Title: Identifying and typing latent neuropathology using pattern recognition on longitudinal data

Keywords: neuroepidemiology, dementia subtypes; Alzheimer's disease and related dementias; vascular dementia; diagnosis

Authors:

Sean Clouston*, PhD, Program in Public Health and Department of Family, Population, and Preventive Medicine, Stony Brook University, Stony Brook, NY, USA

Yun Zhang, MMed, Program in Public Health, Stony Brook University, Stony Brook, NY, USA

***Correspondence to**: Program in Public Health, 101 Nichols Rd., Health Sciences Center, #3-071, Stony Brook University, Stony Brook, NY. Telephone: (631) 444-6593. Email: sean.clouston@stonybrookmedicine.edu. Facsimile: (631) 444-3480.

Acknowledgements: The HRS (Health and Retirement Study) is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan. The authors would like to acknowledge the *Integrative Analysis of Longitudinal Studies on Aging* (NIH/NIA P01 AG043362) network of studies for supporting the development of this work, and to the *National Institute on Aging* for supporting this analysis in particular (NIH/NIA R01 AG058595).

Abstract: Dementia affected 16 million people and caused 110,000 deaths in the U.S. Dementias result from neuropathology and cause decrements to episodic memory. Alzheimer's disease and related dementias (ADRD) or by cerebro-vascular and ischemic disease (VaID) cause most dementias, but these are unreliably identified in population data since diagnoses are difficult and expensive. The current study used pattern recognition on longitudinal trajectories of cognitive decline to identify latent, often unreported, cases of VaID and ADRD. The resulting method was reliably able to both identify and differentiate between these two causes of disease. Incidence of VaID and ADRD identified in this way was 40.23/1,000 (95% CI = [39.40-41.08]) and 27.63/1,000 (95% CI = [26.97-28.29]) person-years respectively and was highly concordant with reported diagnoses. Age was significantly associated with higher incidence of both. This was the first study to use pattern recognition to identify and differentially identify latent neuropathology.

Dementia affects 16 million people in the United States[1], and was recorded as the underlying cause of death in as many as 110,000 deaths in 2017 [2]. More than 99% of all dementias fall into two main types: Alzheimer's disease and related dementias (ADRD) accounting for approximately 60-80% of all cases [3]; and cerebro-vascular and ischemic disease (VaID), which accounts for most of the remaining dementia cases [4]. ADRD and VaID are both preceded and by different types of changes to cognitive functioning, making differential diagnosis difficult. However, ADRD is characterized by progressive losses of capability across domains of "fluid" cognition including episodic memory [5] (**Fig. 1**). The end result is years of "accelerated declines" in cognitive function resulting in progression through milder forms of cognitive impairment to clinical limitations and dementia [6]. In contrast, VaID is characterized by a rapid loss in functioning that may affect similar domains of cognition, but which cause losses over a very short period of time before stabilizing. Recent work has determined that patterns in ADRD might be usefully diagnosed using patterns in the rate of decline [7], this information has not been utilized to identify possible VaID. The objective of the current study was to examine the extent to which ADRD and VaID might be usefully identified using temporal patterns evident in longitudinal data.

Methods

We used waves 1-12 of the *Health and Retirement Study* (HRS), which collects cognitive data biennially starting in 1992 (response rate 81.6%). The HRS is open to enrollment at subsequent waves, and data are publicly available online (http://hrsonline.isr.umich.edu) [8]. Because of the intense analytic requirements for differential diagnostic routines, respondents without at least 4 waves of data were

excluded. The analytic sample therefore included 12,849 respondents who were observed a total of 111,349 times for up to 20 years per person (Tab.1).

Measures

Episodic memory is a critical measure of cognitive functioning that is both sensitive to cognitive aging and AD [9]. To measure episodic memory, respondents were first provided with a list of 10 words and asked to correctly recall to their best ability with each correct one scoring one point. After 10-15 minutes of intermediate distraction questions, respondents were asked again to repeat all 10 words correctly to the interviewer to score the correct answers. The Total Episode Memory Index included the summation of both immediate and delayed verbal recall tests (/20 points). Because the first two waves utilized a 20-item word list, total scores for the two waves were divided by two to match later assessment procedures.

