Routledge
Taylor & Francis Group

Journal of Clinical and Experimental Neuropsychology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ncen20

A comparison of diagnostic performance of word-list and story recall tests for biomarker-determined Alzheimer's disease

Davide Bruno, Ainara Jauregi Zinkunegi, Gwendlyn Kollmorgen, Margherita Carboni, Norbert Wild, Cynthia Carlsson, Barbara Bendlin, Ozioma Okonkwo, Nathaniel Chin, Bruce P. Hermann, Sanjay Asthana, Kaj Blennow, Rebecca Langhough, Sterling C. Johnson, Nunzio Pomara, Henrik Zetterberg & Kimberly D. Mueller

To cite this article: Davide Bruno, Ainara Jauregi Zinkunegi, Gwendlyn Kollmorgen, Margherita Carboni, Norbert Wild, Cynthia Carlsson, Barbara Bendlin, Ozioma Okonkwo, Nathaniel Chin, Bruce P. Hermann, Sanjay Asthana, Kaj Blennow, Rebecca Langhough, Sterling C. Johnson, Nunzio Pomara, Henrik Zetterberg & Kimberly D. Mueller (2023): A comparison of diagnostic performance of word-list and story recall tests for biomarker-determined Alzheimer's disease, Journal of Clinical and Experimental Neuropsychology, DOI: 10.1080/13803395.2023.2240060

To link to this article: https://doi.org/10.1080/13803395.2023.2240060

JOURNAL OF CLINICAL AND EXPERIMENTAL NEUROPSYCHOLOGY
https://doi.org/10.1080/13803395.2023.2240060

Routledge
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

A comparison of diagnostic performance of word-list and story recall tests for biomarker-determined Alzheimer's disease

Davide Bruno[a], Ainara Jauregi Zinkunegi[a], Gwendlyn Kollmorgen[b], Margherita Carboni[c], Norbert Wild[b], Cynthia Carlsson[d,e,f,g], Barbara Bendlin[e,f], Ozioma Okonkwo[e,f], Nathaniel Chin[d,e], Bruce P. Hermann[d,h], Sanjay Asthana[d,e], Kaj Blennow[i,j], Rebecca Langhough[d,e,f], Sterling C. Johnson[d,e,f,g], Nunzio Pomara[k,l], Henrik Zetterberg[e,i,j,m,n,o] and Kimberly D. Mueller[d,e,p]

[a]School of Psychology, Liverpool John Moores University, UK; [b]Roche Diagnostics GmbH, Penzberg, Germany; [c]Roche Diagnostics International Ltd, Rotkreuz, Switzerland; [d]Wisconsin Alzheimer's Institute, School of Medicine and Public Health, University of Wisconsin – Madison, Madison, WI, USA; [e]Wisconsin Alzheimer's Disease Research Center, School of Medicine and Public Health, University of Wisconsin – Madison, Madison, WI, USA; [f]Department of Medicine, University of Wisconsin-Madison, Madison, WI, USA; [g]Geriatric Research Education and Clinical Center, William S. Middleton Veterans Hospital, Madison, WI, USA; [h]Department of Neurology, University of Wisconsin – Madison, Madison, WI, USA; [i]Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, the Sahlgrenska Academy at the University of Gothenburg, Mölndal, Sweden; [j]Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, Mölndal, Sweden; [k]Geriatric Psychiatry Division, Nathan Kline Institute, Orangeburg, NY, USA; [l]School of Medicine, New York University, New York, USA; [m]Department of Neurodegenerative Disease, UCL Institute of Neurology, London, UK; [n]UK Dementia Research Institute at UCL, London, UK; [o]Hong Kong Center for Neurodegenerative Diseases, Clear Water Bay, Hong Kong, China; [p]Department of Communication Sciences and Disorders, University of Wisconsin – Madison, Madison, WI, USA

ABSTRACT

Background: Wordlist and story recall tests are routinely employed in clinical practice for dementia diagnosis. In this study, our aim was to establish how well-standard clinical metrics compared to process scores derived from wordlist and story recall tests in predicting biomarker determined Alzheimer's disease, as defined by CSF ptau/Aβ42 ratio.

