Title: Multiple memory tests in midlife improve prediction of mild cognitive impairment

Daniel E. Gustavson, Ph.D.[a], Mark Sanderson-Cimino, M.A.[a], Jeremy A. Elman, Ph.D.[a], Carol E. Franz, Ph.D.[a], Matthew S. Panizzon, Ph.D.[a], Amy J. Jak, Ph.D.[a, b], Chandra A. Reynolds, Ph.D.[c], Michael J. Lyons, Ph.D.[d], and William S. Kremen, Ph.D.[a, b]

[a] Department of Psychiatry, Center for Behavior Genetics of Aging, University of California, San Diego, La Jolla, CA
[b] Center of Excellence for Stress and Mental Health, Veterans Affairs San Diego Healthcare System, La Jolla, CA
[c] Department of Psychology, University of California, Riverside, Riverside, CA
[d] Department of Psychological and Brain Sciences, Boston University, Boston, MA

Correspondence concerning this article should be addressed to Daniel Gustavson, Department of Psychiatry and Center for Behavioral Genetics of Aging, University of California San Diego, 9500 Gilman Dr. (MC 0738), La Jolla, CA 92093. E-mail: dgustavson@ucsd.edu

Abstract

Objective: Biomarkers predict risk for Alzheimer's disease, but the ability to predict when most people are still biomarker negative would foster earlier identification of risk. We hypothesized that memory testing—particularly combining multiple memory measures—in cognitively normal middle-aged adults would provide good prediction of 6-year progression to mild cognitive impairment (MCI). Our use of multiple measures contrasts with frequent efforts to identify the single best measure for prediction, an underlying assumption of which is that adding a suboptimal measure will worsen prediction.

Method: We examined how Logical Memory, Visual Reproductions, and California Verbal Learning Test (CVLT) predict 6-year progression to MCI in 849 men who were cognitively normal at baseline (mean age=55.69±2.45).

Results: CVLT learning trials was the best individual predictor of amnestic MCI (OR=4.75). However, a latent factor score incorporating 7 measures across the 3 memory tests provided the strongest prediction (OR=9.88). This OR compared favorably to biomarker prediction in a study of substantially older adults.

Interpretation: Use of a single best measure may appear to be time- and cost-effective, but 30 additional minutes of memory testing—even when it includes individually less predictive measures—provides a very worthwhile tradeoff for substantially improved prediction of progression to amnestic MCI. Early identification is important for potentially slowing Alzheimer's disease progression, and these results indicate that neuropsychological tests can be very early sensitive indicators of risk at an age when few individuals will have yet become biomarker positive.

Keywords: mild cognitive impairment, Alzheimer's disease, neuropsychology

The pathogenesis of Alzheimer's disease (AD) begins decades before the onset of dementia, so it is necessary to identify risk factors as early as possible[1-4]. Recent frameworks of AD have emphasized biomarkers (e.g., amyloid beta, tau) to improve early identification and move toward a biological diagnosis[5, 6]. Abnormal biomarker levels are associated with poor cognition and greater cognitive decline in both cognitively normal adults and in those already diagnosed with mild cognitive impairment (MCI), a prodromal phase of AD[7-11]. However, given the protracted period of disease progression, biomarker positivity indicates that a certain amount of disease progression has already taken place. The ability to identify individuals at risk before they become biomarker positive would thus be of great potential value. As such, there is pressing need to identify other tests that are non-invasive, low-cost, and can improve earlier identification of risk for AD[4, 12].

Episodic memory is an effective early predictor of MCI/AD progression[13-19]. This is intuitive as memory is severely impaired in AD, and MCI diagnoses are in part based on impaired memory performance[20]. However, comprehensive neuropsychological assessment of memory (and other cognitive domains) is often lacking in longitudinal studies aimed at detecting which individuals are at the greatest risk for MCI.

Efforts to improve longitudinal prediction of MCI or AD using cognitive measures have often focused on evaluating which measures provide the best prediction compared with other measures[19, 21, 22]. Although it is clearly theoretically and practically useful—and perhaps cost-effective—to identify which specific tests and measures[1] are the most useful at predicting later impairment, an unspoken assumption is that including the lesser predictors in the model will not improve (or may even hurt) predictive ability. In the current study, we tested the hypothesis that

[1] For the purposes of this paper, we use the term "test" to refer to the overarching task (e.g., the Logical Memory test) and "measure" to refer to the scores available within test (e.g., immediate recall, delayed recall).

combining multiple memory measures (within test and across multiple tests) will provide a stronger and more robust prediction than the single best measure. Indeed, we have demonstrated that although multiple memory tests have common genetic and environmental variance, specific tests (and measures within tests) have independent genetic and environmental influences, suggesting that memory tests are not interchangeable[23, 24].

We evaluated our hypothesis in a community sample of middle-aged adults from the Vietnam Era Twin Study of Aging (VETSA) who were all cognitively normal at baseline. They completed multiple memory tests at mean ages 56 and 62. Examining these associations in midlife will be important because improving treatment efficacy may depend on early intervention[12, 25], and cognitive abilities may already be declining by the late 50s[24, 26-28]. In these analyses, we compare two approaches to aggregating measures: *z*-score composites and factor scores (derived from structural equation models). We expected that the factor score approach would have the strongest prediction because they weigh more strongly the measures that are the best indicators of the latent memory construct.

