Bruno, D, Jauregi, A, Pomara, N, Zetterberg, H, Blennow, K, Koscik, R, Carlsson, C, Bendlin, B, Okonkwo, O, Hermann, B, Johnson, S and Mueller, K

Cross-sectional associations of CSF tau levels with Rey's AVLT: A recency ratio study

Abstract

Objective: The preeminent *in-vivo* cerebrospinal fluid (CSF) biomarkers of Alzheimer's disease (AD) are amyloid β 1-42 (Aβ42), p-tau and t-tau. The goal of this study was to examine how well traditional (total and delayed recall) and process-based (recency ratio; Rr) measures derived from Rey's AVLT were associated with these biomarkers.

Method: Data from 235 participants (mean age = 65.5, SD = 6.9), who ranged from cognitively unimpaired to mild cognitive impairment, and for whom CSF values were available, were extracted from the Wisconsin Registry for Alzheimer's Prevention. Bayesian regression analyses were carried out using CSF scores as outcomes, AVLT scores as predictors, and controlling for demographic data and diagnosis.

Results: We found moderate evidence that Rr was associated with both CSF p-tau (BFm = 5.55;) and t-tau (BFm = 7.28), above and beyond the control variables, while it did not correlate with CSF Aβ42 levels. In contrast, total and delayed recall scores were not linked with any of the AD biomarkers, in separate analyses. When comparing all memory predictors in a single regression, Rr remained the strongest predictor of CSF t-tau levels (BFm = 3.57).

Conclusions: Our findings suggest that Rr may be a better cognitive measure than commonly used AVLT scores to assess CSF levels of p-tau and t-tau in non-demented individuals.

Keywords: A/T/N biomarkers, Alzheimer's disease, CSF tau, CSF Aβ42, recency ratio.

Key Points

Question: This paper asks how well process-based measures derived from Rey's AVLT, such as the Recency ratio (Rr), associate with cerebro-spinal (CSF) levels of Alzheimer's disease (AD) biomarkers when compared to traditional measures (e.g., total and delayed recall).

Findings: The results of this study indicate that Rr outperforms AVLT total and delayed recall in predicting CSF levels of t-tau cross-sectionally.

Importance: These findings highlight the importance of process-based measures for the identification of individuals at risk of neurodegeneration.

Next Steps: More research is needed to examine Rr clinical value among individuals presenting with a neurodegenerative disease.

Introduction

In recent years, there have been calls for an unbiased approach to the diagnosis of Alzheimer's disease (AD), leading to the A/T/N classification (Jack et al., 2018). According to this classification, the preeminent *in-vivo* biomarkers of AD reflect amyloid β deposition (*A)*, presence of pathological tau (*T)* and neurodegeneration (*N)*, respectively (Jack et al., 2016; 2018). When using cerebrospinal fluid (CSF) measures, *A* will correspond to levels of amyloid β 1-42 (henceforth, Aβ42), *T* to p-tau levels, and *N* to levels of t-tau.

Alongside biomarkers, however, there is also the need for developing sensitive cognitive measures that reflect the disease associated pathology indexed by biomarker levels (Bock et al., 2021). There are at least two good reasons to do this (Florean et al. 2022). First, cognitive tests are relatively affordable and can be easily available, including in areas where access to laboratory equipment is limited. Relatedly, cognitive tests are relatively unintrusive, thus facilitating data collection on participants who are averse to medical procedures. Second, because subtle neuropsychological changes can be observed already when individuals with elevated levels of AD biomarkers are still asymptomatic for the disease (e.g., Bruno et al. 2021; Mueller et al., 2020).

Among the available neuropsychological tests used for assessment of episodic memory, the Rey Auditory Verbal Leaning test (henceforth, AVLT; Rey, 1958) has been commonly used in older populations. In this test, subjects are read a list of 15 unrelated nouns five times and are asked to free recall these words after each presentation. Then a new 15-word list is tested (interference), followed again by free recall of the originally

presented list. Finally, after about 20-30 minutes, subjects are asked to free recall the original list once again, ending with a recognition test. Most commonly, to evaluate free recall ability, *total recall* is calculated by summing the numbers of correctly recalled items across all five initial (learning) trials. Additionally, *delayed recall* is measured by the number of words recalled correctly after the 20-30 minute delay.