Methods: Data from 295 participants (mean age = 65 ± 9.) were drawn from the University of Wisconsin – Madison Alzheimer's Disease Research Center (ADRC) and Wisconsin Registry for Alzheimer's Prevention (WRAP). Rey's Auditory Verbal Learning Test (AVLT; wordlist) and Logical Memory Test (LMT; story) data were used. Bayesian linear regression analyses were carried out with CSF ptau/Aβ42 ratio as outcome. Sensitivity analyses were carried out with logistic regressions to assess diagnosticity.

Results: LMT generally outperformed AVLT. Notably, the best predictors were primacy ratio, a process score indexing loss of information learned early during test administration, and recency ratio, which tracks loss of recently learned information. Sensitivity analyses confirmed this conclusion.

Conclusions: Our study shows that story recall tests may be better than wordlist tests for detection of dementia, especially when employing process scores alongside conventional clinical scores.

ARTICLE HISTORY
Received 30 January 2023
Accepted 18 July 2023

KEYWORDS
Memory; dementia; story recall; biomarkers; serial position

Introduction

Early detection of neurodegeneration is critical for clinical and research decision making in dementia (Trevethan, 2017). Biomarker research has seen impressive progress of late, allowing for the identification of individuals who are yet to present with clinical symptoms, while showing emerging neuropathology. However, biomarker-based screening, particularly when relying on positron-emission tomography and/or lumbar puncture, can be intimidating, and requires access to highly specialized clinical settings (Manera et al., 2023). As over 60% of people living with dementia are currently in low-to-middle income countries (World Health Organization, 2023), access to affordable, but accurate, screening measures becomes critical.

Testing neuropsychological function is noninvasive, requires minimal training, and is inexpensive. In particular, loss of episodic memory ability is a key feature of Alzheimer's disease (AD) and related dementias symptomatology (Albert et al., 2011; Dubois et al., 2007; De Simone et al., 2019; De Tollis et al., 2021). Episodic memory ability is most commonly assessed with verbal recall tests, and typically these are either word-list or story recall tests (Mansbach et al., 2014; Perri et al., 2013; De Simone et al., 2017). Word-list tests of memory recall employ a list of (semantically unrelated or related)

CONTACT Davide Bruno d.bruno@ljmu.ac.uk Tom Reilly building, Byrom St, Liverpool L3 3AF, United Kingdom

We wish to thank all WRAP and ADRC participants
Supplemental data for this article can be accessed online at https://doi.org/10.1080/13803395.2023.2240060

common nouns which are read sequentially to a person, who, after the list has been read in full, will be asked to repeat all the words they can remember, in any order. In contrast, story recall tests will present a person with one or more stories, which will then be expected to be recalled, again without demands on the order of recall.

Previous research has shown that both word list and story recall tests are useful tools in the diagnosis of dementia and AD (Lemos et al., 2014; Perri et al., 2013; Teichmann et al., 2017; Turchetta et al., 2018). While some studies favor the former (Baek et al., 2012), and others prefer the latter (Park et al., 2017; De Simone et al., 2017), Mansbach et al. (2014) concluded that both types of tests should be included whenever possible, and provided evidence that combining the two yielded a stronger predictor of cognitive diagnosis than using either test alone.

Alongside standard test scores, which typically represent the sum of the units recalled correctly, process scores are also frequently examined (Libon et al., 2013; Milberg et al., 2009). Analysis of process scores, an exemplar being the Boston Process Approach, is based upon the principle that different cognitive processes underlie overall test performance, and that unearthing these processes may be more informative than simply evaluating typical composite scores. An example of a process score applied to story recall comes from De Simone et al. (2017), where the authors reported that forgetting story information over time associated strongly with more elevated risk, and with faster conversion times to a clinical diagnosis of AD. Similarly, Bruno et al. (2018) examined forgetting in word list tests, but focused on the final portion of the list (i.e., recency position). They showed that examining recency forgetting improved longitudinal prediction of early mild cognitive impairment (MCI; see also Bruno et al., 2022). Bruno et al. (2013; see also Talamonti et al., 2019) also proposed that loss of primacy (i.e., information learned early on a list or story), particularly after a delay, may indicate an issue of consolidation. Consistent with this notion, they showed that delayed primacy (as measured with delayed word list recall) was a better predictor of global cognitive decline, compared to the other portions of the serial position curve and standard metrics; and that primacy forgetting was associated to AD neuropathology (Bruno et al., 2021).