Method

Participants

Data were obtained from 1289 men who completed neuropsychological memory tests during the first wave of the VETSA (mean age 56). The final sample included 849 individuals who were cognitively normal at the first wave, participated in wave 2 (mean age 62), and had data for all covariates. All participants were recruited randomly from a previous large-scale study of Vietnam Era Twin Registry participants[29]. All participants served in the United States military at some time between 1965 and 1975; nearly 80% did not serve in combat or in Vietnam[30, 31]. Participants are generally representative of American men in their age group with respect to

health and lifestyle characteristics[32]. All participants provided informed consent and the study was approved by local Institutional Review Boards at the University of California, San Diego and Boston University.

Episodic Memory Measures

Episodic memory was measured at both waves with the Logical Memory (LM) and Visual Reproductions (VR) subtests of the Wechsler Memory Scale-III[33], and the California Verbal Learning Test-II (CVLT)[34]. For LM and VR, we examined immediate recall and delayed recall measures. For the CVLT, we examined the short delay and long delay free recall measures, as well as the total score for the learning trials (1 through 5). In analyses involving MCI, all memory measures were *z*-scored and transformed so that odds ratios (ORs) reflect the increase in odds of MCI for every decrease of 1 *SD* in memory performance (in relation to the full sample of N=1289 at wave 1). These standardization procedures were also conducted for the aggregated memory measures described next.

In addition to examining the predictive ability of each memory measure alone we combined measures in two ways. First, we created *z*-score composites for measures within a given test (e.g., LM immediate recall, LM delayed recall), and comparable measures across tests (e.g., LM immediate recall, VR immediate recall, CVLT short delay free recall). We also created *z*-score composites that combined the 6 short/long delay conditions across all tests (i.e., all measures except CVLT learning trials) or all 7 memory measures (including CVLT learning trials).

Second, we created factor scores from latent memory variables. Latent factors were exported from structural equation models in MPlus version 7.2[35] based in part on those reported in earlier work from this sample[24]. These factor scores are similar to the *z*-score composites, but

measures that have stronger factor loadings on the latent memory factor are weighted more heavily. These latent factors are displayed in Figure 1 and were also based on the full wave 1 sample (N = 1289). Factor scores were created for immediate recall measures, delayed recall measures, immediate and delayed recall measures (i.e., the model presented in Panizzon et al., 2015), and all 7 measures[2]. Each model had good fit to the data based on standard structural equation metrics, including Root Mean Square Error of Approximation values < .06, and the Comparative Fit Index values > .950[36]. There were a few cases where individuals were missing memory data for specific measures (see Table 2). In these cases, *z*-scores and factor scores were created based on all available measures.

Mild cognitive impairment diagnoses

MCI was diagnosed using the Jak-Bondi approach[4, 20, 37]. Impairment in a cognitive domain was defined as having at least two tests >1.5 SDs below the age- and education-adjusted normative means after accounting for "premorbid" cognitive ability by adjusting neuropsychological scores for performance on a test of general cognitive ability that was taken at a mean age of 20 years. The adjustment for age 20 cognitive ability ensures that the MCI diagnosis is capturing a decline in function rather than long-standing low ability. Most prior studies have used an impairment criterion on >1 SD, but they have had considerably older samples. As explained previously[38], we used the conservative criterion on >1.5 SDs because of the expected lower base rate in our late middle-aged sample. The validity of the VETSA MCI diagnoses is supported by previous studies[4, 20, 37] and in the present sample by evidence of

[2] In the final model (Figure 1d), we used modification indices to identify an additional factor loading from the CVLT learning trials directly to the latent Episodic Memory factor. This factor loading was not predicted *a priori* but was intuitive given that this is was only memory measure that directly assessed encoding. Results of the model were similar without this additional factor loading, but the model displayed in Figure 1d was more parsimonious (based on all model fit indices), and the other factor loadings in the model were most similar to those in Figure 1c (without CVLT learning trials).

reduced hippocampal volume in those diagnosed with amnestic MCI[39]. Higher AD polygenic risk scores were also associated with significantly increased odds of MCI in this sample[40], indicating that the MCI diagnosis is genetically-related to AD.

MCI diagnoses at wave 2 were also based on measures that were adjusted to account for practice effects[41], leveraging data from attrition replacement subjects who completed the task battery for the first time at wave 2 (N=179), to estimate the increase in performance expected in returnees who completed the tests twice.

Because we were interested in transition to MCI, analyses included only individuals who were cognitively normal at wave 1 and had data for all covariates. Of the 849 returnees meeting this criterion, 45 (5.3%) progressed to amnestic MCI, and 41 (4.8%) progressed to non-amnestic MCI (impairment in executive function, attention/working memory, visual-spatial, and/or processing speed, but not memory).

Data Analysis

All statistical analyses were conducted using R version 3.5.1. Analyses involving MCI were conducted with mixed effects logistic regression using the lme4 package[42]. In these analyses, we controlled for wave 1 age, the time interval between wave 1 and wave 2 assessments, years of education, race/ethnicity (white non-Hispanic vs. other), wave 1 diabetes status (yes/no), wave 1 hypertension status (yes/no), *APOE*-ε4 allele status (ε4+ vs. ε4-), and wave 1 depression symptoms based on Center for Epidemiologic Studies–Depression scale[43]. Diabetes and hypertension status were based on whether the participant either (a) reported being diagnosed by a doctor, (b) reported they were currently taking medication for diabetes or high blood pressure, and/or (c) whether they had high blood pressure on the day of testing (hypertension only). Finally, twin pair ID was included as a random effect to account for the

clustering of data within families. The lme4 package uses list-wise deletion with missing observations, and reports profile-based 95% confidence intervals (95% CIs).

