle 2. However, if a participant had missing values for a covariate for all cycles, e.g. if no *APOE* genotyping information was available throughout the study, we excluded this participant from the analysis.

**Empirical transition probability matrix**—The clusters obtained from HyDaP were treated as transient states (Supplemental Figure 1). We further added 4 absorbing states: dementia (CDR = 1), dropout due to death, dropout due to illness, and dropout due to other reasons. In contrast to transient states, by definition, once a *subject* enters an absorbing state, the probability of remaining in that state is 100%, i.e. there is no possibility of further transition. We treated dementia and the three types of dropouts as different absorbing states competing with one another. Thus, this approach accounts for the possibility of competing risks between dementia and attrition, thus ensuring that predictors identified for dementia are not artifacts of selective attrition. A detailed interpretation of transition probability is provided in the supplemental file.

Therefore, the method provides the estimated transition probability from a given cluster to any of the clusters (living or potentially transient states), and also the estimated transition probability to one of the four absorbing states (dementia, death, becoming too ill to participate, and other types of dropout), from a given cluster. Based on these transition probabilities, we would be able to identify which cluster(s) have higher probabilities of transitioning to dementia or attrition. Characteristics associated with these cluster(s) are the predictors/risk factors for dementia or attrition. Note that this approach does not test a specific hypothesis and therefore does not assess the associations of individual variables with the outcome or their statistical significance.

We applied the HyDaP algorithm to data from all participants aged 65+. We did not include age, sex, education, and *APOE\*4* carrier status into the clustering algorithm because these variables do not change over the course of the study, i.e. are static for a given person. We also excluded age from clustering since we planned *a priori* to look at two age groups 85–89y and 90+y. Based on the resulting cluster assignment for each participant at each cycle, we then calculated the empirical transition probabilities separately for the subsets aged 85–89y and aged 90+ years.

We further stratified the sample by sex, education, *APOE\*4* carrier status, and age, and calculated the transition probability for each stratum. Thus, we had total of 12 strata: women aged 85–89, women aged 90+, men aged 85–89, men aged 90+, high school or less education and aged 85–89, high school or less education and aged 90+, greater than high school education and aged 85–89, greater than high school education and aged 90+, *APOE\*4* carriers aged 85–89, *APOE\*4* carriers aged 90+, *APOE\*4* non-carriers aged 85–89, and *APOE\*4* non-carriers aged 90+. We calculated the proportion of incident dementia cases within each stratum to examine the effect of sex, education, and *APOE\*4* on risk of incident dementia.

## Results

Of the 1,982 fully assessed participants at baseline, we included 1,439 participants who had complete or partially imputed data on all covariates. Their mean (SD) age was 77.17 (7.34) y, 59.5% were women, 95.8% were of European descent, 87.8% had high school or greater education, and 20.6% carried at least one *APOE\*4* allele. At the 10-year follow-up, 270 were still participating; their mean (SD) age was 83.55 (5.96) years, 62.2% were women, 97.0% were of European descent, 91.1% had high school or higher education, and 18.9% were *APOE\*4* carriers.

Among the 1,439 participants with complete data, we excluded the 9 prevalent dementia cases from the HyDaP model. In the HyDaP sample, we observed 75 incident cases of dementia over ten years of follow-up (Supplemental Table 1) with a median age at dementia onset of 86 years. Among these 75 incident cases, 30 had onset at ages 65–84y, 21 had onset while aged 85–89y, and 24 had onset after age 90y. Our focus in the current analyses is on the 45 incident cases with onset after age 85y.

The 1,430 participants included in the analysis contributed data on multiple assessments such that 9,358 *subjects* (as defined earlier under Methods) were included in the HyDaP algorithm for clustering. Based on the consensus matrix plot, the optimal number of clusters is 4. Supplemental Table 2 shows the distributions of characteristics for each of these 4 clusters; e.g., participants in cluster 1 tend to have more sleep problems; people in cluster 2 are less physically and socially active; participants in cluster 3 in general have profile suggesting better health; cluster 4 contains people with higher waist:hip ratio (abdominal obesity).