In recent years, the field of AD research has seen a push to embrace diagnoses that relied on the analysis of biomarker data, as opposed to clinical symptoms (Porteri et al., 2017), albeit not without controversy (Illán-Gala et al., 2018). Consequently, we noticed that the scientific literature was lacking with regards to examining how sensitive word-list and story recall tests were to biomarker-determined AD (bdAD). One advantage of using bdAD as testing grounds for cognitive assessment is that it avoids potential issues of circularity between memory testing and clinical diagnoses, given that the latter will often rely upon the former. In this study, therefore, we aimed to test how sensitive standard and process scores derived from word-list and story recall tests were to bdAD. We employed the cerebro-spinal fluid (CSF) p-tau/Aβ42 ratio (Salvadó et al., 2022), either continuously or as a cutoff, as the tool for bdAD diagnosis (Van Hulle et al., 2021). We predicted that poorer performance in process metrics, including more forgetting, would associate with increased risk of bdAD classification.

Methods

Participants: Data were drawn from the University of Wisconsin – Madison Alzheimer's Disease Research Center (ADRC) and the Wisconsin Registry for Alzheimer's Prevention (WRAP). To be included in the analysis, participants had to have had at least two assessment visits in either ADRC or WRAP: one lumbar puncture (LP) visit for CSF extraction, and one cognitive evaluation. These visits had to be within a year of each other to ensure assessment validity. The initial reference pool comprised of 2,498 participants, then reduced to 295 participants (mean age = 64.6, ± 9.0) after applying the inclusion criteria above.

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. Table 1 reports demographic variables, CSF levels, and memory scores.

Memory assessment. Word-list recall performance was assessed with the Rey Auditory Verbal Learning Test (AVLT; Rey, 1958). In this test, participants are read a list of 15 unrelated nouns a total of five times, and are asked to free recall these words after each presentation, in any order. Then a new 15-word list is tested (interference), followed again by free recall of the originally presented list. Finally, after about 20–30 min, subjects are asked to free recall the original list once again, ending with a recognition test. To evaluate episodic memory, we scored *total recall* (sum of all the correctly recalled items across all five initial trials), and *delayed recall* (number of words recalled correctly after the 20–30 min delay), which represent the typical test scores extracted from the AVLT. We then scored the recency ratio (Rr), a measure of recency forgetting, following previously reported works (Bruno et al., 2022, 2018). In brief, the final four words of the learning

Table 1. Demographics, CSF and memory data by biomarker-determined diagnosis. LP = lumbar puncture. Imm = Immediate. Del = Delayed. Pr = primacy ratio. Rr = recency ratio. Tr = total ratio.

	Group	N	Mean	SD
Females	Control	147	–	–
	AD	34	–	–
Years of education	Control	225	16.333	2.469
	AD	70	15.814	2.778
APOE risk score	Control	225	1.141	0.733
	AD	70	1.884	0.915
Age at LP	Control	225	62.450	8.201
	AD	70	71.337	8.087
Time elapsed	Control	225	0.308	0.233
	AD	70	0.219	0.196
CSF p-tau/Abeta42 (ng/L)	Control	225	0.069	0.011
	AD	70	0.032	0.009
LMT Imm	Control	225	14.051	3.701
	AD	70	8.593	5.685
LMT Del	Control	225	12.904	3.971
	AD	70	6.829	6.124
LMT Pr	Control	225	0.831	0.262
	AD	70	0.416	0.436
LMT Rr	Control	225	1.030	0.208
	AD	70	1.365	0.767
LMT Tr	Control	225	1.125	0.356
	AD	70	1.751	1.422
AVLT total	Control	225	49.502	9.769
	AD	70	33.814	15.359
AVLT delayed	Control	225	9.724	3.605
	AD	70	4.243	4.874
AVLT Rr	Control	225	1.340	0.779
	AD	70	2.295	1.323
AVLT Tr	Control	225	0.870	0.940
	AD	70	2.306	1.931

list are considered to be within the recency region: a ratio is then calculated between recency performance at the first learning trial (trial 1), and recency performance at the delayed trial. Finally, a + 1 correction is applied to both terms to avoid 0 scores (e.g., trial 1 recency = 3; delayed trial recency = 2; (3 + 1)/(2 + 1) = 4/3 = 1.33). Additionally, we also calculated a total forgetting score (Tr; Bruno et al., 2022), by dividing overall trial 1 recall by overall delayed recall, and applying the same +1 adjustment (e.g., trial 1 = 13; delayed trial = 9; (13 + 1)/(9 + 1) = 14/10 = 1.4).

