Journal of the Neurological Sciences 464 (2024) 123148

Contents lists available at ScienceDirect

Journal of the Neurological Sciences

journal homepage: www.elsevier.com/locate/jns

Story recall performance and AT classification via positron emission tomography: A comparison of logical memory and Craft Story 21

Davide Bruno [a,*], Ainara Jauregi-Zinkunegi [a], Tobey Betthauser [b,c], Cynthia Carlsson [b,c,d,e], Barbara B. Bendlin [c,d], Ozioma Okonkwo [c,d], Nathaniel A. Chin [b,c], Sanjay Asthana [b,c], Rebecca E. Langhough [b,c], Sterling C. Johnson [b,c,d,e], Kimberly D. Mueller [b,c,f]

[a] School of Psychology, Liverpool John Moores University, UK
[b] Wisconsin Alzheimer's Institute, School of Medicine and Public Health, University of Wisconsin – Madison, Madison, WI, USA
[c] Wisconsin Alzheimer's Disease Research Center, School of Medicine and Public Health, University of Wisconsin – Madison, Madison, WI, USA
[d] Department of Medicine, University of Wisconsin—Madison, Madison, WI, USA
[e] Geriatric Research Education and Clinical Center, William S. Middleton Veterans Hospital, Madison, WI, USA
[f] Department of Communication Sciences and Disorders, University of Wisconsin – Madison, Madison, WI, USA

ARTICLE INFO

Keywords:
Tau PET
Amyloid PET
Alzheimer's disease
Serial position
Story recall

ABSTRACT

Background: Early detection of Alzheimer's disease (AD) is one of the critical components of the global response to the growing dementia crisis. Analysis of serial position performance in story recall tests has yielded sensitive metrics for the prediction of AD at low cost. In this study, we examined whether serial position markers in two story recall tests (the logical memory test, LMT, and the Craft Story 21 test, CST) were sensitive to cross-sectional biomarker-based assessment of in vivo neuropathology.

Methods: Participants were selected from the Wisconsin Registry of Alzheimer's Prevention ($n = 288$; WRAP) and the Alzheimer's Disease Research Center ($n = 156$; ADRC), both from the University of Wisconsin–Madison. Average age at PET was 68.9 (6.7) and 67.0 (8.0), respectively. Data included tau and PiB PET, and LMT for WRAP participants and CST for ADRC participants. Two sets of Bayesian analyses (logistic regressions and ANCOVAs) were conducted within each cohort, separately.

Results: Results indicated that the A+T+ classification was best predicted, cross-sectionally, by the recency ratio (Rr), indexing how much of the end of the story was forgotten between initial learning and delayed assessment. Rr outperformed traditional scores and discriminated between A+T+ and A+T−/A-T-, in both cohorts.

Conclusions: Overall, this study confirms that serial position analysis of LMT and CST data, and particularly Rr as an index of recency loss, is a valuable tool for the identification of in vivo tau pathology in individuals free of dementia. Diagnostic considerations are discussed.

1. Introduction

Efforts have been made in recent years to guide the diagnosis of Alzheimer's disease (AD) towards a biological, rather than clinical, framework [1,2,3,4]. According to the most recent working guidelines from the National Institute on Aging and the Alzheimer's Association (https://aaic.alz.org/nia-aa.asp), AD can be diagnosed by in vivo abnormalities of core biomarkers, such as amyloid β or phosphorylated tau. However, biomarker technology and expertise may not always be available at the point of clinical contact, and especially so outside of urban settings [5]. For this reason, access to low cost, easy-to-use tools is vital, provided that such assessments generate reliable associations with core AD biomarkers levels.

Testing neuropsychological function is non-invasive, requires minimal training, is inexpensive, and serves multiple functions outside of diagnosis [6,7]. However, it may seem as though neuropsychological assessment will become increasingly less relevant as biomarker-driven diagnosis develops further. Therefore, we posit here that, as biomarkers constantly evolve, neuropsychological assessments also ought to evolve in step with them. One way to do that is to use neuropsychological assessment to go beyond broad diagnostic applications and focus on the detection of subtle shifts in underlying pathology [8].

* Corresponding author at: Tom Reilly building, Byrom St., Liverpool L3 3AF, UK.
E-mail address: d.bruno@ljmu.ac.uk (D. Bruno).

https://doi.org/10.1016/j.jns.2024.123148
Received 7 June 2024; Received in revised form 5 July 2024; Accepted 20 July 2024
Available online 31 July 2024

Relatedly, process analysis of neuropsychological test performance (or the Boston process approach; [9,10]), proposes that different cognitive processes underlie overall test performance, and that unearthing these processes may be more informative than simply evaluating typical composite scores.

Examples of effective process scores applied to the examination of underlying AD pathology is the analysis of serial position performance. The serial position curve is a common pattern in tests of human memory, where performance tends to be better for stimuli learned at the beginning (primacy) and/or at the end (recency) of a list, as compared to those in the middle (hence the curve shape; e.g., [11]). This recall pattern has been reproduced countless times and has been shown to improve detection of in vivo AD pathology. Loss of primacy recall in stories (i.e., memory for the beginning of a story), for example, was found to predict longitudinal PET amyloid load from an unimpaired baseline [12], while loss of recency (i.e., memory for the end of a story) was found to associate with cerebrospinal tau levels (CSF; [13]). Furthermore, in story recall, both loss of primacy and, to a lesser extent, loss of recency, have been found to cross-sectionally predict biomarker-determined AD [14], as measured by CSF levels of the p-tau/Aβ42 ratio, which combines measures of both proteinopathies, i.e., amyloidosis and tauopathy, and has shown a strong concordance with amyloid PET [15].