Results

Descriptive Statistics

Demographic and clinical characteristics of the sample are displayed in Table 1. There were no significant differences between cognitively normal and amnestic MCI groups in any of these characteristics. Compared to cognitively normal returnees, those diagnosed with non-amnestic MCI at wave 2 were older at baseline (p = .024) and a smaller proportion were white non-Hispanic (p = .031). Descriptive statistics for memory measures at wave 1 are displayed in Table 2.

Six-Year Prediction of Progression to Mild Cognitive Impairment

The results of the primary analyses are displayed in Table 3. Each cell in Table 3 displays an odds ratio from a separate longitudinal logistic regression in which that memory measure (or combination of multiple memory measures) predicts progression to amnestic MCI (Table 3a) or non-amnestic MCI (Table 3b) 6 years later. Higher odds ratios indicate increased odds of progression to MCI estimated for a 1 *SD* decrease on the baseline memory measure.

As shown in Table 3a, each of the 7 individual memory measures at wave 1 significantly predicted progression to amnestic MCI at wave 2. The learning trial measure of the CVLT provided the strongest individual prediction, OR = 4.75, 95% CI [2.59, 12.44], whereas the immediate recall condition of VR was the weakest individual predictor, OR = 1.91, 95% CI [1.21, 3.43]. However, the strongest prediction came from the latent factor score that incorporated all 7 dependent measures at wave 1, OR = 9.88, 95% CI [4.39, 37.72].

Although most confidence intervals overlapped, long delay recall conditions tended to be

stronger predictors of progression to MCI than short delay conditions (11.5% higher ORs on average, range=4.4 to 23.0%) and OR estimates for combinations of multiple measures greatly outperformed the individual measures that comprised them. Specifically, aggregating measures within the same test (e.g., LM immediate and delay recall) resulted in a small increase in ORs (average=21.2% larger ORs, range=2.1 to 55.1%). Aggregating measures of the same type across tests (e.g., all immediate recall measures) resulted in a larger increase in ORs (average=105.9% larger ORs, range=46.0 to 296.9%). Aggregating across all 7 measures resulted in the largest increase in ORs compared to individual measures (average=213.1% larger ORs, range=85.6 to 517.2%). Finally, OR estimates for latent factors scores were always larger than *z*-score composites (average=8.7% larger ORs, range=5.1 to 12.0%), but all 95% CIs overlapped.

A different pattern of results was observed for non-amnestic MCI. Some memory measures at wave 1 predicted progression to non-amnestic MCI at wave 2 (3 out of 7 measures), the strongest of which was VR immediate recall, OR = 2.02, 95% CI [1.33, 3.28]. However, the predictive ability was far weaker than it was for amnestic MCI. Although aggregating measures sometimes resulted in larger ORs than those for individual measures, none of these estimates was stronger than that for VR immediate recall alone. Thus, there was some evidence that baseline memory predicted later non-amnestic impairment, but combining a mix of informative and non-informative predictors did not appear to strengthen prediction.

Discussion

The results of the study confirmed our primary hypothesis that combining multiple measures of memory improved prediction of progression to amnestic MCI across 6 years. These results have strong implications for prospective studies aimed at very early identification of individuals at greatest risk for MCI and AD. They suggest that longitudinal prediction of

amnestic MCI can be substantially improved by administering multiple memory tests and measures at baseline. The use of a single best measure may appear to be time- and cost-effective. However, administering neuropsychological tests of both verbal and non-verbal memory added only about 30 minutes of additional test administration time and resulted in a dramatic increase in the odds ratio.

Predictors of 3-year progression to MCI or dementia in cognitively normal adults in the Australian Imaging, Biomarkers, and Lifestyle Study of Ageing (AIBL) serves as a sample comparison[44]. Aβ positivity was the best individual predictor (OR=4.8), similar to our CVLT trials 1-5 (OR=4.75). AIBL examined CVLT delayed recall (OR=4.2), which was similar to that demonstrated in VETSA here (OR=3.77). The best prediction in AIBL came from combining Aβ and CVLT delayed (OR=15.9) compared with the prediction based on all memory measures in VETSA (OR=9.88). A total of 13% of AIBL progressed to MCI or dementia whereas 5% of VETSA participants progressed to amnestic MCI. However, the memory tests in VETSA are accounting for substantially earlier prediction because the average age was 16 years younger than in AIBL (mean age 56 vs. 72). Importantly, at the baseline age in VETSA, Aβ positivity would likely have little predictive value because so few individuals would be Aβ positive. Thus, non-invasive neuropsychological tests appear to be sensitive for very early risk identification, largely before detection of Aβ at currently defined thresholds.

Although it may seem obvious that memory measures predict later memory impairment, it is still notable that the prediction was 6 years later. Few prospective studies of MCI or AD rigorously control for baseline memory, and many have focused on finding the best individual predictors[19, 21, 22], rather than aggregating across all measures[22, 45-48.] Biomarkers characterize underlying pathology, and because measurable abnormalities appear early in the disease process

they are important and effective predictors of later impairment[7-9]. However, the results described here strongly suggest that we should not lose sight of cognitive tests as low-cost, non-invasive assessments because they indicate that at currently detectable levels, biomarkers are not necessarily the earliest predictors of progression to MCI. This may be especially important in middle-aged samples such as the one utilized here, where we may not expect many individuals to be biomarker positive at baseline.