Story recall performance was assessed with the Logical Memory Task (LMT), a subtest of the Weschler Memory Scale Revised (WMS-R; Wechsler, 1987). LMT comprises two stories, 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. Also on the LMT, 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 permit 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. To note, while both stories were used in WRAP, only story A was employed in ADRC. Two conventional clinical metrics were extracted from LMT (averaging over A and B for WRAP data): Immediate LMT, derived from the total number of idea units recalled immediately after learning the story; and Delayed LMT, derived from the total number of idea units recalled after the 20–25 min delay.

Process scores were primacy ratio (Pr; Bruno et al., 2021), and Rr (Bruno et al., 2018) and Tr (Bruno et al., 2021). Primacy and recency were defined as the first and final eight idea units of the story (out of 25), in keeping with previous studies (Bruno et al., 2021). Rr and Tr were computed as with AVLT, while Pr was calculated following Bruno et al. (2021) by dividing delayed primacy by immediate primacy, with no adjustments. While Pr's formula is inconsistent with the way Rr and Tr are computed, and should arguably be aligned, we opted here for maintaining the original formula for the sake of comparison.

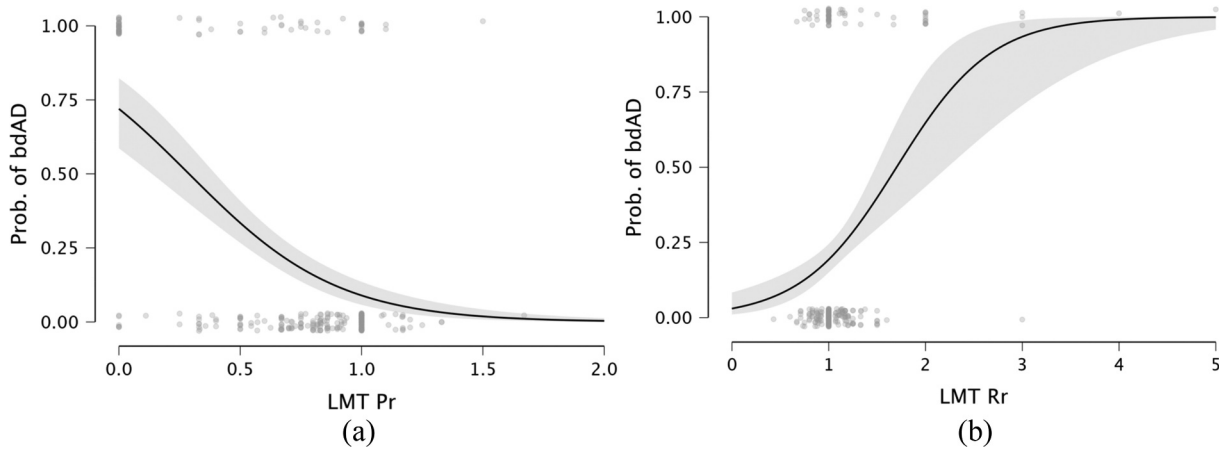

Figure 1. Conditional estimates plot (95% confidence intervals). LMT primacy ratio (Pr) to lumbar puncture for bdAD classification (0: without bdAD; 1: with bdAD). The Y-axis represents the probability of a bdAD classification based on the p-tau/Abeta42 cut off of 0.038. As the primacy ratio score increases, the likelihood of a bdAD classification decreases.

Participants' cognitive data were taken from whichever visit was closest to the visit where the LP was performed (average time between visits was 0.8 years, SD = 1.0) and from whichever visit was last recorded (average time between visits was 5.3 years later, SD = 5.9, for AVLT, and 3.5 years later, SD = 4.0, for LMT).