Inferential Diagnoses

Diagnostic categories were defined using a pattern recognition algorithm that was applied to each respondent in the database [10]. The goal was to identify individuals whose longitudinal pattern of decline was similar to the diagnostic features shown in Figure 1. To accomplish this, a program searched through each individual's data and determined whether each person's data best fit the profile of 1) a linear pattern of change, 2) a piece-wise linear accelerated pattern consistent with ADRD, or 3) a stepwise-linear pattern consistent with VaID. Since the pattern recognition program was particularly sensitive to random variation in the first or last waves, onsets occurring before the second or after the penultimate wave were ignored. Patient group and the best-fitting date of onset were recorded. In a small number of cases, the computer was unable to determine whether

patterns were best fit by a VaID or ADRD patterns of decline, necessitating the creation of an "indeterminate" category.

Respondent sex and date of birth were recorded. Since respondents with cognitive issues sometimes misreport current age, age in years was calculated using date of interview and date of birth. Year of birth was incorporated into longitudinal modeling efforts to account for baseline differences in functioning [11]. Since there is a common reduction in cognitive function in the first time-point due to unfamiliarity with testing circumstances, a dichotomous indicator was incorporated that identified the first wave of assessment in both longitudinal modeling and pattern recognition analyses. Age in months was transformed into years.

Validation

In these data, diagnoses are available from self-reports. Self-reported stroke was recorded is a valid and reliable (AUC=0.99) method for identifying major stroke [12]. Additionally, self-reported Alzheimer's diagnoses were also available in waves 10-12, and were used to differentiate reported ADRD from stroke-related diseases in these data. Alzheimer's diagnoses are notoriously bad, since relying on individuals with memory disorders to report diagnoses has low face validity, and because ADRD diagnoses are undercounted.

Statistical Analyses

Means and standard deviations, and percentages were used to describe the sample. Crude incidence rates were age-stratified. Crude incidence rates (IR) as well as age-standardized incidence rates (aIR) were provided for the entire sample for both ADRD and VaID pathologies.

Results

Sample descriptive statistics (**Tab. 1**) revealed that respondents were in their sixties, on average, and was majority female and White. Bivariate results revealed that individuals with any indication of pathology were older than those without, and that those with ADRD or indeterminate pathology were older than those with Vascular pathology. On average, half of respondents had no observed pathology over the period of observation.

[Table 1]

Table 2 shows that those who were identified as having a VaID-pattern decline were also much more likely to report having had a stroke (AUC=0.86). The incidence rate for VaID-pattern declines was 40.23/1,000 (95% CI = [39.40-41.08]) person-years. Additionally, VaID-pattern cognitive declines were strongly associated with age (Figure 2; HR = 1.046, 95% CI = [1.043-1.049], P<1E-06).

[Table 2; Figure 2]

Table 3 shows that those who were identified as having ADRD-pattern declines were also much more likely to report having had Alzheimer's disease (AUC=0.68). The incidence rate for ADRD-pattern declines was 27.63/1,000 (95% CI = [26.97-28.29]) person-years. Additionally, ADRD-pattern cognitive declines were strongly associated with age (Figure 3; HR = 1.044, 95% CI = [1.042-1.047], P<1E-06).

[Table 3; Figure 3]

Next, we examined the overlap between reported and detected VaID and ADRD (Table 4). The first thing to note is that very few people reported having Alzheimer's disease who did not also report having had a stroke. Secondly, there were a number of

individuals who reported having a stroke who were determined to have ADRD-pattern declines. Finally, most of those who reported having a both a stroke and Alzheimer's disease were, in our data, more likely to be identified as having ADRD-pattern declines. Critically, of those who did not report on either stroke or Alzheimer's disease were often determined to have Alzheimer's-pattern declines and/or strokes.