In addition, the results suggest that latent variable approaches can be employed to improve prediction of MCI. Given the proportion of amnestic MCI in this sample (5.3%), power estimates using Gpower V3.1.9.2 (without covariates) suggest that a study which administered only logical memory delayed recall (OR=3.22) would require 102 subjects to significantly predict progression to amnestic MCI at the .05 level (with 80% power). When using the latent factor based on all 3 memory tests (OR=9.88) the same power analysis indicates that only 35 subjects would be necessary. We conclude that this small amount of additional non-invasive testing time is a very worthwhile investment.

Creating *z*-score composites also effectively led to stronger prediction of progression to amnestic MCI compared with individual measures, especially when measures were aggregated across tests. However, averaging multiple measures ignores the fact that some measures are much better indexes of memory ability than others. Although the average increase in OR was only about 8.7% with factor scores compared to *z*-score, this two-stage process (fitting a structural model of the memory measures, then using the exported factor score in a standard logistic regression) can be easily applied to other datasets and may be valuable in large-scale studies.

In contrast, combining multiple predictors did not increase overall prediction for progression to non-amnestic MCI (though all 95% CIs overlapped). The results argue against general cognitive deficit as a predictor of MCI. Rather, there appears to be some specificity of cognitive predictors of MCI. Thus, combining multiple cognitive measures appears to be capable of dramatic improvement in prediction *if they are domain-relevant*.

Some limitations of the study should be noted. First the sample comprised only men, so it will be important to examine whether these findings generalize to samples with women. Second, amnestic MCI diagnoses were not validated with AD biomarkers. On the other hand, VETSA MCI diagnoses at wave 1 were validated with evidence of reduced hippocampal volume in those diagnosed with amnestic MCI[39], and higher AD polygenic risk scores in those with amnestic MCI[40], the latter supporting their being AD-related.

In summary, baseline measures of memory can strongly predict progression to amnestic MCI. Although all individuals were cognitively normal at baseline, individual differences still effectively predicted later amnestic impairment. Prospective studies of MCI designed to identify those at greatest risk should administer multiple memory tests as early as possible to maximize their ability to predict changes. It is also worth examining whether and when tests in other cognitive domains might further improve prediction. Nevertheless, the results indicate that a small amount of additional testing time could go a long way in increasing power and improving prediction, and that neuropsychological assessment can be a sensitive predictor of risk for MCI even before individuals are likely to have become biomarker positive.

Acknowledgements

The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the NIA/NIH, or the VA. The U.S. Department of Veterans Affairs has provided financial support for the development and maintenance of the Vietnam Era Twin (VET) Registry. Numerous organizations have provided invaluable assistance in the conduct of the VET Registry, including: Department of Defense; National Personnel Records Center, National Archives and Records Administration; Internal Revenue Service; National Opinion Research Center; National Research Council, National Academy of Sciences; the Institute for Survey Research, Temple University. This material was, in part, the result of work supported with resources of the VA San Diego Center of Excellence for Stress and Mental Health Healthcare System. Most importantly, the authors gratefully acknowledge the continued cooperation and participation of the members of the VET Registry and their families as well as the contributions of many staff members and students.

Funding

This research was supported by Grants R01 AG050595, R01 AG022381, and R01 AG059329 from the National Institutes of Health.

Author Contributions

1) Conception of the study: DG, WK. 2) Acquisition and analysis of data: DG, MS, JA, MP. 3) Drafting a significant portion of the manuscript: all authors

References

1. Bateman, R.J., Xiong, C., et al. (2012). Clinical and biomarker changes in dominantly inherited Alzheimer's disease. *New England Journal of Medicine, 367*, 795-804. doi:10.1056/NEJMoa1202753

2. Bennett, D.A., Schneider, J.A., Arvanitakis, Z., Kelly, J.F., Aggarwal, N.T., Shah, R.C., & Wilson, R.S. (2006). Neuropathology of older persons without cognitive impairment from two community-based studies. *Neurology, 66*, 1837-1844. doi:DOI 10.1212/01.wnl.0000219668.47116.e6

3. Aizenstein, H.J., Nebes, R.D., et al. (2008). Frequent amyloid deposition without significant cognitive impairment among the elderly. *Archives of neurology, 65*, 1509-1517. doi:DOI 10.1001/archneur.65.11.1509

4. Kremen, W.S., Jak, A.J., et al. (2014). Early identification and heritability of mild cognitive impairment. *International journal of epidemiology, 43*, 600-10. doi:10.1093/ije/dyt242

5. Jack, C.R., Jr., Bennett, D.A., et al. (2018). NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia, 14*, 535-562. doi:10.1016/j.jalz.2018.02.018

6. Sperling, R.A., Aisen, P.S., et al. (2011). Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia, 7*, 280-92. doi:10.1016/j.jalz.2011.03.003

7. Ho, J.K., Nation, D.A., & Alzheimer's Disease Neuroimaging, I. (2018). Neuropsychological profiles and trajectories in preclinical Alzheimer's Disease. *Journal*

of the International Neuropsychological Society, 24, 693-702. doi:10.1017/S135561771800022X