Biomarker determination. All CSF samples were assayed at the Clinical Neurochemistry Laboratory, University of Gothenburg, under strict quality control procedures. CSF markers (phosphorylated tau 181, henceforth p-tau; and amyloid β 1–42, henceforth Aβ42) were measured using the exploratory Roche NeuroToolKit assays, a panel of automated robust prototype immunoassays (Roche Diagnostics International Ltd, Rotkreuz, Switzerland), as previously described (Van Hulle et al., 2021).

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 (McKhann et al., 2011).

Analysis plan. We first carried out a Bayesian regression analysis with CSF p-tau/Aβ42 ratio as outcome; all memory scores (standard and process), as predictors (in the same analysis to allow for interactions); and the following covariates: age; time elapsed between memory assessment and the lumbar puncture; years of education; gender; visit number (to account for potential practice effects); and *APOE* risk score. The null model comprised all control variables. 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.

Second, we performed sensitivity analyses with the best predictors from the Bayesian regression by running Frequentist bivariate regression analyses and extracting diagnostic data. Outcome in this case was a CSF p-tau/Aβ42 ratio cutoff of 0.038, as previously determined for this cohort (Van Hulle et al., 2021): a low ratio, up to and including 0.038, identified individuals without bdAD, whereas a higher ratio identified individuals with bdAD. Of the 295 participants included in the analysis, 225 were controls and 70 displayed bdAD. Covariates in these analyses were the same as in the Bayesian regressions. Analyses were conducted using JASP (0.17.1; https://jasp-stats.org/).

Results

Table 1 reports demographics, CSF and memory data by bdAD.

The best fitting model, with a BF_{10} over 1 billion, included Immediate LMT, Delayed LMT, LMT Pr and LMT Rr. To note, the best model did not include any AVLT predictors, thus suggesting that LMT scores may be better overall when predicting bdAD classification. Inclusion probabilities favored Pr (>0.999; $BF_{inclusion}$>1 billion) and Rr (>0.999; $BF_{inclusion}$ >1,000,000), over LMT Immediate (0.998; $BF_{inclusion}$ = 207), and LMT Delayed (0.992; $BF_{inclusion}$ = 53; see Supplementary Materials for the full results). Therefore, the logistic sensitivity analyses included Pr and Rr as predictors. Pr reduced the AIC from 226 to 207, $p < 0.001$. Pr also was a significant predictor ($p < 0.001$; ORs = 0.084). Rr reduced AIC to 215, $p < 0.001$, and also was a significant predictor ($p = 0.002$; ORs = 4.790). When combining both Pr and Rr as predictors in the same analysis, multicollinearity did not appear to be an issue (VIF<1.5). AIC was a little lower than with Pr alone (205). LMT Pr remained significant ($p < 0.001$, ORs = 0.128), while Rr displayed a trend ($p = 0.051$, ORs = 2.940). These results suggest that LMT Pr may be slightly better

overall as a predictor of bdAD classification in these data compared to LMT Rr, but that LMT Rr should not be completely discounted. Diagnostic metrics (including covariates) for Pr alone show high specificity (0.942), lower sensitivity (0.686), moderate PPV (0.787), and high NPV (0.906). Removing covariates and using 0.26 as Pr cutoff increased specificity to 0.960, reduced sensitivity to 0.471; PPV was 0.786 and NPV was 0.854 (Figure 1a). Diagnostic metrics (including covariates) for Rr alone show high specificity (0.960), lower sensitivity (0.614), good PPV (0.827), and high NPV (0.889). Removing covariates and using 1.70 as a cut-point increased specificity to 0.996, reduced sensitivity to 0.229; PPV was 0.941 and NPV was 0.806 (Figure 1b).

Discussion

Tests that are noninvasive yet relatively accurate are important for clinical decision making (Trevethan, 2017). With this paper, we aimed to test how well conventionally used test scores compared to process scores derived from word-list and story recall tests predicted biomarker-determined AD. We did this by using the CSF p-tau/Aβ42 ratio, either continuously or as a cutoff in sensitivity analyses, as outcome. Our results support the use of story-recall process scores, specifically Pr, which indices loss of primacy information over time, and to a lesser extent Rr, which tracks loss of recency information, as valuable predictors of bdAD classification. Moreover, we noted that LMT-derived process scores provided better fits than AVLT-derived metrics, in these data.