All in all, the findings above indicate that serial position markers in story recall are associated with amyloid burden as measured by PET and CSF, and tau as measured in CSF biomarkers, above and beyond conventional clinical scores of the same test – however, it remains to be established whether serial position markers are also sensitive to tau PET burden. New PET ligands, and particularly, ^{18}F-MK-6240 [16], have emerged as sensitive biomarkers of neurofibrillary tangles in the preclinical stages of AD [17]. The favourable imaging properties and spatial distributions shown by ^{18}F-MK-6240 [17] enable the investigation of regional tau-tracer intake, which is crucial, as the locations of early neuropathological changes offer valuable insights into the early clinical characteristics of AD [18]. Thus, examining if serial position scores in story recall are sensitive to elevations of both amyloid and tau PET is useful, especially when assessing individuals who might have started developing early neuropathological changes difficult to detect with other means.

The aims of the current study were two. First, we wished to examine whether serial position markers in story recall were sensitive to cross-sectional PET-based biomarker positivity of amyloid (A) and tau (T) load, as determined by ^{11}C-PiB and ^{18}F-MK-6240, respectively, in individuals free of dementia [19,20]. Second, as we maintain that our serial position-based scoring should be applicable across different tests, we set out to evaluate whether serial position metrics were sensitive to PET A and T loads when using different story recall tests. To achieve this, we employed two commonly used story recall tests, the Logical Memory Task (LMT) and the Craft Story 21 test (CST; see [21]), across different cohorts, to study prediction of biomarker status. Moreover, CST is non-proprietary and, as such, a lower-cost option to clinicians, in compliance with the National Alzheimer's Coordinating Center Uniform Data Set Neuropsychological Battery suggestions [22]. Finally, considering previous findings [12–14], we anticipated that serial position markers, and specifically, higher loss of primacy or recency from immediate learning to delayed testing, would be associated with increased probability of an A+T+ classification, outperforming traditional metrics, and that these markers would also discriminate between AT groups.

2. Methods

Participants for LMT Study: Data were drawn from the Wisconsin Registry of Alzheimer's Prevention (WRAP; University of Wisconsin – Madison). To be included in the analysis, participants had to have measures of both Pittsburgh compound-B (PiB) PET, to assess amyloid distribution volume ratios (DVR), and ^{18}F-MK-6240 PET for tau standardized uptake value ratio (SUVR), alongside story recall data, derived from the LMT [23]. LMT and PET assessments had to be within two years of each other. Participants were excluded if they presented as outliers by having clinically high levels of PET tau load, but not PET amyloid load. In addition, participants had to be classified as cognitively unimpaired (stable or declining) or with mild cognitive impairment at cognitive assessment. Cognitive statuses were assessed using a two-tiered consensus conference approach (for details, see [24,25]), based on core clinical criteria developed by the National Institute on Aging and the Alzheimer's Association [26,27]. All in all, these criteria left us with 288 individuals, of whom 240 were classified as cognitively unimpaired stable, 30 as cognitively unimpaired declining, and 18 with mild cognitive impairment, at the time of cognitive assessment (see Table 1 for more demographics data). All activities for this study were approved by the ethics committees of the authors' universities and competed in accordance with the Declaration of Helsinki. All participants provided informed consent prior to testing.

Participants for CST Study: Data were drawn from the University of Wisconsin – Madison Alzheimer's Disease Research Center (ADRC). The same inclusion and exclusion criteria as above were applied. One difference with WRAP is that in ADRC, participants were classified as cognitively unimpaired as a single classification, or with mild cognitive impairment, at cognitive assessment. Cognitive statuses were also determined by a multi-disciplinary consensus conference review that was blind to AD biomarkers statuses (e.g., PET or CSF data), based on core clinical criteria developed by the National Institute on Aging and the Alzheimer's Association [26,27]. Furthermore, we also removed one extreme outlier based on inspection of q-q plots. These criteria left us with 156 participants, of whom 140 were classified as cognitively unimpaired and 17 with mild cognitive impairment, at the time of cognitive assessment (see Table 2 for more demographics data). All activities for this study were approved by the ethics committees of the authors' universities and competed in accordance with the Declaration of Helsinki. All participants provided informed consent prior to testing.