Table 1 shows the proportion of incident dementia cases in each stratum by sex, education or *APOE\*4* carrier status. *APOE\*4* was associated with higher risk of dementia among aged 85–89 ( $p = 0.03$ ), however, this did not withstand correction for multiple comparisons ( $n = 2$  comparisons). There is no evidence suggesting that incident dementia is related to either sex or education.

The transition probability matrices among those aged 85–89 years and among those aged 90+ years old can be found in Figures 1 and 2, respectively. A participant may potentially change conditions from one assessment cycle to the next; therefore, can either remain in the same cluster, or transit from one cluster to one of the three other clusters, or transit from one cluster to one of the four absorbing states (dementia, death, or other dropout category). For example, in Figure 1, an 85–89 year-old person who started at cluster 1 has a 60.8% chance of remaining in cluster 1, a 17.1% chance of transitioning to cluster 2, and a 0.7% chance of transitioning to dementia in the next cycle. After stratifying by sex (men/women), *APOE\*4* allele (present/absent), or education (high school /> high school), the transition probability matrices among those 85–89y old and among 90+ years old are displayed in Supplemental Figures 2–4. An 85–89 year-old person in cluster 2 has a higher probability of transitioning to incident dementia compared to a person in clusters 1, 3 or 4 (Figure 1). A person 90y old also has a higher probability transitioning to incident dementia at the next cycle if s/he is in cluster 2 compared to being in clusters 1, 3, or 4 (Figure 2). Thus, we compare the cluster with higher probability of transitioning to incident dementia (cluster 2) to the combined clusters with lower probability of transitioning to incident dementia (clusters 1, 3 and 4).

In general, the characteristics for the clusters most likely to progress to incident dementia between those aged 85–89y and aged 90+y were fairly similar (Table 2 and 3). They all had poor to fair self-rated health, took more prescription drugs, were more likely to report subjective memory complaints, to report having been diagnosed with heart disease, thyroid disease, arthritis, hypertension, to have higher systolic and diastolic blood pressures, to have hearing impairment; they were also *less* likely to be married, to currently consume alcohol, to use computers, to exercise, to be physically active, and to say they received sufficient help from others. In addition, among those aged 85–89y, those at increased risk of incident dementia tended to have more depression symptoms, to be less likely to smoke currently, and less likely to state that they had any confidantes. The results after stratification by gender, education and *APOE\*4* carrier status remained consistent with the unstratified results (data not shown).



## Discussion

In this population-based cohort followed over 10 years, we had previously reported that a traditional hypothesis-based regression model failed to identify any significant associations between putative risk/protective factors and incident dementia with onset after the median onset age of 87y. We included as “predictors” both variables which could be independent risk factors, such as demographics, blood pressure, and *APOE\*4* genotype, and also variables which could be subclinical signs or epiphenomena of dementia, such as subjective cognitive concerns and functional limitations. The machine learning method is atheoretical and does not distinguish between variables based on their assumed relationship with the outcome; that distinction takes place during interpretation results.

Using Markov modeling with the HyDaP clustering machine learning method, among those aged 85+y, we found characteristics suggesting a “sicker” profile, consistent with risk factors described for dementia at younger-old ages. In addition, only in the subgroup aged 85–89y, incident dementia was also predicted by depression symptoms, not smoking, social isolation, and the *APOE\*4* genotype.

Many previous studies have used Cox proportional hazards regression models to examine the association between potential risk factors and time to onset of dementia using only baseline characteristics. To incorporate longitudinal data on covariates while taking attrition into account, one can use joint modeling including longitudinal and survival submodels. However, in most studies the absolute number of incident dementia cases with onset in the ninth and tenth decades is not large and thus do not provide sufficient power to detect significant effects in joint modeling analyses. An alternative approach, which we have previously used<sup>16</sup>, is to model slope of cognitive decline rather than incident cases, but that approach asks a different question and could have different results. Our Markov modeling with HyDaP, being data-driven and not involving hypothesis testing, does not face issues of statistical significance or power. It can first partition subjects into different subgroups (clusters) and then model the observations collected at successive assessments, studying their longitudinal behavior while simultaneously accounting for competing risks of attrition. As a result, these new methods allow us to identify subgroups with higher risk of transitioning to dementia.