An alternative approach to traditional scoring of neuropsychological tests is the examination of *process scores* (Kaplan, 1988). Process scores make assumptions on the underlying neurocognitive processes leading to test performance, and aim to determine how and why an individual performed the way they did. Process scores have been shown to identify asymptomatic individuals at risk of subsequent cognitive decline (Bruno et al., 2013; Gicas et al., 2020; Talamonti et al., 2020; Thomas et al., 2018; 2020). An example of a free recall process score that can be extracted from AVLT data is the recency ratio (Rr; Bruno et al., 2016). Rr leverages the observation that individuals with AD present with good immediate memory recall for items learned at the end of a list (recency items, *e.g.*, Foldi et al., 2003), while displaying very poor recency performance after a delay (Carlesimo et al., 1995). This ratio between immediate and delayed recency performance has been used to track, among other things, the level of risk in progressing to mild cognitive impairment (MCI) from a healthy baseline (Bruno et al., 2018; Egeland, 2021), and to AD from MCI (Turchetta et al., 2020). Rr has been shown to correlate also with CSF neurogranin, a biomarker of post-synaptic dysfunction (Bruno et al., 2021), and to aid the differential diagnosis of AD (Turchetta et al., 2018). Higher Rr scores indicate more recency forgetting and, consequently, more overall risk of cognitive impairment.

Bruno et al. (2018) showed that 82% of examined individuals who were cognitively intact at baseline and had Rr scores of 1 or above eventually converted to a classification of early MCI.

The goal of this study was to examine how well traditional (total and delayed recall) and process-based (Rr) measures, derived from Rey's AVLT, were associated with cross-sectional CSF A/T/N biomarkers levels. To do this, we examined data from the Wisconsin Registry of Alzheimer's Prevention (WRAP), a population study of middle-aged individuals with a family history of AD, based at the University of Wisconsin – Madison (Johnson et al., 2018). Our goal was to mimic the conditions of a clinical examination as much as possible by selecting the smallest available time lapse between CSF and AVLT visits, and by considering data which, outside of the CSF biomarkers, would be readily available to a clinician, such as age and sex of the individual – for this reason, we opted not to include genetic markers of AD (e.g., APOE status), as these are not typically available in routine examinations.

Methods

Participants. Data were extracted from WRAP, an ongoing longitudinal cohort study based at the University of Wisconsin–Madison, USA. For the present study, WRAP participants were selected based on having completed at least two assessment visits: one for cognitive screening and one for a lumbar puncture leading to CSF extraction. All participants completed the AVLT and were classified as cognitively unimpaired – stable (CUS), cognitively unimpaired – declining (CUD), or with MCI via a consensus conference diagnosis. To classify individuals based on their cognitive status, WRAP uses

a two-tiered consensus conference approach. First, an algorithm that identifies cases where impairment may exist is applied, based on whether or not they meet one or more of the following criteria: (1) the participant performs 1.5 SDs below the mean on factor scores or individual measures of memory, executive function, language, working memory, or attention (Koscik et al., 2014; Clark et al., 2016; Langhough Koscik et al., 2021); (2) cognitive performance on one or more tests falls below values used in other studies as cut-points for clinical MCI diagnoses (e.g. WMS-R Logical Memory II, Wechsler, 1987: story A score <9: AD Neuroimaging Initiative, Petersen et al., 2010); or (3) an abnormal informant report indicates subjective cognitive or functional decline. Second, a consensus diagnosis is determined by a team including physicians, clinical neuropsychologists, and clinical nurse practitioners, based on cognitive, medical history, lifestyle, subjective cognitive complaints, and informant data, for each visit (Langhough Koscik et al., 2021). The CUD label is assigned when participants are performing lower than expected on internal norms, and the consensus review team has ruled out other causes, including worse diagnoses such as MCI or dementia. The MCI diagnosis follows the core clinical criteria (excluding biomarkers) from Albert et al. (2011; but see also Winblad et al. 2004), adopted by the National Institute on Aging-Alzheimer's Association. The core clinical criteria from McKhann et al. (2011) were used for dementia, without reference to biomarkers. From the total pool of 1551 volunteers, 242 participants had CSF data from at least one visit, and 235 participants fulfilled all of the above inclusion criteria (see Table 1 for demographic information). Of these, 200 were CUS, 28 were CUD, and seven had MCI. All activities for this study were approved by

the Institutional Review Board of the University of Wisconsin – Madison, and completed in accordance with the Helsinki Declaration. All participants provided informed consent prior to testing.