Memory assessment – LMT. The Logical Memory Task (LMT) is a subtest of the Weschler Memory Scale Revised (WMS-R; [23]), and comprises two stories ("A" and "B"), each with 25 items ("idea units"). Each story is read aloud to the participant and then the participant is asked to recall each story immediately after presentation, and again after a 25–30 min delay. Participants are free to recall the items in any order they prefer. Scoring procedures from the WMS-R manual were applied. Although the scoring criteria permits some alteration from the original item (e.g., "slid off the table" is allowed instead of "fell off the table"), certain items must be recalled verbatim, e.g., numerical expressions or proper names. Two conventional clinical metrics were extracted from LMT (averaging over A and B): LMT Total Immediate recall, derived from the total number of idea units recalled immediately

Table 1
Demographic variables and memory scores by whole sample and AT groups: number of females; education (mean and SD, in years); *APOE* risk score (mean and SD); age at cognitive assessment (mean and SD, in years); elapsed time between assessment and PET (mean and SD, in years); LMT visit number at assessment (median, and min/max); Imm: immediate recall; Del: delayed recall; Rr: recency ratio; Tr: total ratio; Pr: primacy ratio.

	Total (N = 288)	A-T- (*N* = 203)	A+T- (*N* = 48)	A+T+ (*N* = 37)
Gender (Females)	194	136	32	26
Education	16.6 (2.8)	16.5 (2.8)	16.7 (2.7)	17.0 (2.8)
APOE risk score	1.2 (0.8)	1.0 (0.6)	1.6 (0.9)	2.0 (0.9)
Age	68.9 (6.7)	68.0 (6.9)	70.8 (5.5)	71.5 (5.6)
Elapsed time	0.5 (0.5)	0.5 (0.5)	0.6 (0.6)	0.4 (0.5)
LMT visit	6 (1–8)	6 (1–8)	7 (1–8)	7 (1–8)
LMT Total Imm	13.9 (3.5)	14.1 (3.3)	14.1 (3.9)	12.2 (3.6)
LMT Total Del	12.7 (3.9)	13.0 (3.6)	13.1 (4.2)	10.6 (4.5)
LMT Rr	1.1 (0.2)	1.0 (0.2)	1.0 (0.2)	1.2 (0.5)
LMT Tr	1.1 (0.4)	1.1 (0.2)	1.2 (0.5)	1.3 (0.8)
LMT Pr	1.2 (0.3)	1.2 (0.3)	1.2 (0.4)	1.2 (0.4)

after learning the story; and LMT Total Delayed recall, derived from the total number of idea units recalled after the delay. Process scores were primacy ratio (Pr; [12]), recency ratio (Rr; [12]), and total forgetting (or total ratio; Tr; [12]). Primacy and recency were defined as the first and final eight idea units of the story (out of 25), in keeping with previous studies [12]. All three scores were calculated by dividing immediate performance by delayed performance, after applying a+1 correction at each term to compensate for possible 0 scores. Participants' cognitive data were taken from whichever visit was closest to the PET visit.

Memory assessment – CST. The CST [28] comprises a single story assessed immediately and after a delay of approximately 20 min. Responses are scored either verbatim if exactly reported, or paraphrased if the general meaning is captured. We focused on the 25 paraphrased idea items of the story and broke them down into primacy, middle and recency bins following the same mould as with LMT (eight, nine and eight, respectively), as LMT also employs paraphrased responses (see also [21]).

Positron Emission Tomography. Participants underwent PET scans with [C-11] Pittsburgh Compound-B (PiB), acquiring scan data with a dynamic 70-min protocol [29]. Amyloid burden was assessed as a global cortical average PiB distribution volume ratio (DVR) and the threshold for PiB PET positivity was set at PiB $\geq$ 1.19 [30]. ^{18}F-MK-6240 PET standard uptake value ratios (SUVRs) were acquired using a 20-min dynamic protocol, that began 70 min after the bolus injection [19,17]. Tau-positive PET scans were defined by SUVR positivity threshold at 2 standard deviations above the mean of the PiB(−) group in the entorhinal cortex (entorhinal MK-6240 SUVR > 1.27; see [17]).

Four biomarker groups were established (A-T-, A-T+, A+T-, and A+T+) based on combinations of amyloid and tau PET biomarker positivity, yet due to the low number of individuals with A-T+ classification, 14 in WRAP and 6 in ADRC, these were excluded from the analyses. As a result, the final number of individuals included in the analyses was 288 in WRAP and 157 in ADRC. To explore whether serial position metrics of story recall are associated with the probability of an A+T+ classification, participants were initially stratified into two groups: 1) A-T- and A-T+ merged into a single group; and 2) A+T+ as the other group. This decision to merge participants followed previous reports showing that faster cognitive decline occurs in individuals with both elevated amyloid and tau biomarkers, compared to those with one or no elevated biomarkers [17,31].

Genotyping. DNA was extracted from whole blood. Samples were aliquoted on 96-well plates for determination of *APOE* genotypes. An *APOE* risk score was calculated based on the odds ratios of the ε2/ε3/ε4 genotype, as previously reported [27].

Analysis plan. For each story recall test in its corresponding sample (LMT in WRAP, CST in ADRC), the same statistical analyses were conducted separately, First, we carried out cross-sectional Bayesian logistic regression analyses with AT classification as outcome (A-T- and A+T- vs. and A+T+). Predictors were traditional and serial position metrics from LMT or CST, depending on cohort, and control variables were years of education, gender, *APOE* risk score, age at cognitive assessment, visit number to account for practice effects, and elapsed time between cognitive assessment and tau PET. This analysis was carried out to identify the best predictors of AT outcomes. Credible intervals (CIs) were set to 95%. The prior was set to JZS, and the model prior was set to Uniform. One thousand Markov chain Monte-Carlo simulations were conducted to determine parameters and compensate from possible violations of normality. Second, we carried out Bayesian ANCOVAs with the same covariates; AT classification as independent variable (this time with three separate groups: A-T-, A+T- and A+T+); and whichever sensitive LMT/CST metrics emerged from the initial regression as dependent variables, in separate analyses. Model priors were set to Uniform. Analyses were conducted using JASP (0.18.3; https://jasp-stats.org/).