Previous studies using regression models have also been unable to identify risk factors for dementia among “oldest old” individuals, whether defined as age 80+, 85+ or 90+ years. The European DESCRIPA (Development of screening guidelines and criteria for predementia Alzheimer’s disease)<sup>17</sup> study found that the LIBRA (Lifestyle for Brain Health) Index (based solely on modifiable risk factors: depression, diabetes, physical activity, hypertension, obesity, smoking, hypercholesterolemia, coronary heart disease, and mild/moderate alcohol use) increased risk of dementia up until age 80, but not beyond<sup>5</sup>. The Dutch IPCI (Integrated Primary Care Information) study found that the risk for dementia for all risk factors decreased with increasing age and was no longer significant in individuals aged 90y. Adjusting for mortality as a competing risk did not change the results<sup>7</sup>. The 90+ Study has contributed much original work to the understanding of dementia in the tenth decade of life. Participating in social activities, using caffeine and Vitamin C appeared to be

associated with reduced risk of dementia<sup>6</sup> while poor physical performance (balance, gait, hand grip) was associated with higher risk<sup>18</sup>.

In the 90+ study, self-reported late-onset hypertension with onset age between 80–89 appeared to reduce the risk of dementia consistent with our previous finding that higher systolic pressures were associated with lower risk of dementia before age 87y<sup>4,19</sup>. However, it is at odds with our current finding of reported hypertension and higher blood pressures being associated with higher probability of transitioning to dementia among those aged 85+y. Our current finding that thyroid disease and arthritis were potential risk factors also appear consistent with a 90+ Study finding that these and other autoimmune diseases were associated with hippocampal sclerosis at autopsy<sup>20</sup>. A novel retrospective neuropathology study from the Oregon Alzheimer Center examined preceding cognitive trajectories in participants with a mean age at death of 93.7y<sup>21</sup>. These authors found that memory trajectory was associated with *APOE\*4*, verbal fluency trajectory with gross infarcts, and MMSE trajectory with *APOE\*4*, gross infarcts, Braak scores, moderate neuritic plaques, and hippocampal sclerosis.

The lack of observable independent risk factors for the clinical dementia syndrome in the oldest old can be attributed to the increasing dissociation, with aging, between neuropathology and cognition/ clinical symptoms<sup>22,23</sup>. Further, since neuropathology in the oldest individuals has been shown to be mixed, including cerebrovascular disease, Alzheimer-type pathology, hippocampal sclerosis, and TDP-43 proteinopathy<sup>24</sup>, it would be very difficult to find common risk factors that predicted all of these pathological entities.

Our findings should be interpreted in the context of the strengths and limitations of our study design. The MYHAT cohort was relatively large in size at inception and provided longitudinal follow-up over ten years; however, during this period, attrition reduced cohort size by 77%. The population-based approach to recruitment enhances the external validity (generalizability) of the study results in general but reduces the extent that in-depth measurements (such as imaging and fluid biomarkers) can be collected as they would be in a research setting such as an Alzheimer Disease Research Center<sup>25</sup>. The novel machine learning tool that we used allowed all available data on each participant to be used and was thus able to detect factors associated with subsequent dementia that we were previously unable to detect by traditional regression methods. It was not restricted in scope by *a priori* hypotheses, and thus not constrained by our current imperfect understanding of risk factors; for the same reason, some of the factors it revealed as associated with incident dementias, such as subjective memory complaints, could be consequences or epiphenomena of, rather than independent risk factors for, dementia.

We note that our statistical approach requires either that missing data be handled via imputation prior to applying the HyDaP algorithm, or that participants with incomplete data points be excluded. Further, while this was not a problem in our study, Markov assumptions dictate that the model can only handle equally spaced follow-up assessments with the transition probability from one state to another being fixed over time. If this assumption is not met, it becomes necessary to partition the entire study period into subintervals and estimate Markov transition probabilities within each subinterval. Finally, we emphasize that



this machine learning approach does not generate effect sizes and levels of statistical significance for each identified predictor, or a measure of overall model fit. We do not have sufficient incident dementia cases with onset after age 85 to allow our sample to be split into a training sample and a validation sample; We attempted a validation study using a dataset from a population study performed in the 1990s in an adjacent community, the MoVIES project<sup>26</sup> of individuals born up to 40 years earlier than MYHAT participants. We obtained good validation for study participants aged 85–89 but not for those aged 90 and older. The two studies were similar but not identical in aims and methods (e.g. frequency of follow up) and not all the same variables were collected. Thus, the validation was expected to be unsatisfied. We look forward to other groups validating our findings by replication in their samples.