Procedure. WRAP procedures have been described previously (e.g., Sager, Hermann & La Rue, 2005; Johnson et al., 2018), but, briefly, WRAP is an ongoing longitudinal study based in Madison, WI, USA, of middle-aged individuals, who attend regular visits, typically every 2 years. Each participant completed self-report questionnaires on demographics, health history and lifestyle, in addition to clinical assessments, and a neuropsychological test battery. Some participants also underwent a lumbar puncture. The neuropsychological test battery included the Rey AVLT, described above.

CSF collection. CSF was extracted using a Sprotte 24- or 25-gauge spinal needle, under fasting conditions. During each lumbar puncture visit, 22 mL of CSF was extracted, which was then combined, mixed, centrifuged and aliquoted into tubes of 1.5 mL capacity. These tubes were stored within 30 minutes at -80◦C (for more details on the CSF procedure, see Van Hulle et al., 2021).

Biomarker measurements. All CSF samples were assayed at the Clinical Neurochemistry Laboratory, University of Gothenburg, using the same batch of reagents under strict quality control procedures. Elecsys β-amyloid(1-42) CSF, Elecsys Phospho-Tau (181P) CSF and Elecsys Total-Tau CSF, were performed on a cobas e 601 analyzer, as previously described (Van Hulle et al., 2021).

Data analysis plan. To fulfil our study aims, we first carried out Bayesian linear regression analyses to test models with AVLT scores against null models including

control variables (specified below). Bayesian analyses allow for the estimation of model plausibility, which permits comparison of models with different combinations of predictors, and for the determination of effect sizes with credible intervals (e.g., Teipel et al., 2021). For all analyses, the model prior was set to Uniform, where all models are *a-priori* equally likely, and the prior on parameters was set to the default Jeffreys-Zellner-Siow (JZS) prior probability distribution, which allows the Bayes factor to be the same regardless of unit of measurement. Credible intervals were set to 89%, which is considered more stable than 95% (Kruschke, 2014). To address potential issues with non-normally distributed residuals in the regressions, Markov chain-Monte Carlo (MCMC) sampling to each analysis was applied 1,000 times. The outcome variables were CSF levels of Aβ42, p-tau and t-tau, in separate analyses. Control variables (forming the null models) were: age at the lumbar puncture, time elapsed between lumbar puncture and memory assessment, sex, consensus diagnosis at lumbar puncture, and the Wide-Range Achievement Test-3 (WRAT-3) Reading Subtest (raw score) at the baseline visit as a culturally-reduced measure of cognitive reserve, within the North American context (Manly et al., 2002). Predictor variables were total recall, delayed recall and Rr: we first looked at these predictors in separate analyses, and then compared them directly in a single regression. Total recall is calculated by adding the numbers of correctly free recalled words across all five learning trials. Delayed recall is the total number of free recalled items at the delayed recall trial. Finally, Rr is calculated by dividing the number of correctly recalled recency words (i.e., the last four words presented) from the learning trial (trial 1 of AVLT) by the corresponding number of correctly recalled recency words

in the delayed recall trial. A correction also was applied ((immediate recency score + 1) / (delayed recency score + 1)) to avoid missing data due to zero scores, as in Bruno et al., 2018. Analyses were conducted using JASP (0.14; https://jasp-stats.org/).

Transparency and openness. Data can be requested here: https://wrap.wisc.edu/data-requests/.

Table 1. Demographics, CSF measures and memory tests scores (mean and standard deviation) for the study participants. Time elapsed was calculated as an absolute value. N = sample size; LP = lumbar puncture; CSF = cerebro-spinal fluid; Rr = recency ratio; AVLT = auditory verbal learning test.

Characteristic	Total	CUS	CUD	MCI
N	235	200	28	7
Sex (females)	156 (66%)	135 (68%)	17 (61%)	4 (57%)
Age at LP (years)	65.5 (6.9)	65.1 (6.9)	67.2 (6.6)	70.4 (5.3)
Time elapsed (years)	1.4 (1.2)	1.5 (1.2)	1.2 (1.1)	1.9 (1.1)
WRAT-3 raw score	51.4 (4.6)	51.2 (4.7)	52.4 (3.4)	50.6 (4.6)
CSF Aβ42 (ng/L)	909.9 (405.9)	905.3 (393.0)	891.3 (484.5)	1115.8(438.7)
CSF P-tau (ng/L)	18.9 (7.4)	18.9 (7.0)	17.8 (7.7)	26.2 (10.8)
CSF T-tau (ng/L)	214.1 (75.7)	213.3 (73.1)	198.7 (72.4)	299.3 (115.1)
Rr	1.3 (0.8)	1.3 (0.7)	1.5 (0.8)	2.9 (1.5)
AVLT total recall	51.9 (8.4)	53.2 (7.9)	45.5 (6.6)	40.6 (9.9)
AVLT delayed recall	10.7 (3.1)	11.2 (2.7)	8.5 (2.6)	4.3 (4.8)