3. Results

3.1. Logical memory test in WRAP

Table 1 reports demographic variables and memory scores by whole sample (WRAP) and AT group.

Bayesian logistic regression. The logistic analysis (37 of 288 were classified as positive) yielded a best fitting model (BF_{10} = 168.144, extreme evidence) with two predictors: LMT Total Immediate recall ($BF_{inclusion}$ = 1.46) and LMT Rr ($BF_{inclusion}$ = 2.65). $BF_{inclusion}$ scores show that model odds increase when including either variable 1.5 and 2.7 times, respectively. More LMT Total Immediate recall was associated with lower risk of an A+T+ classification (mean coefficient = −0.066, SD = 0.104, CIs −0.323 to 0.062), and higher LMT Rr scores, indicating more recency forgetting, were associated with more risk of an A+T+ classification (mean coefficient = 1.059, SD = 1.100, CIs −0.381 to 3.514).

Bayesian ANCOVAs. Following on from the logistic regression, we carried out two ANCOVAs with LMT Total Immediate recall and LMT Rr as dependent variables, separately. There was moderate evidence that LMT Total Immediate recall was influenced by the AT classification (BF_{10} = 5.957; moderate evidence). As shown in Fig. 1 (left), LMT Total Immediate recall was highest for A-T-, in the middle for A+T-, and lowest for A+T+. Post-hoc comparisons showed that LMT Total Immediate recall levels discriminated successfully between A+T+ and A-T- (BF_{10} = 29.375), but not as well between A+T+ and A+T- (BF_{10} = 2.571) or between A-T- and A+T- (BF_{10} = 0.173). Analogously, the AT classification was found to affect Rr (BF_{10} = 35.345; strong evidence). As shown in Fig. 1 (right), LMT Rr was lowest for A+T-, in the middle for A-T-, and highest for A+T+. Post-hoc comparisons showed that LMT Rr discriminated between A+T+ and both A+T- (BF_{10} = 3.035) and A-T- (BF_{10} = 120.210), but not between A-T- and A+T- (BF_{10} = 0.336).

Interim summary. The analysis on WRAP data shows that LMT Total Immediate recall and LMT Rr both provide sensitive measures of the difference between individuals classified as A+T+ and people classified as either A-T- or A+T-.

3.2. Craft Story 21 in ADRC

Table 2 reports demographic variables and memory scores by whole sample (ADRC) and AT group.

Bayesian logistic regression. The logistic analysis (25 of 157 were classified as positive) yielded a best fitting model (BF_{10} = 6.667, moderate evidence) with CST Rr as the only predictor ($BF_{inclusion}$ = 6.240). Higher CST Rr scores (mean coefficient = 1.727, SD = 1.480, CIs −0.578 to 4.603) were associated with more A+T+ risk.

Bayesian ANCOVAs. Following on from the logistic regression, we focused only on CST Rr for the ANCOVA. There was moderate evidence that CST Rr was influenced by the AT classification (BF_{10} = 5.217). As shown in Fig. 2, CST Rr was lowest for A+T-, in the middle for A-T-, and

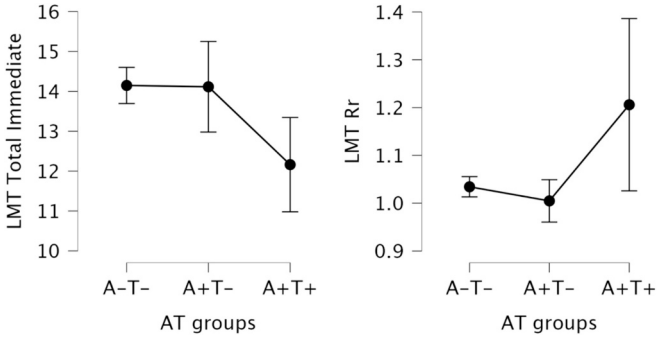

Fig. 1. Means and error variance of LMT Total Immediate recall (left) and Rr (right) scores by AT classification.

Table 2
Demographic variables and memory scores by whole sample and AT groups: number of females; education (mean and SD, in years); *APOE* risk score (mean and SD); age at cognitive assessment (mean and SD, in years); elapsed time between assessment and PET (mean and SD, in years); CST visit number at assessment (median, and min/max); Imm para: immediate paraphrased recall; Del para: delayed paraphrased recall; Rr: recency ratio; Tr: total ratio; Pr: primacy ratio.