[Table 4]

Discussion

This study used a pattern recognition algorithm to identify disease-specific shapes of cognitive decline as an objective metric indicative of the incidence of ADRD and VaID. This is the first study to examine the shape of longitudinal trajectories in order to identify and to further differentiate subtypes of dementia-causing neuropathology. This is also, to the authors' knowledge, the first study to utilize pattern recognition on longitudinal epidemiologic data to identify any latent disease process. The methods used were able to reliably identify both ADRD and VaID-pattern declines in a sample of U.S. residents aged 50 and older with at least five observations. In this sample, this method was able to detect self-reported major strokes with high reliability (AUC=0.864, 95% CI = [0.859-0.869]) and ADRD with moderate reliability (AUC=0.677, 95% CI = [0.657-0.696]). Further work is warranted to refine identification methods, and to further determine risk factors for ADRD and cerebrovascular-pattern cognitive declines.

The pattern recognition method used in this study identified a large number of individuals with ADRD-pattern cognitive declines who did not report having a diagnosed dementia. This study expected to find many more ADRD-pattern declines than might be reported by respondents. ADRD is drastically under-diagnosed in population, with some

estimates suggesting that as many as 50-80% patients with ADRD-related symptoms and neuropathology never receive a diagnosis [13]. Additionally, this study identified ADRD-patterns that may be still in preclinical phases since clinical diagnoses of dementia rely not on poor or degrading memory but rather on evidence of cognitive impairment accompanied by extant limitations in the ability to navigate social space [14]. Nevertheless, further work seeking to validate these methods in identifying risk of latent ADRD neuropathology is critical in future work.

VaID is caused by ischemic lesions ranging in size and impacts; however, the prevalence of these lesions and the type, location, and size needed to cause cognitive damage or impairment is unclear[15]. It is a huge challenge to make clinical and population diagnosis, as golden standard diagnosis relies on post-mortem evidence [16]. Lacking evidence from neuroimaging, individuals or their representative are asked to determine the timing of symptoms in order to make a diagnosis [17]. As a result, only major strokes are reliably diagnosed, most small strokes as well as many moderately large ones are never found. Since this process is difficult to disentangle from ADRD, most studies do not make a clear distinction between causes [18]. This can result in lack of sufficient evidence for risk factors that are unique to any particular disease [19]. Although recognized as a common cause of dementia in population-based cohorts [20], the lack of clear criteria and the low likelihood of detection make prevalence and incidence estimates from nationally representative studies suspect.

Early dementia diagnoses create an opportunity to lower patient risks, prevent complications, and also help individuals to better prepare for their future [21]. Thus, while there is currently no treatment, there is a need to develop methods to reliably identify

dementia and its subtypes[22]. No studies have yet been found in clinical efforts sought to examine biomarkers[23], neural characteristics[24] or neuropsychological screening methods[25], let alone population screening. Years of preclinical cognitive decline in longitudinal data fits dementia pathology patterns and in turn imply possibilities to utilize population-based data mining techniques to make a diagnose. Distinct from models replying on neuropsychological test cutoffs and using random slopes methods, we emphasized patterns of within-person change over time.

Strengths & Limitations

This study would be one of the earliest to provide modeling that is specifically focused on modeling both healthy aging and decline patterns and to differentiate subtypes of cognitive impairment. It associates longitudinal data closely with clinical pathology theories to seek the prevalence and distribution of the dementia and subtypes for the first time. Our method also helps estimate ADRD prevalence while avoiding an existing bias that relies on cutoffs which includes people who are consistently relatively low cognitive or in development of ADRD but not actually. A bias our model avoids is linear assumptions as within-person random slopes models made in precious studies. It might show a more convenient and less expensive way of limited clinical utility like cognitive neuropsychology or magnetic resonance imaging results to identify individuals with dementia or predict risk of dementia. While neuroimaging is increasingly feasible, it has not been reliably used to determine prevalence of the disease and remains a method for validation of findings rather than for determining the extent of disease in a population. We found more stroke cases than were otherwise reported, and it is therefore critical that

future efforts seek to determine to what extent strokes identified in these types of data are corroborated by neuroimaging markers.