In summary, in a large population-based longitudinal cohort followed annually for ten years, we have used a novel machine learning method based on Markov modeling to identify potential predictors for dementia in individuals aged 85+y. Overall, they suggest poor health, inactivity, and social isolation as predictors; if independently replicated, this approach might help identify high-risk groups for closer monitoring and early intervention. In addition, the results demonstrate the value of a method that can detect associations within a single cohort that could not be detected with traditional regression approaches. Most population-based studies are not large enough, suffer natural attrition over time, and do not have sufficient incident cases in the oldest age groups to power traditional hypothesis-driven modeling approaches. Thus, non-traditional atheoretical approaches with appropriate external validation provide an alternative way to generate testable hypotheses about rare events such as incident dementia in the oldest old.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgement

YJ performed the analysis and drafted the manuscript. CCC conceptualized the analysis. EJ contributed to the data acquisition. MG designed, administered the study and drafted the manuscript. All authors critically revised the manuscript.

The authors are grateful to the MYHAT study participants and to all MYHAT staff for their time and efforts.

### Funding

This work was supported in part by research grant R01 AG023651 from the National Institute on Aging, US Department of Health and Human Services.

## References

1. Fratiglioni L, Viitanen M, von Strauss E, et al. Very old women at highest risk of dementia and Alzheimer's disease: incidence data from the Kungsholmen Project, Stockholm. *Neurology*. 1997;48(1):132–138. [PubMed: 9008508]
2. Rocca WA, Cha RH, Waring SC, et al. Incidence of dementia and Alzheimer's disease: a reanalysis of data from Rochester, Minnesota, 1975–1984. *American Journal of Epidemiology*. 1998;148(1):51–62. [PubMed: 9663404]