	Total (N = 156)	A-T- (*N* = 115)	A+T- (*N* = 17)	A+T+ (*N* = 24)
Gender (Females)	105	77	13	14
Education	16.2 (2.4)	16.2 (2.4)	16.9 (2.5)	15.5 (2.5)
APOE risk score	1.3 (0.9)	1.0 (0.7)	1.6 (1.0)	2.3 (1.0)
Age	67.0 (8.0)	65.5 (7.8)	69.1 (7.2)	72.5 (7.3)
Elapsed time	0.4 (0.3)	0.4 (0.4)	0.2 (0.4)	0.3 (0.2)
CST visit	6 (1–13)	5 (1–13)	6 (2–11)	6 (1–13)
CST Imm para	16.8 (4.6)	17.4 (4.2)	16.6 (6.0)	14.3 (4.3)
CST Del para	15.9 (4.9)	16.6 (4.4)	15.4 (5.9)	12.8 (5.0)
CST Rr	1.1 (0.3)	1.0 (0.2)	1.0 (0.1)	1.1 (1.3)
CST Tr	1.1 (0.5)	1.1 (0.1)	1.1 (0.1)	1.4 (1.2)
CST Pr	1.1 (0.4)	1.1 (0.4)	1.1 (0.2)	1.2 (0.6)

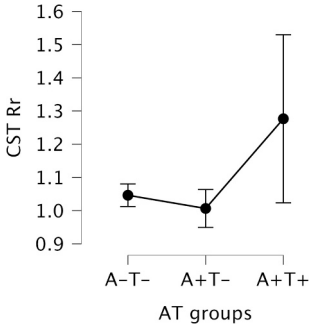

Fig. 2. Means and error variance of CST Rr scores by AT classification.

highest for A+T+. Post-hoc comparisons showed that CST Rr was higher for A+T+ than for A-T- ($BF_{10} = 37.580$; very strong evidence), but there were no differences between A-T- and A+T- ($BF_{10} = 0.360$) or between A+T- and A+T+ ($BF_{10} = 1.137$). Fig. 2 reports the means and error variance for CST Rr, by AT group.

Interim summary. The analysis on ADRC data shows that CST Rr provides comparatively the most sensitive measure of the difference between individuals classified as A+T+ and people classified as either A-T- or A+T-.

4. Discussion

In this study, we set out to examine whether serial position markers in story recall were sensitive to cross-sectional biomarker-based assessment of in vivo neuropathology, as determined by ^{18}F-MK-6240 and ^{11}C-PiB PET imaging, and to compare this sensitivity across two commonly used story recall tests: the logical memory test and the Craft Story 21 test. Our results, across two cohorts, confirmed that serial position analysis of story recall data, and particularly Rr as an index of recency loss, is a valuable tool for the identification of in vivo neurodegenerative pathology in older individuals free of dementia.

The present study provides novel evidence on the usefulness of serial position markers derived from story recall, especially Rr, in detecting cross-sectional PET-based biomarker positivity of amyloid and tau in older individuals free of dementia. Specifically, the analyses with CST showed that Rr was the most sensitive measure to the difference between individuals classified as A+T+ and those classified as either A-T- or A+T-, outperforming CST traditional metrics. In analyses with LMT, the best model was a combination of Rr and LMT Total immediate recall, but Rr was comparatively the stronger association. Altogether, results indicate that higher Rr scores, derived from either LMT or CST, are associated with higher odds of A+T+ classification in older adults free of dementia.

Previous studies have shown that a faster cognitive decline occurs in individuals with both elevated amyloid and tau biomarkers, compared to those with one or no elevated biomarkers [17,31]. Considering amyloid-β and tau accumulations begin years before clinical impairment, we believe it was necessary to examine whether serial position markers also discriminate between individuals with different AT classifications. Current analyses showed that in LMT, Rr discriminated between A+T+ and both A+T- and A-T-, while in CST, Rr discriminated between A+T+ and A+T-. Overall, these findings suggest that Rr, derived from either story recall test, might provide useful clinical information in older adults.

Analyses on clinical cut offs were consistent with these findings. When setting Rr at 2, in both LMT-WRAP and CST-ADRC, positive predictive values (PPV) and negative predictive values (NPV) were high: 100% and 88%, respectively, for LMT; and 100% and 86%, respectively, for CST. These figures suggest that a positive test result is always a true positive (hit) rather than a false positive (false alarm) and that a negative test result is mostly due to a true negative (correct rejection) rather than a false negative (miss). However, the low prevalence of amyloid and tau PET positivity, likely due to the fact that participants were free of dementia at cognitive assessment, makes the NPV generally more reliable than the PPV [32]. As such, we also examined the prevalence-free measures specificity (how likely we are to detect an AT- person with the test) and sensitivity (how likely we are to detect an AT+ person with the test). Again, specificity was high in both tests (100% for both), but this time sensitivity was low (5% and 8%, respectively). Therefore, we can conclude from our results that, with either test when applying Rr, a negative score (i.e., a score below 2) should give us confidence that the person will not be classified as A+T+ based on PET imaging. Hence, if pre-selecting for further examination, participants with Rr scores at 2 or above will be more likely to yield positive results.