Results

Rr ranged from 0.20 to 5 in CUS, from 0.40 to 4 in CUD, and from 1.25 to 5 in MCI. AVLT total recall ranged from 28 to 71 in CUS, from 30 to 58 in CUD, and from 28 to 59 in MCI. AVLT delayed recall ranged from 4 to 15 in CUS, from 4 to 13 in CUD, and from 0 to 13 in MCI

To test the hypothesis that Rr was associated with CSF levels of Aβ42, the null model was compared with the model including Rr. This analysis yielded no support to the hypothesis that Rr predicts Aβ42 in this sample, as the null model performed better than the model with Rr: the null model had a Bayesian factor (BF_M) of 2.030, meaning that the odds in favour of the null model were about 2 times higher than the odds of the model with Rr. In contrast, the model with Rr had a BF_M of 0.493. However, the Rr model (BF_M = 5.546) outperformed the null model (BF_M = 0.180) with p-tau as the outcome (i.e., the Rr model is over five times as likely as the null model), showing moderate evidence in favour of the Rr model, and suggesting that Rr predicts CSF p-tau levels in this sample better than the control variables. Rr had a posterior mean of 1.15, a SD of 0.71, and lower/higher 89% credible intervals of 0 and 2.053, respectively (note that credible intervals are determined by MCMC sampling). These values indicate that one added Rr point increases CSF p-tau levels by about 1.15 points, corresponding to ~ 6% of the mean CSF p-tau level detected in this sample. Analogously, the Rr model (BF_M = 7.283) outperformed the null model (BF_M = 0.137) with t-tau as the outcome, suggesting that Rr predicts CSF t-tau levels in this sample better than the control variables. Rr had a posterior mean of 12.882, SD = 7.223, and lower/higher 89% credible intervals of 0 and 21.502, respectively. These values indicate that one added Rr point increases CSF t-tau

levels by 12.882 points, corresponding to roughly 6% of the mean t-tau level detected in this sample. In contrast, models with total or delayed recall routinely underperformed the null model.

When comparing models with Rr, total recall and delayed recall directly in a single regression, we observed that the Rr model remained the best model with CSF t-tau as outcome. The Rr model achieved a BF_M of 3.573, showing moderate evidence in favour of that model, with better odds than the second best model, which combined Rr and total recall, $BF_M = 2.549$ ($BF_{10} = 0.790$, meaning that the observed data are 0.790 times as likely to occur under the model with Rr and total recall, as compared to the Rr-only model). A model with all memory scores also performed worse than the model with only Rr: $BF_M = 0.970$, $BF_{10} = 0.360$. However, no model reached a threshold above anecdotal evidence (i.e., $BF_M = 3$) with CSF p-tau levels, topping out at $BF_M = 2.786$ when using Rr alone as a predictor. This result suggests that, while the model with Rr performed better than the null model and models with other memory predictors, the evidence is insufficient to draw firm conclusions. Finally, the null model performed best with CSF Aβ42 levels ($BF_M = 3.824$), beating the model with Rr ($BF_M = 1.475$, $BF_{10} = 0.493$), which was second best.

Put Figure 1a, 1b and 1c about here

Discussion

The goal of this study was to examine how well traditional (total and delayed recall) and process-based (Rr) measures, derived from Rey's AVLT, were associated with A/T/N biomarkers of AD in CSF: Aβ42, p-tau and t-tau. Our results indicate that Rr outperformed both total and delayed AVLT recall measures. Rr was associated with both CFS p- and t-tau levels with moderate evidence, whereas there were no correlations between total and delayed recall, and any of the A/T/N biomarkers. When combining all memory predictors in a single analysis, the odds in favour of the Rr model remained the highest when predicting CSF t-tau levels, while evidence in favour of the Rr model predicting CSF p-tau levels was only anecdotal.