3. Corrada MM, Brookmeyer R, Paganini-Hill A, et al. Dementia incidence continues to increase with age in the oldest old: the 90+ study. *Annals of neurology*. 2010;67(1):114–121. [PubMed: 20186856]
4. Ganguli M, Lee C-W, Snitz BE, et al. Rates and risk factors for progression to incident dementia vary by age in a population cohort. *Neurology*. 2015;84(1):72–80. [PubMed: 25471390]
5. Vos SJ, Van Boxtel MP, Schiepers OJ, et al. Modifiable risk factors for prevention of dementia in midlife, late life and the oldest-old: validation of the LIBRA index. *Journal of Alzheimer's Disease*. 2017;58(2):537–547.
6. Paganini-Hill A, Kawas CH, Corrada MM. Lifestyle factors and dementia in the oldest-old: the 90+ study. *Alzheimer disease and associated disorders*. 2016;30(1):21. [PubMed: 25710250]
7. Legdeur N, van der Lee SJ, de Wilde M, et al. The association of vascular disorders with incident dementia in different age groups. *Alzheimer's research & therapy*. 2019;11(1):47.
8. Ganguli M, Snitz B, Bilt JV, et al. How much do depressive symptoms affect cognition at the population level? The Monongahela–Youghiogheny Healthy Aging Team (MYHAT) study. *International journal of geriatric psychiatry*. 2009;24(11):1277–1284. [PubMed: 19340894]
9. Folstein MF, Folstein SE, McHugh PR. “Mini-mental state”: a practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research*. 1975;12(3):189–198. [PubMed: 1202204]
10. Mungas D, Marshall S, Weldon M, et al. Age and education correction of Mini-Mental State Examination for English-and Spanish-speaking elderly. *Neurology*. 1996;46(3):700–706. [PubMed: 8618670]
11. Ganguli M, Vander Bilt J, Lee C-W, et al. Cognitive test performance predicts change in functional status at the population level: The MYHAT Project. *Journal of the International Neuropsychological Society*. 2010;16(5):761–770. [PubMed: 20609270]
12. Morris J The clinical dementia rating (cdr): Current version and scoring rules. *Neurology*. 1993;43(11):2412–2414.
13. Ganguli M, Chang C-CH, Snitz BE, et al. Prevalence of mild cognitive impairment by multiple classifications: The Monongahela-Youghiogheny Healthy Aging Team (MYHAT) project. *The American Journal of Geriatric Psychiatry*. 2010;18(8):674–683. [PubMed: 20220597]
14. Ganguli M, Jia Y, Hughes TF, et al. Mild cognitive impairment that does not progress to dementia: a population-based study. *Journal of the American Geriatrics Society*. 2019;67(2):232–238. [PubMed: 30444944]
15. Wang S, Yabes JG, Chang C-CH. Hybrid Density-and Partition-based Clustering Algorithm for Data with Mixed-type Variables. arXiv preprint arXiv:190502257. 2019.
16. Ganguli M, Fu B, Snitz BE, et al. Vascular risk factors and cognitive decline in a population sample. *Alzheimer disease and associated disorders*. 2014;28(1):9. [PubMed: 24126216]
17. Visser P, Verhey F, Boada M, et al. Development of screening guidelines and clinical criteria for predementia Alzheimer's disease. *Neuroepidemiology*. 2008;30(4):254–265. [PubMed: 18515975]
18. Bullain SS, Corrada MM, Perry SM, et al. Sound Body Sound Mind? Physical Performance and the Risk of Dementia in the Oldest-Old: The 90+ Study. *Journal of the American Geriatrics Society*. 2016;64(7):1408–1415. [PubMed: 27377238]
19. Corrada MM, Hayden KM, Paganini-Hill A, et al. Age of onset of hypertension and risk of dementia in the oldest-old: The 90+ Study. *Alzheimer's & Dementia* 2017;13(2):103–110.
20. Trieu T, Sajjadi SA, Kawas CH, et al. Risk factors of hippocampal sclerosis in the oldest old: the 90+ study. *Neurology*. 2018;91(19):e1788–e1798. [PubMed: 30315072]
21. Nguyen MT, Mattek N, Woltjer R, et al. Pathologies Underlying Longitudinal Cognitive Decline in the Oldest Old. *Alzheimer Disease & Associated Disorders*. 2018;32(4):265–269. [PubMed: 30052535]
22. M Corrada MJ Berlau D, H Kawas C. A population-based clinicopathological study in the oldest-old: the 90+ study. *Current Alzheimer Research*. 2012;9(6):709–717. [PubMed: 22471863]
23. James BD, Bennett DA, Boyle PA, et al. Dementia from Alzheimer disease and mixed pathologies in the oldest old. *Jama*. 2012;307(17):1798–1800. [PubMed: 22550192]
24. Kawas CH, Kim RC, Sonnen JA, et al. Multiple pathologies are common and related to dementia in the oldest-old: The 90+ Study. *Neurology*. 2015;85(6):535–542. [PubMed: 26180144]

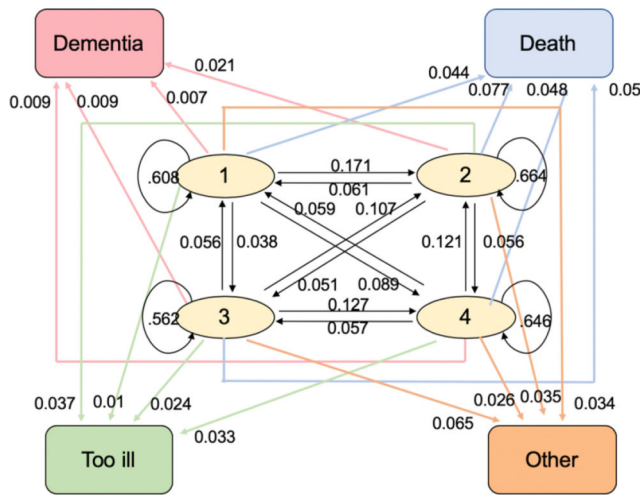
25. Kukull WA, Ganguli M. Generalizability: the trees, the forest, and the low-hanging fruit. *Neurology*. 2012;78(23):1886–1891. [PubMed: 22665145]
26. Ganguli M, Dodge H, Chen P, et al. Ten-year incidence of dementia in a rural elderly US community population: the MoVIES Project. *Neurology*. 2000;54(5):1109–1116. [PubMed: 10720283]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