While the exact neurobiological mechanisms linking loss of recency to tau load are yet to be fully elucidated, our present findings are consistent with previous reports. Rr has been shown, in fact, both with word-lists tests and LMT, to be a valuable predictor of AD biomarkers [13], and tauopathy in particular [33]. We have previously suggested that loss of primacy, i.e., the early portion of a list or story, may be more sensitive specifically to changes in brain amyloid deposition, as proposed in Bruno et al. [12,14], possibly due to primacy performance being more sensitive to associative memory functions and related associative cortex, where amyloid deposition is most common (e.g., [34]). In contrast, Rr appears to be more responsive to tau-related neuronal damage, which is typically associated with medial-temporal lobe areas. However, these points are still largely a matter of speculation and further investigation, examining more closely neurocognitive activity with brain imaging, would be required to address these ideas. Nevertheless, the present study shows that Rr is sensitive to cross-sectional PET-based biomarker positivity of amyloid and tau load, not only when derived from the LMT, but also when deriving Rr from a different story recall test, CST, in a different cohort.

As noted, females represented the majority of our sample (69%), but post hoc gender-differential secondary analyses showed that associations between Rr and AT classification were independent of gender (see also [35]). Unlike what we previously reported with primacy (see [36]), it is possible that recency forgetting may be less gender-dependent and, as such, a potentially better cognitive marker in mixed-gender contexts.

The choice of limiting the analyses exclusively on individuals free of dementia was both theoretically motivated, as well as practical. We think that finding sensitive cognitive markers for people who are showing no obvious signs of dementia is valuable because, in the first instance, it would help reassure those individuals who worry they may be on a trajectory to AD but are otherwise well – and a high negative predictive value, as we report, is useful for that purpose –, and second, it

gives us screening measures that can be applied even without a broader clinical context – to this point, we also controlled for demographics and genetic risk factors, further consolidating this conclusion. However, as mentioned, this was also a practical choice because the number of individuals with dementia in this cohort was small, thus making it impossible to conduct meaningful separate analyses on these individuals, and otherwise making the overall cohort heterogeneous. Nevertheless, the lack of individuals with dementia in this study is also a limitation, as knowing how these predictors behave within more clearly clinically defined groups would be very informative.

Another limiting factor of this study is that the participants were overwhelmingly white Caucasians. We agree with a building international consensus that emphasises the importance of including more diversity within AD research (e.g., [37]): ethnic diversity, but also diversity based on language, culture, and economic background.

In summary, this study of two different cohorts showed that serial position markers of story recall, and particularly Rr as an index of recency loss, are associated with cross-sectional PET-based biomarker positivity of amyloid and tau. Specifically, Rr, derived from either LMT or CST, was the strongest predictor of A+T+ classification, outperforming traditional LMT and CST scores. The present study also showed that Rr from both tests, discriminates between A+T+ and A+T- individuals. We believe that serial position analysis of story recall data, and particularly Rr, may be a valuable tool for the identification of in vivo tau pathology.

Author notes

We wish to thank all WRAP and ADRC participants.

This secondary analysis of WRAP and ADRC data was funded by a NIH-NIA (R01 AG070940-01) grant to KDM, in which DB and REL are co-investigators.

The full results from the Bayesian regression analysis and the R code used in JASP for these analyses are available upon request.

The ethical regulations that govern WRAP and ADRC prevent unrestricted public archiving of anonymised study data. Data can be requested from the respective Executive Committees at: https://wrap.wisc.edu/data-requests-2/ and https://www.adrc.wisc.edu/apply-resources. Data will be released to internal and external investigators following confirmation of IRB approval together with an evaluation by WRAP and ADRC of scientific merit and resource availability.

No author reports any conflicts of interests or disclosures.

CRediT authorship contribution statement

Davide Bruno: Writing – original draft, Funding acquisition, Formal analysis, Conceptualization. **Ainara Jauregi-Zinkunegi:** Writing – review & editing, Formal analysis, Data curation. **Tobey Betthauser:** Methodology. **Cynthia Carlsson:** Funding acquisition. **Barbara B. Bendlin:** Funding acquisition. **Ozioma Okonkwo:** Funding acquisition. **Nathaniel A. Chin:** Data curation. **Sanjay Asthana:** Project administration, Funding acquisition. **Rebecca E. Langhough:** Funding acquisition, Conceptualization. **Sterling C. Johnson:** Project administration, Funding acquisition. **Kimberly D. Mueller:** Funding acquisition, Conceptualization.

References

[1] G. Chételat, J. Arbizu, H. Barthel, V. Garibotto, I. Law, S. Morbelli, A. Drzezga, Amyloid-PET and 18F-FDG-PET in the diagnostic investigation of Alzheimer's disease and other dementias, Lancet Neurol. 19 (11) (2020) 951–962.

[2] G.B. Frisoni, M. Boccardi, F. Barkhof, K. Blennow, S. Cappa, K. Chiotis, B. Winblad, Strategic roadmap for an early diagnosis of Alzheimer's disease based on biomarkers, Lancet Neurol. 16 (8) (2017) 661–676.

[3] I. Illán-Gala, J. Pegueroles, V. Montal, E. Vilaplana, M. Carmona-Iragui, D. Alcolea, Alzheimer's Disease Neuroimaging Initiative, Challenges associated with biomarker-based classification systems for Alzheimer's disease, Alzheimer's & Dementia 10 (2018) 346–357.

[4] C. Porteri, E. Albanese, C. Scerri, M.C. Carrillo, H.M. Snyder, B. Martensson, for the Roadmap, G. T. F, The biomarker-based diagnosis of Alzheimer's disease. 1—ethical and societal issues, Neurobiol. Aging 52 (2017) 132–140.