Our method is designed to dementia patterns identification rather than accurate dementia diagnosis, so it is impossible to differentiate between different etiologies of disease. This study was limited in relying on self-reported diagnoses, undergoing dementia or stroke. Although "fluid" cognition is recognized as the main injury domains, stroke cases with slow or small losses to episodic memory, or concentrated on other domains of cognitive functioning, may also be missed. Finally, this study is also limited by the reliance on datasets with at least five follow ups occurring over a relative long period, implying that the number of databases able to use this type of method may be limited.

Conclusions

Diagnosing dementia and stroke is expensive and difficult, resulting in a large burden of disease that remains undiagnosed [26]. This is problematic for research since those lacking diagnoses are likely to differ in a number of ways, including by level of access to healthcare or by level of connection with family members or friends needed to identify symptoms. This study sought to inferentially identify the incidence of ADRD and cerebrovascular diseases pathology by utilizing longitudinal decline patterns. The results are promising, but suggest that more work is warranted replicating, extending, and validating the current model. Given the potential importance of understanding timing of declines, future research is warranted to understand both healthy and pathological forms of cognitive aging. This study contributes to attempts to use simple cognitive variables to predicting subjects with a higher risk of developing dementia in the absence of

Clinical tests among population.

References

1. 2018 Alzheimer's disease facts and figures. *Alzheimer's & Dementia.* 2018;14(3):367-429.
2. Centers for Disease Control and Prevention. *Leading Causes of Death.* Atlanta, GA: Centers for Disease Control and Prevention;2017.
3. Alzheimer's Association. 2016 Alzheimer's disease facts and figures. *Alzheimer's & Dementia.* 2016;12(4):459-509.
4. Gorelick PB, Scuteri A, Black SE, et al. Vascular Contributions to Cognitive Impairment and Dementia: A Statement for Healthcare Professionals From the American Heart Association/American Stroke Association. *Stroke.* 2011;42(9):2672-2713.
5. Richards M, Deary IJ. A life course approach to cognitive capability. In: Kuh D, Cooper R, Hardy R, Richards M, Ben-Shlomo Y, eds. *A Life Course Approach to Healthy Ageing.* Oxford: Oxford University Press; 2014:32-45.
6. Bruscoli M, Lovestone S. Is MCI really just early dementia? A systematic review of conversion studies. *International Psychogeriatrics.* 2004;16(02):129-140.
7. Clouston S. Determinants of cognitive pathology. American Psychopathological Association; 2015; New York, NY.
8. Health and Retirement Study (HRS). Health and Retirement Study, public use dataset. . In: University of Michigan with funding from the National Institute on Aging (U01 AG 009740), ed. Ann Arbor, MI2014.
9. Baddeley A. Working memory. *Science.* 1992;255(5044):556.
10. Clouston SA, Glymour M, Terrera GM. Educational inequalities in aging-related declines in fluid cognition and the onset of cognitive pathology. *Alzheimers Dement (Amst).* 2015;1(3):303-310.
11. Flynn JR. IQ gains over time: Toward finding the causes. *The rising curve: Long-term gains in IQ and related measures.* 1998:25-66.
12. Okura Y, Urban LH, Mahoney DW, Jacobsen SJ, Rodeheffer RJ. Agreement between self-report questionnaires and medical record data was substantial for diabetes, hypertension, myocardial infarction and stroke but not for heart failure. *Journal of clinical epidemiology.* 2004;57(10):1096-1103.
13. Connolly A, Gaehl E, Martin H, Morris J, Purandare N. Underdiagnosis of dementia in primary care: variations in the observed prevalence and comparisons to the expected prevalence. *Aging & mental health.* 2011;15(8):978-984.
14. McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers & Dementia.* 2011;7(3):263-269.
15. Vitali P, Migliaccio R, Agosta F, Rosen HJ, Geschwind MD. Neuroimaging in Dementia. *Seminars in neurology.* 2008;28(4):467-483.