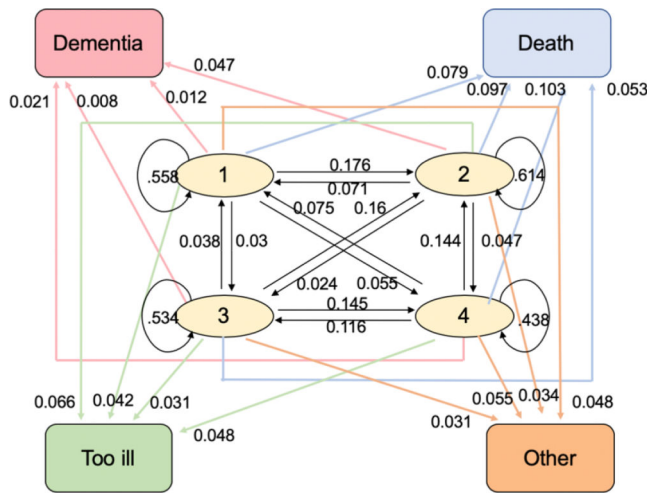


		Cycle (t+1)							
Cycle t	State	1	2	3	4	Dementia (CDR $\geq$ 1)	Dropout - Death	Dropout - Too ill	Dropout - Others
	1	0.608	0.171	0.038	0.089	0.007	0.044	0.01	0.034
2	0.061	0.664	0.051	0.056	0.021	0.077	0.037	0.035	
3	0.056	0.107	0.562	0.127	0.009	0.05	0.024	0.065	
4	0.059	0.121	0.057	0.646	0.009	0.048	0.033	0.026	

**Figure 1.**

Transition Probability Matrix and Plot for Aged 85–89

Note: circles represent the four clusters (transient states) and squares represent the four absorbing states. Arrows indicate the direction of transition and the numbers on the arrow show the corresponding transition probability.



		Cycle (t+1)								
		State	1	2	3	4	Dementia (CDR ≥ 1)	Dropout - Death	Dropout - Too ill	Dropout - Others
Cycle t	1	.558	0.176	0.03	0.055	0.012	0.079	0.042	0.048	
	2	0.071	0.614	0.024	0.047	0.047	0.097	0.066	0.034	
	3	0.038	0.16	0.534	0.145	0.008	0.053	0.031	0.031	
	4	0.075	0.144	0.116	0.438	0.021	0.103	0.048	0.055	

**Figure 2.**  
 Transition Probability Matrix and Plot for Aged 90+  
 Note: circles represent the four clusters (transient states) and squares represent the four absorbing states. Arrows indicate the direction of transition and the numbers on the arrow show the corresponding transition probability.

**Table 1.**

Proportion of incident dementia cases in each stratum.

	Women	Men	<i>P</i>	w/o <i>APOE</i> *4	w/ <i>APOE</i> *4	<i>P</i>	HS	>HS	<i>P</i>
<b>85–89</b>									
# of dementia	13	8	0.928	13	8	0.032	11	10	0.294
Total	1137	728		1499	366		1182	683	
Percentage	0.011	0.011		0.009	0.022		0.009	0.015	
<b>90+</b>									
# of dementia	18	6	0.303	21	3	0.992	21	3	0.075
Total	576	309		775	110		631	254	
Percentage	0.031	0.019		0.027	0.027		0.033	0.012	

Abbreviation: HS = high school graduate

\**P*-values were derived from a two-sample proportional Z test; alpha-level = 0.025 with Bonferroni correction for 2 comparisons

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Table 2.**

Characteristics of clusters that more likely to progress to dementia versus clusters that less likely to progress to dementia among aged 85–89