High Rr scores are thought to depend on a reduction in long-term memory, due to a loss of consolidation ability, while reliance on phonological/echoic short-term memory remains relatively intact (Bruno et al., 2018; Turchetta et al., 2020). Consolidation ability is typically associated with function of the medial-temporal lobe (Wixted, 2004; Wixted & Cai, 2013), an area that is implicated early in neurodegeneration and, particularly, in tauopathy (Maass et al., 2019; Tennant et al., 2021). Therefore, it is possible that Rr may be more sensitive than total and delayed AVLT recall scores to detecting development of tangle pathology and neurodegeneration in the medial-temporal lobe. To note, unlike total and delayed recall, Rr takes into consideration *loss* of information from immediate to delayed recall, as it tracks a difference between performances, and hence may be more attuned to neurodegenerative damage. Consistent with this idea is the previous finding that Rr was sensitive to CSF neurogranin levels, a biomarker of post-synaptic dysfunction

and possible neurodegeneration, in cognitively intact older individuals with major depression (Bruno et al., 2021).

It should be noted that, upon discovering that Rr was associated with CSF p-tau levels, it was not surprising to find that Rr also correlated with CSF t-tau levels. This is because CSF p- and t-tau levels are highly correlated in individuals who do not have dementia (e.g., Van Hulle et al., 2021). Indeed, in our own data, CSF p- and t-tau levels achieved a Pearson's R value of 0.981, indicating extremely high association. Therefore, it would be informative to replicate these analyses in individuals presenting with more pronounced symptoms of a neurodegenerative disease.

The pattern of results observed in this study (i.e., Rr did not predict CSF Aβ42 levels, but did predict CSF p- and t-tau levels) contrasts with that observed in a previous report (Bruno et al., 2019). In Bruno et al. (2019), we observed that Rr (but not total or delayed recall in the AVLT) was sensitive to CSF Aβ42 levels in people with MCI (n=16), but was not sensitive to CSF tau levels in the same group. One difference across these reports is that the sample size of the MCI group in Bruno et al. (2019) was more than twice as large as that in the present study (n=7). It is therefore conceivable that if we were to test a larger sample of people with MCI, all A/T/N biomarkers might be correlated with Rr. Another difference to note, albeit not large, lies in the average ages of the cognitively unimpaired group across papers: they were marginally younger (62.5, SD=9.2) in the Bruno et al.'s paper, where Rr was not found to predict AD CSF biomarkers levels, and older in the present data (65.5, SD = 6.9), although this difference is probably unlikely to have much of an impact on the results. Nevertheless, more

research is needed to elucidate the predictive power of Rr compared to traditional AVLT cognitive scores, with regards to AD biomarkers.

Another AVLT serial position marker that has been used to identify asymptomatic individuals at risk for cognitive decline is primacy (i.e., memory for items at the beginning of a list; e.g., Gicas et al., 2020; LaRue et al., 2008), particularly when measured after a delay (Bruno et al., 2013; Talamonti et al., 2020). Therefore, we also wanted to consider, *post-hoc*, whether delayed primacy had any impact on CSF levels of AD biomarkers in this cohort. We defined primacy as the first four items on the list, as per Bruno et al. (2013), and focused on performance in the delayed trial. We then ran the same regression analyses as above but using delayed primacy as the sole predictor (while maintaining the same covariates). None of the delayed primacy models reached the BFm threshold of "3", and the best result was an anecdotal effect of delayed primacy on p-tau (BFm = 2.072). We can therefore conclude that, within these data, Rr outperformed delayed primacy when predicting CSF levels of AD biomarkers.

A potential limitation of the present paper is that the sample constituted primarily White participants. Evidence from previous studies indicates that different racial and ethnic groups show differences in brain morphology, such as hippocampal volume (DeCarli et al., 2008), white matter hyperintensity volume (Brickman et al., 2008; Divers et al., 2013), and total cerebral brain volume (Stavitsky et al., 2010). Thus, current findings on Rr as a marker of A/T/N biomarkers need to be examined in diverse groups of research participants, and outside of North America, using culturally-appropriate versions of the same tests. Another limitation, partly noted above, is the small number of

individuals with MCI. Small samples are more susceptible to measurement noise, which may also explain the odd observation whereby CSF Aβ42 are higher rather than lower in the MCI groups compared to the other groups. Future tests should address this issue and be carried out with a larger sample. Finally, another possible limitation, which would also probably benefit from a larger sample size, is the fact that delayed recall scores (see Table 1) are much lower in the MCI group compared to the CUS group. This occurrence might make delayed recall less reliable as a predictor.