16. Scheltens P, Blennow K, Breteler MM, et al. Alzheimer's disease. *Lancet.* 2016;388(10043):505-517.
17. Pohjasvaara T, Mantyla R, Ylikoski R, Kaste M, Erkinjuntti T. Comparison of different clinical criteria (DSM-III, ADDTC, ICD-10, NINDS-AIREN, DSM-IV) for the diagnosis of vascular dementia. National Institute of Neurological Disorders and Stroke-Association Internationale pour la Recherche et l'Enseignement en Neurosciences. *Stroke.* 2000;31(12):2952-2957.
18. Korczyn AD, Vakhapova V, Grinberg LT. Vascular dementia. *Journal of the neurological sciences.* 2012;322(1-2):2-10.
19. Jellinger KA. The enigma of vascular cognitive disorder and vascular dementia. *Acta Neuropathologica.* 2007;113(4):349-388.
20. Kalaria RN. The pathology and pathophysiology of vascular dementia. *Neuropharmacology.* 2018;134(Pt B):226-239.
21. Doubal FN, Ali M, Batty GD, et al. Big data and data repurposing - using existing data to answer new questions in vascular dementia research. *BMC Neurology.* 2017;17(1):72.
22. Albert MS, DeKosky ST, Dickson D, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 2011;7(3):270-279.
23. Buchhave P, Minthon L, Zetterberg H, Wallin AK, Blennow K, Hansson O. Cerebrospinal fluid levels of beta-amyloid 1-42, but not of tau, are fully changed already 5 to 10 years before the onset of Alzheimer dementia. *Archives of general psychiatry.* 2012;69(1):98-106.
24. Eikelenboom P, Bate C, Van Gool WA, et al. Neuroinflammation in Alzheimer's disease and prion disease. *Glia.* 2002;40(2):232-239.
25. Nasreddine ZS, Phillips NA, Bedirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc.* 2005;53(4):695-699.
26. Amjad H, Roth DL, Sheehan OC, Lyketsos CG, Wolff JL, Samus QM. Underdiagnosis of Dementia: an Observational Study of Patterns in Diagnosis and Awareness in US Older Adults. *Journal of general internal medicine.* 2018;33(7):1131-1138.

Table 1. Descriptive characteristics at individual baseline of eligible participants for the entire sample and separated by inferential diagnosis

	Whole Sample		No Pathology		Vascular Pattern		Alzheimer's Disease and Related Dementia Pattern		Mixed or Indeterminate Pattern	
Characteristic	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Age in years	61.84	8.73	59.97	7.79	63.42	10.03	63.46	8.69	65.05	8.84
Characteristic	**%**		**%**		**%**		**%**		**%**	
Race/Ethnicity										
White	75.44%		74.63%		75.15%		76.53%		77.76%	
Black	13.90%		14.30%		13.64%		13.53%		13.13%	
Other	2.09%		1.99%		2.33%		2.24%		1.72%	
Hispanic	8.57%		9.09%		8.89%		7.71%		7.39%	
Female sex	57.61%		55.87%		58.47%		59.23%		61.62%	
Group Size	(n=18,102)		(n=11,142 [53.44%])		(n=6,514 [31.24%])		(n=2,901 [13.91%])		(n=293 [1.41%])	

Table 2. Concordance between self-reported stroke and strokes detected using pattern recognition

	Reported Stroke	No Reported Stroke
VaID Pattern	6,045	1,031
No Pattern Evident	556	5,198
AUC (95% CI)	0.864	(0.859-0.869)
RR (95% CI)	7.658	(7.184-8.164)
Sensitivity		0.916
Specificity		0.834
Positive Predictive Value		0.854
Negative Predictive Value		0.903

Note: VaID: Cerebrovascular and ischemic disease.

Table 3. Concordance between self-reported Alzheimer's and detected ADRD-pattern declines