	Less likely to progress to dementia	More likely to progress to dementia	<i>P</i> *
N	1157 **	708 **	
<b>SELF-REPORTS OF DIAGNOSES BY HEALTH CARE PROFESSIONALS EVER OR DURING PRECEDING YEAR</b>			
Cardiovascular disease, n (%)	86 ( 7.4)	76 (10.7)	0.018
Irregular heartbeat, n (%)	301 (26.0)	217 (30.6)	0.034
Cerebrovascular disease, n (%)	32 ( 2.8)	27 ( 3.8)	0.263
Hypertension, n (%)	745 (64.4)	545 (77.0)	<0.001
Diabetes, n (%)	222 (19.2)	154 (21.8)	0.201
Thyroid disease, n (%)	194 (16.8)	192 (27.1)	<0.001
Arthritis, n (%)	770 (66.6)	589 (83.2)	<0.001
Rheumatoid arthritis, n (%)	20 ( 1.7)	19 ( 2.7)	0.218
<b>ASSESSMENTS DURING STUDY VISITS</b>			
Depression (mCESD >3), n (%)	49 ( 4.2)	68 ( 9.6)	<0.001
# of prescription medication >3, n (%)	651 (56.3)	498 (70.3)	<0.001
Systolic blood pressure (mean (SD))	131.50 (15.51)	134.31 (16.67)	<0.001
Diastolic blood pressure (mean (SD))	71.10 (8.28)	72.05 (9.06)	0.02
Waist:Hip ratio (mean (SD))	0.90 (0.09)	0.89 (0.09)	0.007
Hearing impairment (%)			0.012
None	734 (63.4)	401 (56.6)	
Right ear didn't hear	85 ( 7.3)	60 ( 8.5)	
Left ear didn't hear	126 (10.9)	77 (10.9)	
Bilateral didn't hear	212 (18.3)	170 (24.0)	
<b>SELF-RATINGS BY PARTICIPANTS DURING STUDY VISITS</b>			
Self-rated health, n (%)			<0.001
Poor	14 ( 1.2)	21 ( 3.0)	
Fair	121 (10.5)	149 (21.0)	
Good	569 (49.2)	406 (57.3)	
Very good	327 (28.3)	108 (15.3)	
Excellent	126 (10.9)	24 ( 3.4)	
Subjective memory complaints >0, n (%)	729 (63.0)	545 (77.0)	<0.001
<i>Hearing loss</i>			
Using hearing aid, n (%)	380 (32.8)	195 (27.5)	0.019
Hear Conversation, n (%)			0.479
No	11 ( 1.0)	4 ( 0.6)	
Yes, without hearing aid	1003 (86.7)	625 (88.3)	
Yes ONLY with hearing aid	143 (12.4)	79 (11.2)	
<i>Sleep complaints</i>			
Difficulty falling asleep, n (%)	392 (33.9)	197 (27.8)	0.007
Difficulty back asleep, n (%)	423 (36.6)	372 (52.5)	<0.001

	<b>Less likely to progress to dementia</b>	<b>More likely to progress to dementia</b>	
Wake up too early, n (%)	319 (27.6)	156 (22.0)	0.009
Daytime sleepiness, n (%)	471 (40.7)	492 (69.5)	<0.001
Snoring, n (%)	336 (29.0)	171 (24.2)	0.025
<i>Lifestyle</i>			
Current drinking, n (%)	765 (66.1)	165 (23.3)	<0.001
Current smoking, n (%)	38 ( 3.3)	10 ( 1.4)	0.02
Exercise, n (%)	822 (71.0)	247 (34.9)	<0.001
Physical activity, n (%)	906 (78.3)	220 (31.1)	<0.001
Using computer, n (%)	425 (36.7)	87 (12.3)	<0.001
<i>Social connectedness/isolation</i>			
Married, n (%)	539 (46.6)	129 (18.2)	<0.001
Confidantes, n (%)	1148 (99.2)	698 (98.6)	0.277
Receive Sufficient help, n (%)	1101 (95.2)	624 (88.1)	<0.001

Acronyms: mCESD – modified Center for Epidemiologic Studies Depression Scale

\* *P*-values for continuous variables were derived from ANOVA; *p*-values for categorical variables were derived from chi-square test.

\*\* From Figure 1, cluster 2 has the highest transition probability to incident dementia, thus we compare cluster 2 to clusters 1,3, and 4 combined.