8. Jansen, W.J., Ossenkoppele, R., et al. (2018). Association of cerebral amyloid-beta aggregation with cognitive functioning in persons without dementia. *JAMA psychiatry, 75*, 84-95. doi:10.1001/jamapsychiatry.2017.3391

9. Papp, K.V., Mormino, E.C., Amariglio, R.E., Munro, C., Dagley, A., Schultz, A.P., Johnson, K.A., Sperling, R.A., & Rentz, D.M. (2016). Biomarker validation of a decline in semantic processing in preclinical Alzheimer's disease. *Neuropsychology, 30*, 624-30. doi:10.1037/neu0000246

10. Mitchell, A.J. (2008). The clinical significance of subjective memory complaints in the diagnosis of mild cognitive impairment and dementia: a meta-analysis. *International Journal of Geriatric Psychiatry, 23*, 1191-202. doi:10.1002/gps.2053

11. Petersen, R.C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of internal medicine, 256*, 183-194. doi:DOI 10.1111/j.1365-2796.2004.01388.x

12. Sperling, R.A., Mormino, E., & Johnson, K. (2014). The evolution of preclinical Alzheimer's disease: Implications for prevention trials. *Neuron, 84*, 608-22. doi:10.1016/j.neuron.2014.10.038

13. Edmonds, E.C., Delano-Wood, L., Galasko, D.R., Salmon, D.P., Bondi, M.W., & Alzheimer's Disease Neuroimaging, I. (2015). Subtle cognitive decline and biomarker staging in preclinical Alzheimer's disease. *Journal of Alzheimer's Disease, 47*, 231-42. doi:10.3233/JAD-150128

14. Gomar, J.J., Bobes-Bascaran, M.T., Conejero-Goldberg, C., Davies, P., Goldberg, T.E., & Alzheimer's Disease Neuroimaging, I. (2011). Utility of combinations of biomarkers,

cognitive markers, and risk factors to predict conversion from mild cognitive impairment to Alzheimer disease in patients in the Alzheimer's disease neuroimaging initiative. *Archives of General Psychiatry, 68*, 961-9. doi:10.1001/archgenpsychiatry.2011.96

15. Ewers, M., Walsh, C., et al. (2012). Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. *Neurobiology of Aging, 33*, 1203-1214. doi:10.1016/j.neurobiolaging.2010.10.019

16. Gomar, J.J., Conejero-Goldberg, C., Davies, P., Goldberg, T.E., & Alzheimer's Disease Neuroimaging, I. (2014). Extension and refinement of the predictive value of different classes of markers in ADNI: Four-year follow-up data. *Alzheimer's & Dementia, 10*, 704-12. doi:10.1016/j.jalz.2013.11.009

17. Insel, P.S., Mattsson, N., Mackin, R.S., Scholl, M., Nosheny, R.L., Tosun, D., Donohue, M.C., Aisen, P.S., Jagust, W.J., Weiner, M.W., & Alzheimer's Disease Neuroimaging, I. (2016). Accelerating rates of cognitive decline and imaging markers associated with beta-amyloid pathology. *Neurology, 86*, 1887-96. doi:10.1212/WNL.0000000000002683

18. Jedynak, B.M., Lang, A., Liu, B., Katz, E., Zhang, Y., Wyman, B.T., Raunig, D., Jedynak, C.P., Caffo, B., Prince, J.L., & Alzheimer's Disease Neuroimaging, I. (2012). A computational neurodegenerative disease progression score: Method and results with the Alzheimer's Disease Neuroimaging Initiative cohort. *Neuroimage, 63*, 1478-86. doi:10.1016/j.neuroimage.2012.07.059

19. Tabert, M.H., Manly, J.J., Liu, X., Pelton, G.H., Rosenblum, S., Jacobs, M., Zamora, D., Goodkind, M., Bell, K., Stern, Y., & Devanand, D.P. (2006). Neuropsychological

prediction of conversion to Alzheimer disease in patients with mild cognitive impairment. *Arch Gen Psychiatry, 63*, 916-24. doi:10.1001/archpsyc.63.8.916

20. Bondi, M.W., Edmonds, E.C., Jak, A.J., Clark, L.R., Delano-Wood, L., McDonald, C.R., Nation, D.A., Libon, D.J., Au, R., Galasko, D., & Salmon, D.P. (2014). Neuropsychological criteria for mild cognitive impairment improves diagnostic precision, biomarker associations, and progression rates. *Journal of Alzheimer's Disease, 42*, 275-89. doi:10.3233/JAD-140276

21. De Jager, C.A., Hogervorst, E., Combrinck, M., & Budge, M.M. (2003). Sensitivity and specificity of neuropsychological tests for mild cognitive impairment, vascular cognitive impairment and Alzheimer's disease. *Psychological Medicine, 33*, 1039-50.

22. Lange, K.L., Bondi, M.W., Salmon, D.P., Galasko, D., Delis, D.C., Thomas, R.G., & Thal, L.J. (2002). Decline in verbal memory during preclinical Alzheimer's disease: examination of the effect of APOE genotype. *J Int Neuropsychol Soc, 8*, 943-55.