To summarise, the current results showed that Rr, the ratio between AVLT immediate and delayed performance scores at the recency position is a sensitive measure of CSF levels of p-tau and t-tau. Higher Rr scores, showing disproportionate loss of recency recall from immediate to delayed testing, are associated with an increase in CSF p- and t-tau levels, but not with CSF Aβ42 levels, when controlling for demographics and diagnosis. In contrast, neither the AVLT total or delayed recall scores significantly predicted CSF levels of AD biomarkers. We wish to draw two conclusions. First, as argued previously (Bruno et al., 2016; 2018; 2019; 2021), we suggest that serial position values should be included in databases examining AD and other types of dementia. Second, we posit that Rr is a worthwhile measure to add to the clinician's arsenal (see also Egeland, 2021) when evaluating individuals suspected to be on a trajectory towards neurodegeneration, and, as such, that it should be evaluated also as a possible addition to the MCI diagnostic criteria.

References

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., ... & Phelps, C. H. (2011). 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. *Alzheimer's & dementia*, *7*(3), 270-279.

Bock, J. R., Russell, J., Hara, J., & Fortier, D. Optimizing Cognitive Assessment Outcomes for Alzheimer's Disease by Matching Wordlist Memory Test Features to Scoring Methodology. *Frontiers in Digital Health*, 161.

Brickman, A. M., Schupf, N., Manly, J. J., Luchsinger, J. A., Andrews, H., Tang, M. X., ... & Brown, T. R. (2008). Brain morphology in older African Americans, Caribbean Hispanics, and whites from northern Manhattan. *Archives of neurology*, *65*(8), 1053-1061.

Bruno, D., Gleason, C. E., Koscik, R. L., Pomara, N., Zetterberg, H., Blennow, K., & Johnson, S. C. (2019). The recency ratio is related to CSF amyloid beta 1-42 levels in MCI-AD. *International journal of geriatric psychiatry*, *34*(3), 415-419.

Bruno, D., Koscik, R. L., Woodard, J. L., Pomara, N., & Johnson, S. C. (2018). The recency ratio as predictor of early MCI. *International psychogeriatrics*, *30*(12), 1883-1888.

Bruno, D., Mueller, K. D., Betthauser, T., Chin, N., Engelman, C. D., Christian, B., ... & Johnson, S. C. (2021). Serial position effects in the Logical Memory Test: Loss of primacy predicts amyloid positivity. *Journal of Neuropsychology*, *15*(3), 448-461.

Bruno, D., Reichert Plaska, C., Clark, D. P., Zetterberg, H., Blennow, K., Verbeek, M. M., & Pomara, N. (2021). CSF α-synuclein correlates with CSF neurogranin in late-life depression. *International Journal of Neuroscience*, *131*(4), 357-361.

Bruno, D., Reichert, C., & Pomara, N. (2016). The recency ratio as an index of cognitive performance and decline in elderly individuals. *Journal of Clinical and Experimental Neuropsychology*, *38*(9), 967-973.

Carlesimo, G. A., Sabbadini, M., Fadda, L., & Caltagirone, C. (1995). Different components in word-list forgetting of pure amnesics, degenerative demented and healthy subjects. *Cortex*, *31*(4), 735-745.

Clark, L. R., Koscik, R. L., Nicholas, C. R., Okonkwo, O. C., Engelman, C. D., Bratzke, L. C., ... & Johnson, S. C. (2016). Mild cognitive impairment in late middle age in the

Wisconsin registry for Alzheimer's prevention study: prevalence and characteristics using robust and standard neuropsychological normative data. *Archives of Clinical Neuropsychology*, *31*(7), 675-688.

DeCarli, C., Reed, B. R., Jagust, W. J., Martinez, O., Ortega, M., & Mungas, D. (2008). Brain behavior relationships amongst African Americans, caucasians and Hispanics. *Alzheimer disease and associated disorders*, *22*(4), 382.

Divers, J., Hugenschmidt, C., Sink, K. M., Williamson, J. D., Ge, Y., Smith, S. C., ... & Freedman, B. I. (2013). Cerebral white matter hyperintensity in African Americans and European Americans with type 2 diabetes. *Journal of Stroke and Cerebrovascular Diseases*, *22*(7), e46-e52.