**Table 3.**

Characteristics of clusters that more likely to progress to dementia versus clusters that less likely to progress to dementia among aged 90+

	Less likely to progress to dementia	More likely to progress to dementia	<i>P</i> *
N	441**	444**	
<b>SELF-REPORTS OF DIAGNOSES BY HEALTH CARE PROFESSIONALS EVER OR DURING PRECEDING YEAR</b>			
Cardiovascular disease, n (%)	43 (9.8)	51 (11.5)	0.466
Irregular heartbeat, n (%)	105 (23.8)	152 (34.2)	0.001
Cerebrovascular disease, n (%)	10 (2.3)	11 (2.5)	0.999
Hypertension, n (%)	300 (68.0)	370 (83.3)	<0.001
Diabetes, n (%)	100 (22.7)	114 (25.7)	0.335
Thyroid disease, n (%)	75 (17.0)	91 (20.5)	0.214
Arthritis, n (%)	299 (67.8)	380 (85.6)	<0.001
Rheumatoid arthritis, n (%)	4 (0.9)	4 (0.9)	0.999
<b>ASSESSMENTS DURING STUDY VISITS</b>			
Depression (mCESD >3), n (%)	238 (54.0)	290 (65.3)	0.001
# of prescription medication >3, n (%)	296 (67.1)	353 (79.5)	<0.001
Systolic blood pressure (mean (SD))	130.61 (16.63)	134.41 (16.94)	0.001
Diastolic blood pressure (mean (SD))	69.10 (8.30)	70.30 (8.64)	0.036
Waist:Hip ratio (mean (SD))	0.90 (0.08)	0.89 (0.08)	0.038
Hearing impairment (%)			<0.001
None	236 (53.5)	190 (42.8)	
Right ear didn't hear	41 (9.3)	49 (11.0)	
Left ear didn't hear	66 (15.0)	50 (11.3)	
Bilateral didn't hear	98 (22.2)	155 (34.9)	
<b>SELF-RATINGS BY PARTICIPANTS DURING STUDY VISITS</b>			
Self-rated health, n (%)			<0.001
Poor	4 (0.9)	11 (2.5)	
Fair	44 (10.0)	86 (19.4)	
Good	203 (46.0)	252 (56.8)	
Very good	134 (30.4)	72 (16.2)	
Excellent	56 (12.7)	23 (5.2)	
Subjective memory complaints >0, n (%)	296 (67.1)	353 (79.5)	<0.001
<i>Hearing loss</i>			
Using hearing aid, n (%)	203 (46.0)	145 (32.7)	<0.001
Hear Conversation, n (%)			0.001
No	10 (2.3)	25 (5.6)	
Yes, without hearing aid	329 (74.6)	352 (79.3)	
Yes ONLY with hearing aid	102 (23.1)	67 (15.1)	
<i>Sleep complaints</i>			
Difficulty falling asleep, n (%)	194 (44.0)	145 (32.7)	0.001
Difficulty back asleep, n (%)	173 (39.2)	231 (52.0)	<0.001

Wake up too early, n (%)	133 (30.2)	120 ( 27.0)	0.339
Daytime sleepiness, n (%)	197 (44.7)	297 ( 66.9)	<0.001
Snoring, n (%)	117 (26.5)	85 ( 19.1)	0.011
<i>Lifestyle</i>			
Current drinking, n (%)	252 (57.1)	84 ( 18.9)	<0.001
Current smoking, n (%)	8 ( 1.8)	12 ( 2.7)	0.507
Exercise, n (%)	284 (64.4)	142 ( 32.0)	<0.001
Physical activity, n (%)	324 (73.5)	146 ( 32.9)	<0.001
Using computer, n (%)	133 (30.2)	38 ( 8.6)	<0.001
<i>Social connectedness/isolation</i>			
Married, n (%)	150 (34.0)	52 ( 11.7)	<0.001
Confidantes, n (%)	427 (96.8)	444 (100.0)	<0.001
Receive Sufficient help, n (%)	402 (91.2)	409 ( 92.1)	0.693

Acronyms: mCESD – modified Center for Epidemiologic Studies Depression Scale

\* *P*-values for continuous variables were derived from ANOVA; *p*-values for categorical variables were derived from chi-square test.

\*\* From Figure 2, cluster 2 has the highest transition probability to incident dementia, thus we compare cluster 2 to clusters 1,3, and 4 combined.