23. Kremen, W.S., Panizzon, M.S., et al. (2014). Genetic complexity of episodic memory: a twin approach to studies of aging. *Psychol Aging, 29*, 404-17. doi:10.1037/a0035962

24. Panizzon, M.S., Neale, M.C., et al. (2015). Genetic and environmental architecture of changes in episodic memory from middle to late middle age. *Psychology and Aging, 30*, 286-300. doi:10.1037/pag0000023

25. Golde, T.E., Schneider, L.S., & Koo, E.H. (2011). Anti-abeta therapeutics in Alzheimer's disease: the need for a paradigm shift. *Neuron, 69*, 203-13. doi:10.1016/j.neuron.2011.01.002

26. Gustavson, D.E., Panizzon, M.S., Elman, J.A., Franz, C.E., Reynolds, C.A., Jacobson, K.C., Friedman, N.P., Xian, H., Lyons, M.J., & Kremen, W.S. (2018). Stability of genetic

and environmental influences on executive functions in midlife. *Psychology and Aging, 33*, 219-231. doi:10.1037/pag0000230

27. Salthouse, T.A. (2009). When does age-related cognitive decline begin? *Neurobiol Aging, 30*, 507-14. doi:10.1016/j.neurobiolaging.2008.09.023

28. Salthouse, T.A. (2012). Are individual differences in rates of aging greater at older ages? *Neurobiol Aging, 33*, 2373-81. doi:10.1016/j.neurobiolaging.2011.10.018

29. Tsuang, M.T., Bar, J.L., Harley, R.M., & Lyons, M.J. (2001). The Harvard Twin Study of Substance Abuse: What we have learned. *Harvard Review of Psychiatry, 9*, 267-279. doi:10.1093/hrp/9.6.267

30. Kremen, W.S., Panizzon, M.S., Xian, H., Barch, D.M., Franz, C.E., Grant, M.D., Toomey, R., & Lyons, M.J. (2011). Genetic architecture of context processing in late middle age: More than one underlying mechanism. *Psychology and Aging, 26*, 852-63. doi:10.1037/a0025098

31. Kremen, W.S., Thompson-Brenner, H., Leung, Y.M., Grant, M.D., Franz, C.E., Eisen, S.A., Jacobson, K.C., Boake, C., & Lyons, M.J. (2006). Genes, environment, and time: The Vietnam Era Twin Study of Aging (VETSA). *Twin Research and Human Genetics, 9*, 1009-22. doi:10.1375/183242706779462750

32. Schoenborn, C.A. & Heyman, K.M. (2009). Health characteristics of adults aged 55 years and over: United States, 2004-2007. *National Health Statistics Report, 16*, 1-31.

33. Wechsler, D., *Wechsler Memory Scale (WMS-III)*. 1997, San Antonio, TX: Psychological Corporation.

34. Delis, D.C., Kramer, J.H., Kaplan, E., & Ober, B.A., *California Verbal Learning Test (CVLT-2). 2nd ed.*. 2000, San Antonio, TX: Psychological Corporation.

35. Muthén, L.K. & Muthén, B.O., *Mplus Version 7 user's guide*. 2010-2012, Los Angeles, CA: Muthén & Muthén.

36. Hu, L.T. & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling-a Multidisciplinary Journal, 6*, 1-55. doi:10.1080/10705519909540118

37. Jak, A.J., Bondi, M.W., Delano-Wood, L., Wierenga, C., Corey-Bloom, J., Salmon, D.P., & Delis, D.C. (2009). Quantification of five neuropsychological approaches to defining mild cognitive impairment. *The American Journal of Geriatric Psychiatry, 17*, 368-75. doi:10.1097/JGP.0b013e31819431d5

38. Granholm, E.L., Panizzon, M.S., Elman, J.A., Jak, A.J., Hauger, R.L., Bondi, M.W., Lyons, M.J., Franz, C.E., & Kremen, W.S. (2017). Pupillary Responses as a Biomarker of Early Risk for Alzheimer's Disease. *J Alzheimers Dis, 56*, 1419-1428. doi:10.3233/JAD-161078

39. Jak, A.J., Panizzon, M.S., et al. (2015). Hippocampal atrophy varies by neuropsychologically defined MCI among men in their 50s. *American Journal of Geriatric Psychiatry, 23*, 456-465. doi:10.1016/j.jagp.2014.08.011

40. Logue, M.W., Panizzon, M.S., et al. (2019). Use of an Alzheimer's disease polygenic risk score to identify mild cognitive impairment in adults in their 50s. *Molecular Psychiatry, 24*, 421-430. doi:10.1038/s41380-018-0030-8

41. Elman, J.A., Jak, A.J., et al. (2018). Underdiagnosis of mild cognitive impairment: A consequence of ignoring practice effects. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 10*, 372-381. doi:10.1016/j.dadm.2018.04.003

42. Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*. doi:10.18637/jss.v067.i01

43. Radloff, L.S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied psychological measurement, 1*, 385-401.

44. Rowe, C.C., Bourgeat, P., et al. (2013). Predicting Alzheimer disease with beta-amyloid imaging: results from the Australian imaging, biomarkers, and lifestyle study of ageing. *Annals of neurology, 74*, 905-13. doi:10.1002/ana.24040

45. Mickes, L., Wixted, J.T., Fennema-Notestine, C., Galasko, D., Bondi, M.W., Thal, L.J., & Salmon, D.P. (2007). Progressive impairment on neuropsychological tasks in a longitudinal study of preclinical Alzheimer's disease. *Neuropsychology, 21*, 696-705. doi:10.1037/0894-4105.21.6.696

46. Rabin, L.A., Pare, N., Saykin, A.J., Brown, M.J., Wishart, H.A., Flashman, L.A., & Santulli, R.B. (2009). Differential memory test sensitivity for diagnosing amnestic mild cognitive impairment and predicting conversion to Alzheimer's disease. *Aging, Neuropsychology, and Cognition, 16*, 357-76. doi:10.1080/13825580902825220

47. Jonaitis, E.M., Koscik, R.L., et al. (2019). Measuring longitudinal cognition: Individual tests versus composites. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 11*, 74-84. doi:10.1016/j.dadm.2018.11.006

48. Gustavson, D.E., Elman, J.A., et al. (Manuscript submitted for publication). Semantic fluency predicts six-year progression to mild cognitive impairment in middle-aged men.