Egeland, J. (2021). Following HN over 21 years: recency change and reduced retention predict later impairment in memory, and recency ratio may combine both effects. *Neurocase*, *27*(2), 147-154.

Florean, I., Penolazzi, B., Menichelli, A., Pastore, M., Cattaruzza, T., Mazzon, G., & Manganotti, P. (2022). Using the ATN system as a guide for the neuropsychological assessment of Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology*, 1-18.

Foldi, N. S., Brickman, A. M., Schaefer, L. A., & Knutelska, M. E. (2003). Distinct serial position profiles and neuropsychological measures differentiate late life depression from normal aging and Alzheimer's disease. *Psychiatry research*, *120*(1), 71-84.

Gicas, K. M., Honer, W. G., Wilson, R. S., Boyle, P. A., Leurgans, S. E., Schneider, J. A., & Bennett, D. A. (2020). Association of serial position scores on memory tests and hippocampal-related neuropathologic outcomes. *Neurology*, *95*(24), e3303-e3312.

Jack Jr, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., ... & Silverberg, N. (2018). NIA-AA research framework: toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia*, *14*(4), 535-562.

Jack, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Feldman, H. H., Frisoni, G. B., ... & Dubois, B. (2016). A/T/N: an unbiased descriptive classification scheme for Alzheimer disease biomarkers. *Neurology*, *87*(5), 539-547.

Johnson, S. C., Koscik, R. L., Jonaitis, E. M., Clark, L. R., Mueller, K. D., Berman, S. E., ... & Sager, M. A. (2018). The Wisconsin Registry for Alzheimer's Prevention: a review of findings and current directions. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *10*, 130-142.

Kaplan, E. (1988). The process approach to neuropsychological assessment. *Aphasiology*, *2*(3-4), 309-311.

Koscik, R. L., La Rue, A., Jonaitis, E. M., Okonkwo, O. C., Johnson, S. C., Bendlin, B. B., ... & Sager, M. A. (2014). Emergence of mild cognitive impairment in late middle-aged adults in the wisconsin registry for Alzheimer's prevention. *Dementia and geriatric cognitive disorders*, *38*(1-2), 16-30.

Koscik, R. L., Berman, S. E., Clark, L. R., Mueller, K. D., Okonkwo, O. C., Gleason, C. E., ... & Johnson, S. C. (2016). Intraindividual cognitive variability in middle age predicts cognitive impairment 8–10 years later: results from the Wisconsin Registry for Alzheimer's Prevention. *Journal of the International Neuropsychological Society*, *22*(10), 1016-1025.

Kruschke, J. (2014). Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan.

Langhough Koscik, R., Hermann, B. P., Allison, S., Clark, L. R., Jonaitis, E. M., Mueller, K. D., ... & Johnson, S. C. (2021). Validity Evidence for the Research Category, "Cognitively Unimpaired–Declining," as a Risk Marker for Mild Cognitive Impairment and Alzheimer's Disease. *Frontiers in Aging Neuroscience*, 404.

Maass, A., Berron, D., Harrison, T. M., Adams, J. N., La Joie, R., Baker, S., ... & Jagust, W. J. (2019). Alzheimer's pathology targets distinct memory networks in the ageing brain. *Brain*, *142*(8), 2492-2509.

Manly, J. J., Jacobs, D. M., Touradji, P., Small, S. A., & Stern, Y. (2002). Reading level attenuates differences in neuropsychological test performance between African American and White elders. *Journal of the International Neuropsychological Society*, *8*(3), 341-348.

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack Jr, C. R., Kawas, C. H., ... & Phelps, C. H. (2011). 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. *Alzheimer's & dementia*, *7*(3), 263-269.

Mueller, K. D., Koscik, R. L., Du, L., Bruno, D., Jonaitis, E. M., Koscik, A. Z., ... & Johnson, S. C. (2020). Proper names from story recall are associated with beta-amyloid in cognitively unimpaired adults at risk for Alzheimer's disease. *Cortex*, *131*, 137-150.

Petersen, R. C., Aisen, P. S., Beckett, L. A., Donohue, M. C., Gamst, A. C., Harvey, D. J., ... & Weiner, M. W. (2010). Alzheimer's disease neuroimaging initiative (ADNI): clinical characterization. *Neurology*, *74*(3), 201-209.

Rey, A. (1958). L'examen clinique en psychologie.