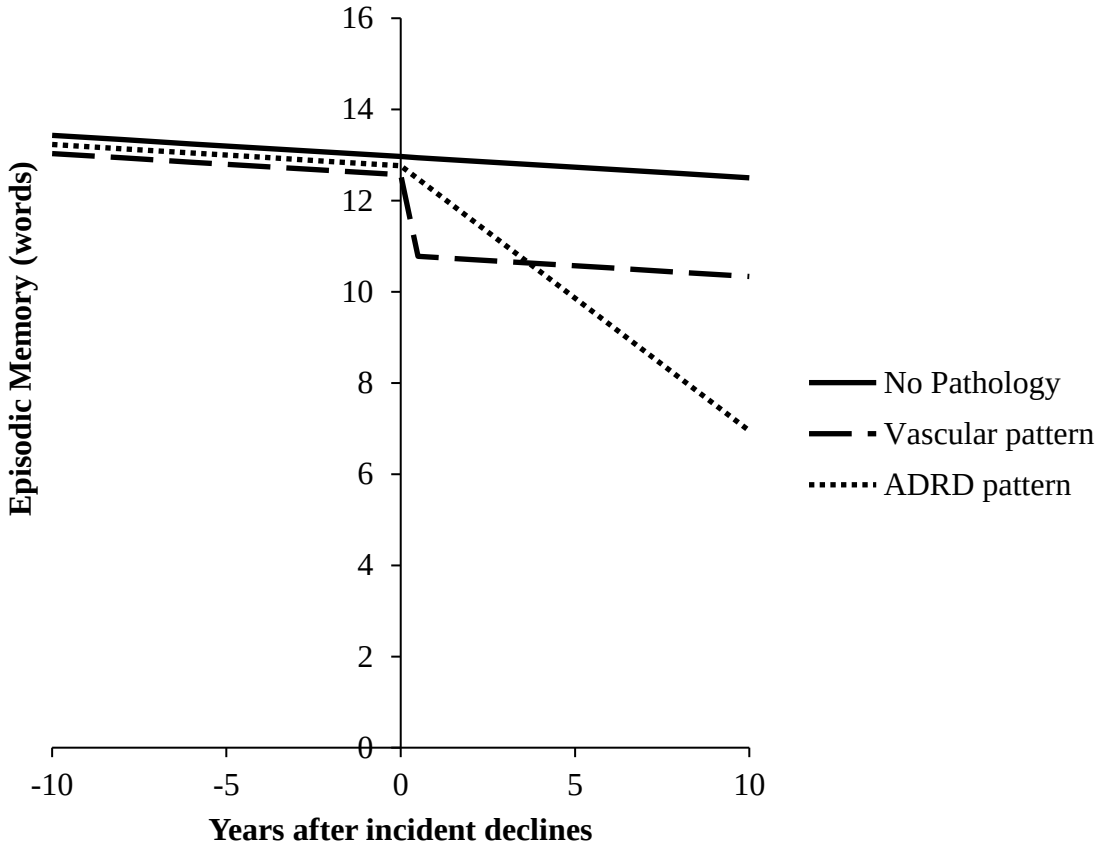
	Reported Alzheimer's	No Reported Alzheimer's
ADRD Pattern	395	4,928
No Pattern Evident	150	8,333
AUC (95% CI)	0.677	(0.657-0.696)
RR (95% CI)	3.845	(3.22-4.592)
Sensitivity		0.725
Specificity		0.628
Positive Predictive Value		0.074
Negative Predictive Value		0.982

Note: ADRD: Alzheimer's disease or a related dementia.

Table 4. Overlap between diagnoses reported and those that were identified using pattern recognition

Reported Stroke	No Pathology	Vascular Disease Pattern	Alzheimer's Disease and Related Dementia Patterning	Mixed or Indeterminate	Total
None reported	4,329	2	1,534	0	5,865
Stroke reported	1,742	2,123	1,632	1,005	6,502
Alzheimer's disease reported	33	0	25	0	58
Both reported	33	115	204	72	424
Missing report	2,285	854	1,798	317	5,254
Total (without missing)	8,422	2,240	3,395	1,077	12,849

Figure 1. Differential pathological characterization based on aging pattern in episodic memory



Note: ADRD: Alzheimer's disease or a related dementia. VaID: Cerebrovascular and ischemic disease. Declines occurred at a rate of -0.047 (95% CI = [-0.044, -0.049]) among those without pathology, but decayed faster for those with ADRD pattern declines (B = -0.535 (95% CI = [-0.525, -0.545]). The impact of strokes was relatively large, with those experiencing VaID-pattern declines showing a deficit of -1.764 words (95% CI = [-1.721, -1.807]) on average afterwards.