Table 1

Demographic and Clinical Characteristics and Covariates Included in Analyses Involving Mild Cognitive Impairment (MCI)

Demographic Variable	Remained Cognitively Normal (N = 763)		Progressed to Amnestic MCI (N = 45)		Progressed to Non-Amnestic MCI (N = 41)		*p* (CN vs. aMCI)	*p* (CN vs. nMCI)
	M	*SD*	*M*	*SD*	*M*	*SD*		
Lifetime Education (years)	14.03	2.16	13.58	1.75	13.49	1.83	0.191	0.098
Age (at wave 1)	55.69	2.45	56.37	2.48	56.68	2.41	0.102	**0.024**
Age Interval (wave 2 - wave 1)	5.75	0.70	5.76	0.67	5.65	0.54	0.907	0.400
Depression Symptoms (at wave 1)	7.69	7.14	7.32	7.03	9.71	9.66	0.858	0.152
Ethnicity (% white non-Hispanic)	91.34	-	84.44	-	80.49	-	0.119	**0.031**
ApoE status (% e4 positive)	30.80	-	33.33	-	21.95	-	0.480	0.193
Diabetes (% yes at wave 1)	10.35	-	11.11	-	17.07	-	0.868	0.267
Hypertension (% yes at wave 1)	58.32	-	64.44	-	73.17	-	0.230	0.069

Note: The final two columns display the *p* value for comparisons between individuals who remained cognitively normal (N=763) and either progressed to amnestic MCI (aMCI; N=45) or non-amnestic MCI (nMCI; N=41), controlling for clustering of data within families. Significant group differences are displayed in bold ($p < .05$). Lifetime education was the number of years of school completed. Depression symptoms were measured with the Center for Epidemiologic Studies – Depression Scale[43], with scores above 15 indicating risk for clinical depression.

Table 2

Descriptive Statistics for Episodic Memory Measures at Baseline (Mean Age 56)

Memory Variable	N	M	SD	Range	Skewness	Kurtosis
Logical Memory						
Immediate Recall	1281	23.47	6.16	4, 44	-0.04	-0.10
Delayed Recall	1279	20.01	6.63	0, 41	-0.10	-0.13
Visual Reproductions						
Immediate Recall	1284	78.24	12.41	21, 103	-0.58	0.48
Delayed Recall	1283	54.75	19.51	0, 100	-0.15	-0.44
California Verbal Learning Test						
Learning Trials	1270	42.84	8.51	18, 74	0.09	-0.06
Short Delay Free Recall	1270	8.64	2.74	1, 16	0.07	-0.21
Long Delay Free Recall	1269	9.06	2.89	0, 16	0.00	-0.30

Note: In all analyses involving MCI, dependent measures were standardized and reverse scored so that odds ratios reflect increase in risk of MCI at lower levels of cognitive ability. Shown here are descriptive statistics for all individuals at wave 1 (were used to standardize data and create *z*-score composites and latent factor scores). Distributional characteristics remained acceptable for the subset of 849 individuals in the analyses involving progression to MCI (range skewness = -.45 to .22; range kurtosis = -.36 to .05).

Table 3

Odds Ratios and 95% Confidence Intervals from Logistic Regressions of Mild Cognitive Impairment (MCI) Predicted by Baseline Memory Measures

Dependent Measure	Logical Memory	Visual Reproductions	CVLT	*z*-score composites	Latent factor scores
A. Prediction of Amnestic MCI					
Immediate recall only	**3.01**	**1.91**	**3.61**	**5.27**	**5.67**
	[1.79, 6.44]	**[1.21, 3.43]**	**[2.00, 9.39]**	**[2.78, 15.29]**	**[2.92, 16.98]**
Delayed recall only	**3.22**	**2.35**	**3.77**	**5.86**	**6.45**
	[1.89, 6.83]	**[1.50, 4.38]**	**[2.14, 8.48]**	**[3.09, 15.83]**	**[3.30, 17.99]**
Learning trials only	-	-	**4.75**	-	-
			[2.59, 12.44]		
Immediate & delay recall[a]	**3.28**	**2.40**	**4.26**	**6.57**	**6.91**
	[1.93, 7.05]	**[1.51, 4.56]**	**[2.29, 10.92]**	**[3.32, 19.91]**	**[3.47, 20.20]**
Immediate & delay recall + learning trials[a]	-	-	**5.60**	**8.82**	**9.88**
			[2.83, 16.06]	**[4.04, 32.61]**	**[4.39, 37.72]**
B. Prediction of Non-Amnestic MCI					
Immediate recall only	1.10	**2.02**	1.29	**1.71**	**1.63**
	[.71, 1.72]	**[1.33, 3.28]**	[.85, 1.99]	**[1.08, 2.85]**	**[1.03, 2.69]**
Delayed recall only	1.11	**1.55**	**1.78**	**1.72**	**1.86**
	[.71, 1.77]	**[1.01, 2.47]**	**[1.15, 3.01]**	**[1.07, 2.96]**	**[1.16, 3.21]**
Learning trials only	-	-	1.03	-	-
			[.66, 1.57]		
Immediate & delay recall[a]	1.11	**1.96**	**1.53**	**1.78**	**1.91**
	[.71, 1.76]	**[1.27, 3.25]**	**[1.01, 2.47]**	**[1.11, 3.04]**	**[1.19, 3.31]**
Immediate & delay recall + learning trials[a]	-	-	1.36	**1.64**	1.56
			[.89, 2.15]	**[1.02, 2.76]**	[.99, 2.60]