Sager, M. A., Hermann, B., & La Rue, A. (2005). Middle-aged children of persons with Alzheimer's disease: APOE genotypes and cognitive function in the Wisconsin Registry for Alzheimer's Prevention. *Journal of geriatric psychiatry and neurology*, *18*(4), 245-249.

Stavitsky, K., Du, Y., Seichepine, D., Laudate, T. M., Beiser, A., Seshadri, S., ... & Au, R. (2010). White matter hyperintensity and cognitive functioning in the racial and ethnic minority cohort of the Framingham Heart Study. *Neuroepidemiology*, *35*(2), 117-122.

Talamonti, D., Koscik, R., Johnson, S., & Bruno, D. (2020). Predicting early mild cognitive impairment with free recall: The primacy of primacy. *Archives of Clinical Neuropsychology*, *35*(2), 133-142.

Teipel, S. J., Dyrba, M., Ballarini, T., Brosseron, F., Bruno, D., Buerger, K., ... & Heneka, M. T. (2021). Association of cholinergic basal forebrain volume and functional connectivity with markers of inflammatory response in the Alzheimer's disease spectrum. *Journal of Alzheimer's Disease*, (Preprint), 1-16.

Tennant, V.R., Harrison, T.M., Adams, J.N., La Joie, R., Winer, J.R. & Jagust, W.J. (2021). Fusiform Gyrus Phospho-Tau is Associated with Failure of Proper Name Retrieval in Aging. *Annals of Neurology*.

Thomas, K. R., Bangen, K. J., Weigand, A. J., Edmonds, E. C., Wong, C. G., Cooper, S., ... & Alzheimer's Disease Neuroimaging Initiative. (2020). Objective subtle cognitive difficulties predict future amyloid accumulation and neurodegeneration. *Neurology*, *94*(4), e397-e406.

Thomas, K. R., Edmonds, E. C., Eppig, J., Salmon, D. P., Bondi, M. W., & Alzheimer's Disease Neuroimaging Initiative. (2018). Using neuropsychological process scores to identify subtle cognitive decline and predict progression to mild cognitive impairment. *Journal of Alzheimer's Disease*, *64*(1), 195-204.

Turchetta, C. S., De Simone, M. S., Perri, R., Fadda, L., Caruso, G., De Tollis, M., ... & Carlesimo, G. A. (2020). Forgetting Rates on the Recency Portion of a Word List Predict Conversion from Mild Cognitive Impairment to Alzheimer's Disease. *Journal of Alzheimer's Disease*, *73*(4), 1295-1304.

Turchetta, C. S., Perri, R., Fadda, L., Caruso, G., De Simone, M. S., Caltagirone, C., & Carlesimo, G. A. (2018). Forgetting rate on the recency portion of a word list

differentiates mild to moderate Alzheimer’s disease from other forms of dementi. *Journal of Alzheimer's Disease*, *66*(2), 461-470.

Van Hulle, C., Jonaitis, E. M., Betthauser, T. J., Batrla, R., Wild, N., Kollmorgen, G., Andreasson, U., Okonkwo, O., Bendlin, B. B., Asthana, S., Carlsson, C. M., Johnson, S. C., Zetterberg, H., & Blennow, K. (2021). An examination of a novel multipanel of CSF biomarkers in the Alzheimer's disease clinical and pathological continuum. *Alzheimer's & dementia : the journal of the Alzheimer's Association*, *17*(3), 431–445. https://doi.org/10.1002/alz.12204

Wechsler, D. (1987). *WMS-R: Wechsler memory scale-revised*. Psychological Corporation.

Winblad, B., Palmer, K., Kivipelto, M., Jelic, V., Fratiglioni, L., Wahlund, L. O., ... & Petersen, R. C. (2004). Mild cognitive impairment–beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *Journal of internal medicine*, *256*(3), 240-246.

Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annu. Rev. Psychol.*, *55*, 235-269.

Wixted, J. T., & Cai, D. J. (2013). Memory consolidation. *Oxf. Handb. Cogn. Neurosci*, *1*, 436-456.

Figure 1. Plots of the correlations between residuals recency ratio (X-axis) and residuals CSF levels (Y-axis). The residuals were calculated by partialling out all control variables (sex, age at LP, time elapsed, and WRAT-3 raw score.

Figure 1a. Plot of the correlation between recency ratio and CSF p-tau.

Figure 1b. Plot of the correlation between recency ratio and CSF t-tau.

Figure 1c. Plot of the correlation between recency ratio and CSF Aβ42.