Figure 2. Age-specific incidence of vascular-pattern cognitive declines

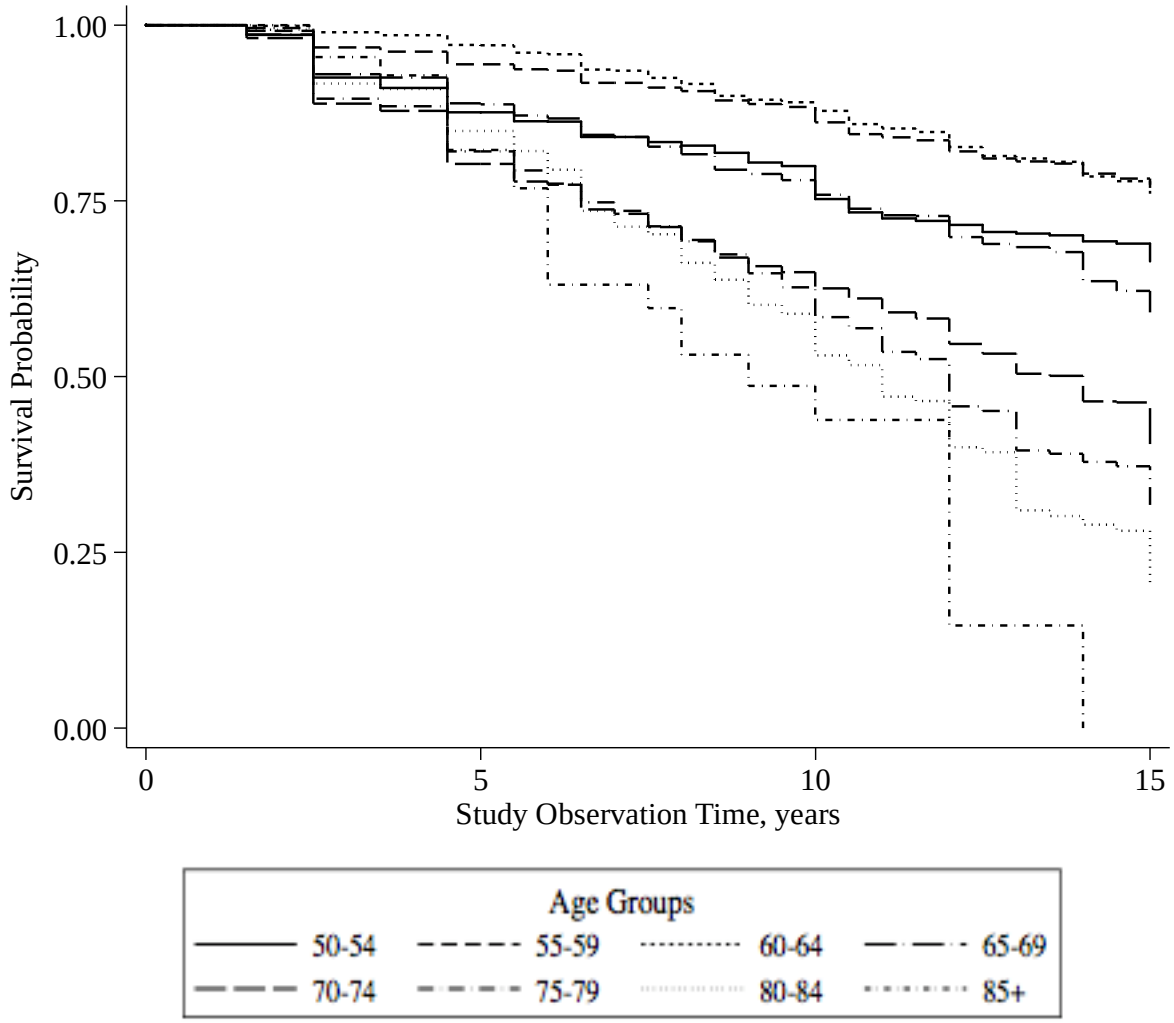


Figure 3. Age-specific incidence of ADRD-pattern cognitive declines.