[5] V. Manera, E. Rovini, P. Wais, Early detection of neurodegenerative disorders using behavioral markers and new technologies: new methods and perspectives, Front. Aging Neurosci. (2023) 15, https://doi.org/10.3389/fnagi.2023.1149886.

[6] P.D. Harvey, Clinical applications of neuropsychological assessment, Dialogues Clin. Neurosci. 14 (2012) 91–99.

[7] P.D. Harvey, Clinical applications of neuropsychological assessment, Dialogues Clin. Neurosci. 14 (1) (2012) 91–99.

[8] K.D. Mueller, L. Du, D. Bruno, T. Betthauser, B. Christian, S. Johnson, R.L. Koscik, Item-level story recall predictors of amyloid-beta in late middle-aged adults at increased risk for Alzheimer's disease, Front. Psychol. 13 (2022) 908651.

[9] D.J. Libon, R. Swenson, M. Lamar, C.C. Price, G. Baliga, A. Pascual-Leone, R. Au, S. Cosentino, S.L. Andersen, The Boston process approach and digital neuropsychological assessment: past research and future directions, J. Alzheimer's Dis. 87 (4) (2022) 1419–1432, https://doi.org/10.3233/JAD-220096.

[10] W.P. Milberg, N. Hebben, E. Kaplan, I. Grant, K. Adams, The Boston process approach to neuropsychological assessment, Neuropsychol. Assess. Neuropsychiatr. Neuromed. Disord. 3 (2009) 42–65.

[11] B.B. Murdock Jr., The serial position effect of free recall, J. Exp. Psychol. 64 (5) (1962) 482–488, https://doi.org/10.1037/h0045106.

[12] D. Bruno, K.D. Mueller, T. Betthauser, N. Chin, C.D. Engelman, B. Christian, S. C. Johnson, Serial position effects in the logical memory test: loss of primacy predicts amyloid positivity, J. Neuropsychol. 15 (3) (2021) 448–461.

[13] D. Bruno, A.J. Zinkunegi, G. Kollmorgen, I. Suridjan, N. Wild, C. Carlsson, K. D. Mueller, The recency ratio assessed by story recall is associated with cerebrospinal fluid levels of neurodegeneration biomarkers, Cortex 159 (2023) 167–174.

[14] D. Bruno, A. Jauregi Zinkunegi, G. Kollmorgen, M. Carboni, N. Wild, C. Carlsson, K. D. Mueller, A comparison of diagnostic performance of word-list and story recall tests for biomarker-determined Alzheimer's disease, J. Clin. Exp. Neuropsychol. (2023) 1–7.

[15] M.R. Campbell, S. Ashrafzadeh-Kian, R.C. Petersen, M.M. Mielke, J.A. Syrjanen, A. C. van Harten, A. Algeciras-Schimnich, P-tau/aβ42 and aβ42/40 ratios in csf are equally predictive of amyloid pet status, Alzheimer's & Dementia 13 (1) (2021) e12190.

[16] E.D. Hostetler, A.M. Walji, Z. Zeng, P. Miller, I. Bennacef, C. Salinas, J.L. Evelhoch, Preclinical characterization of 18F-MK-6240, a promising PET tracer for in vivo quantification of human neurofibrillary tangles, J. Nucl. Med. 57 (10) (2016) 1599–1606.

[17] T.J. Betthauser, R.L. Koscik, E.M. Jonaitis, S.L. Allison, K.A. Cody, C.M. Erickson, S. C. Johnson, Amyloid and tau imaging biomarkers explain cognitive decline from late middle-age, Brain 143 (1) (2020) 320–335.

[18] A.P. Mecca, M.K. Chen, R.S. O'Dell, M. Naganawa, T. Toyonaga, T.A. Godek, C. H. van Dyck, Association of entorhinal cortical tau deposition and hippocampal synaptic density in older individuals with normal cognition and early Alzheimer's disease, Neurobiol. Aging 111 (2022) 44–53.

[19] T.J. Betthauser, K.A. Cody, M.D. Zammit, D. Murali, A.K. Converse, T.E. Barnhart, B.T. Christian, In vivo characterization and quantification of neurofibrillary tau PET radioligand 18F-MK-6240 in humans from Alzheimer disease dementia to young controls, J. Nucl. Med. 60 (1) (2019) 93–99.

[20] S.S. Simon Simon, E. Varangis, S. Lee, Y. Gu, Y. Gazes, Q.R. Razlighi, C. Habeck, Y. Stern, *In vivo* tau is associated with change in memory and processing speed, but not reasoning, in cognitively unimpaired older adults, Neurobiol. Aging 133 (2024) 28–38.

[21] C.O. Nester, J. Qin, C. Wang, M.J. Katz, R.B. Lipton, L.A. Rabin, Concordance between logical memory and craft story 21 in community-dwelling older adults: the role of demographic factors and cognitive status, Arch. Clin. Neuropsychol. 38 (7) (2023) 1091–1105.

[22] S. Weintraub, L. Besser, H.H. Dodge, M. Teylan, S. Ferris, F.C. Goldstein, B. Giordani, J. Kramer, D. Loewenstein, D. Marson, D. Mungas, Version 3 of the Alzheimer Disease Centers' neuropsychological test battery in the Uniform Data Set (UDS), Alzheimer Disease & Associated Disorders 32 (1) (2018) 10–17.