Note: Each cell displays an odds ratio (OR) from a separate analysis in which that memory measure (or combination of measures) predicts progression to amnestic MCI (A) or non-amnestic MCI (B). Significant ORs are displayed in bold ($p < .05$). Like all individual measures, *z*-score composites and factor scores were scored such that higher ORs indicate greater risk for amnestic MCI (aMCI) or non-amnestic MCI (nMCI) at -1 *SD* for that variable.
[a] indicates measures in this row were also based on *z*-score composites (e.g., LM immediate and LM delayed recall in the first column), except for the final column which was based on factor scores (from the models displayed in Figure 1c and 1d).

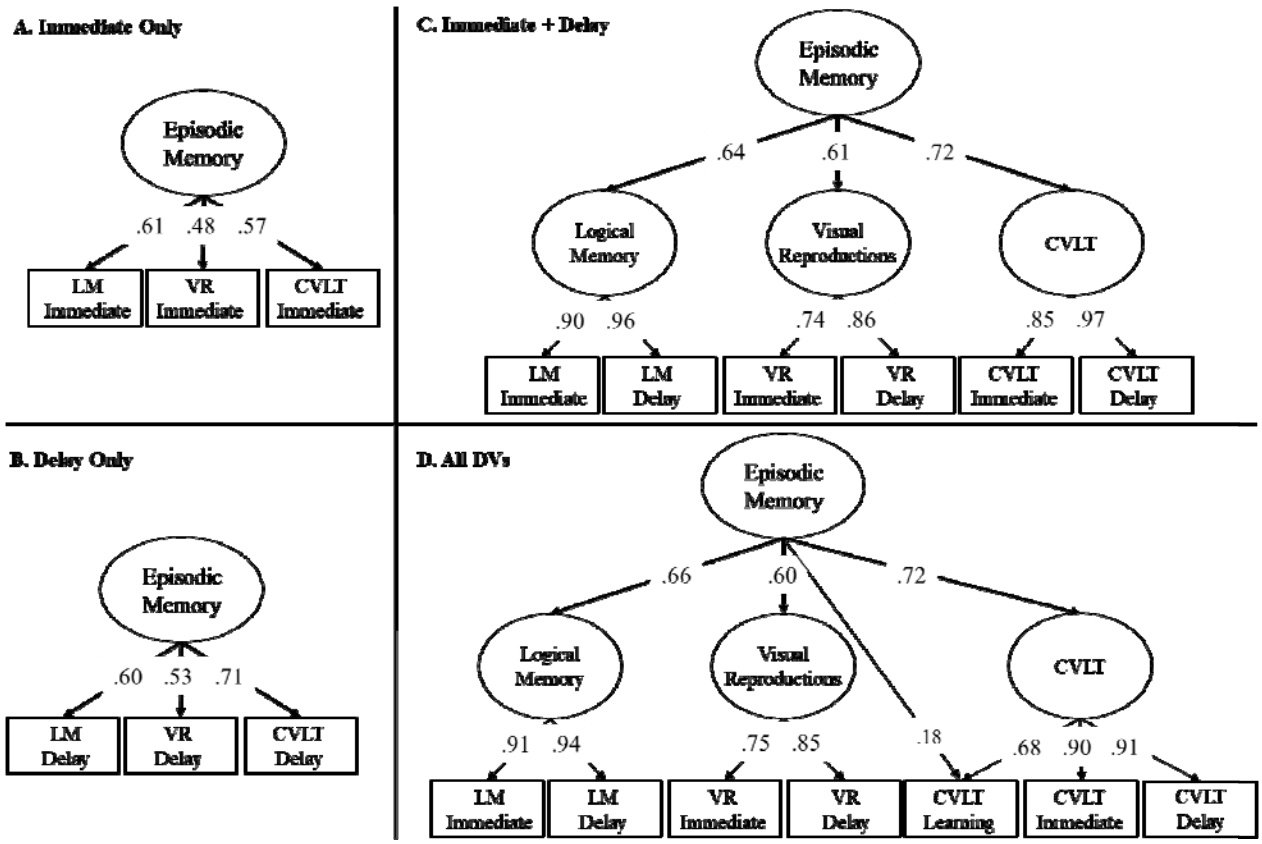

Figure 1: Structural equation models used to create factor scores for episodic memory at wave 1. In each model, variance explained in a given memory measure (rectangle) by latent factors (ovals) can be computed by squaring the factor loading on that factor. Factor scores for the highest-level Episodic Memory latent factor in each model were exported in Mplus and used as a continuous variable in separate logistic regression analyses involving MCI. All factor loadings were significant ($p < .05$) and all models fit the data well.