[23] D. Wechsler, WMS-R: Wechsler Memory Scale-Revised, Psychological Corporation, 1987.

[24] S.C. Johnson, R.L. Koscik, E.M. Jonaitis, L.R. Clark, K.D. Mueller, S.E. Berman, B. B. Bendlin, C.D. Engelman, O.C. Okonkwo, K.J. Hogan, S. Asthana, C.M. Carlsson, B.P. Hermann, M.A. Sager, The Wisconsin registry for Alzheimer's prevention: a review of findings and current directions, Alzheimers Dement. 10 (2018) 130–142, https://doi.org/10.1016/j.dadm.2017.11.007.

[25] R. Langhough Koscik, B.P. Hermann, S. Allison, L.R. Clark, E.M. Jonaitis, K. D. Mueller, S.C. Johnson, Validity evidence for the research category,"cognitively unimpaired–declining," as a risk marker for mild cognitive impairment and Alzheimer's disease, Front. Aging Neurosci. 13 (2021) 688478.

[26] M.S. Albert, S.T. DeKosky, D. Dickson, B. Dubois, H.H. Feldman, N.C. Fox, A. Gamst, D.M. Holtzman, W.J. Jagust, R.C. Petersen, P.J. Snyder, M.C. Carrillo, B. Thies, C.H. Phelps, 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. 7 (3) (2011) 270e279.

[27] G.M. McKhann, D.S. Knopman, H. Chertkow, B.T. Hyman, C.R. Jack, C.H. Kawas, W.E. Klunk, W.J. Koroshetz, J.J. Manly, R. Mayeux, R.C. Mohs, J.C. Morris, M. N. Rossor, P. Scheltens, M.C. Carrillo, B. Thies, S. Weintraub, C.H. Phelps, 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 Dement. 7 (3) (2011) 263–269, https://doi.org/10.1016/j.jalz.2011.03.005.
[28] S. Craft, J. Newcomer, S. Kanne, S. Dagogo-Jack, P. Cryer, Y. Sheline, A. Alderson, Memory improvement following induced hyperinsulinemia in Alzheimer's disease, Neurobiol. Aging 17 (1) (1996) 123–130.
[29] S.C. Johnson, B.T. Christian, O.C. Okonkwo, J.M. Oh, S. Harding, G. Xu, M. A. Sager, Amyloid burden and neural function in people at risk for Alzheimer's disease, Neurobiol. Aging 35 (3) (2014) 576–584.
[30] A.M. Racine, L.R. Clark, S.E. Berman, R.L. Koscik, K.D. Mueller, D. Norton, S. C. Johnson, Associations between performance on an abbreviated CogState battery, other measures of cognitive function, and biomarkers in people at risk for Alzheimer's disease, J. Alzheimers Dis. 54 (4) (2016) 1395–1408.
[31] R.A. Sperling, E.C. Mormino, A.P. Schultz, R.A. Betensky, K.V. Papp, R. E. Amariglio, K.A. Johnson, The impact of amyloid-beta and tau on prospective cognitive decline in older individuals, Ann. Neurol. 85 (2) (2019) 181–193.
[32] P. Ranganathan, R. Aggarwal, Common pitfalls in statistical analysis: understanding the properties of diagnostic tests–part 1, Perspect. Clin. Res. 9 (1) (2018) 40.
[33] D. Bruno, A. Jauregi Zinkunegi, N. Pomara, H. Zetterberg, K. Blennow, R.L. Koscik, C. Carlsson, B. Bendlin, O. Okonkwo, B.P. Hermann, S.C. Johnson, K.D. Mueller, Cross-sectional associations of CSF tau levels with Rey's AVLT: a recency ratio study, Neuropsychology 37 (6) (2023) 628–635, https://doi.org/10.1037/neu0000821.
[34] M.J. Grothe, S.J. Teipel, Alzheimer's Disease Neuroimaging Initiative, Spatial patterns of atrophy, hypometabolism, and amyloid deposition in Alzheimer's disease correspond to dissociable functional brain networks, Hum. Brain Mapp. 37 (1) (2016) 35–53.
[35] A. Jauregi Zinkunegi, D. Bruno, T.J. Betthauser, R. Langhough, S. Asthana, N. A. Chin, K.D. Mueller, A comparison of story-recall metrics to predict hippocampal volume in older adults with and without cognitive impairment, Clin. Neuropsychol. (2023) 1–18.
[36] D. Bruno, K.M. Gicas, A. Jauregi Zinkunegi, K.D. Mueller, M. Lamar, Delayed primacy recall performance predicts post mortem Alzheimer's disease pathology from unimpaired ante mortem cognitive baseline, Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring 16 (1) (2023) e12524.
[37] S. Franzen, J.E. Smith, E. van den Berg, M. Rivera Mindt, R.L. van Bruchem-Visser, E.L. Abner, L.S. Schneider, N.D. Prins, G.M. Babulal, J.M. Papma, Diversity in Alzheimer's disease drug trials: the importance of eligibility criteria, Alzheimers Dement. 18 (4) (2022) 810–823.