The identification of Pr as a sensitive process score for the identification of bdAD is consistent with previous work showing that LMT-derived Pr was sensitive to a longitudinal classification as amyloid-positive, based on Pittsburgh compound B PET tracing (Bruno et al., 2021). The population tested in that paper partly overlapped with the participants included in the present manuscript, although overall sample size was nearly doubled here. We previously argued that loss of primacy may signal a failure to consolidate contextual information (Bruno et al., 2021), which in turn serves as a strong cue for recalling the learned material (e.g., Howard & Kahana, 2002; Howard et al., 2006). However, this hypothesis cannot be tested within the current dataset, and will require further examination. Nevertheless, beyond theoretical concerns, our findings suggest that LMT Pr is a valuable tool for identifying individuals whose memory may be declining but who are unlikely to have AD presently or in the subsequent few years. In fact, when we examine Pr values cross-sectionally, scores above 0.26 allowed to identify controls correctly 95% of the time (false positive rate ~ 5%). Similarly, scores above 0.32 allowed for 96% correct classification of individuals without longitudinal bdAD. In turn, scores below the cutoffs identified people with bdAD approximately three out of four times.

Previous work has shown Rr, both with word-lists and story recall tests, to be a valuable predictor of AD biomarkers (Bruno et al., 2022, 2021), and tauopathy in particular (Bruno et al., 2022). Indeed, Rr was also a strong predictor of p-tau/Ab42 levels in the present analyses as well, although it was superseded by Pr in binary logistic regressions. It is possible that while Pr is more sensitive to changes in brain amyloid deposition, as proposed in Bruno et al. (2021), Rr is more responsive to tau-related neuronal damage. However, this is currently speculation and further investigation, examining more closely neurocognitive activity with brain imaging, would be required to address this hypothesis.

One notable limitation of this study is that AVLT and LMT baselines were not always the same. We attempted to account for this discrepancy by controlling for both time differences in analyses where both tests were included. However, this is not an ideal methodological choice. Nevertheless, LMT scores still yielded better correlations with CSF p-tau/Aβ42 compared to AVLT scores when we examined the nearest cognitive assessments to lumbar puncture – in which case, LMT and AVLT were administered during the same session. However, exactly why LMT outperformed AVLT in these analyses is not clear and warrants further inspection, including whether these results can be generalized to other story-recall or word-list learning tasks. Moreover, several other process scores, such as intrusions or learning slopes, were not included in the present analyses. Further research may want to address this limitation, and compare Pr and Rr to other established process scores.

Finally, as noted, while both stories were used in WRAP, only story A was employed in ADRC. Therefore, we also carried out separate analyses in WRAP only (n = 212) and ADRC only (n = 83) participants. With WRAP only participants, the overall results were unchanged: the best fitting model included Immediate LMT, Delayed LMT, LMT Pr and LMT Rr, with a BF_{10} over 1 billion. Inclusion probabilities favored Pr (>0.999; $BF_{inclusion}$>20,000,000) and Rr (>0.999; $BF_{inclusion}$>10,000,000), over LMT Immediate (0.995; $BF_{inclusion}$ = 60), and LMT Delayed (0.931; $BF_{inclusion}$ = 11). Again, the logistic sensitivity analyses included Pr and Rr as predictors. Pr reduced the AIC from 174 to 154, $p < 0.001$. Pr also was a significant predictor ($p < 0.001$; ORs = 0.057). Rr reduced AIC to 161, $p < 0.001$, and also was a significant predictor ($p = 0.003$; ORs = 8.353). In contrast, when examining ADRC only participants, no model predicted CSF p-tau/Aβ42 better than the null model. However, it is unclear

whether this lack of prediction is due to using a single story in LMT, or due to the substantially reduced sample size. Further evidence is required to elucidate this question.

In summary, with this study, we showed that LMT-based process scores that account for forgetting of serial position information, and primacy ratio in particular, may be useful tools to detect individuals with suspected biomarker-defined AD. In particular, we showed that these metrics associated with pathology better than traditional scores, such as total and delayed recall. While these tests may not be fully diagnostic, they can aid clinicians by gathering valuable information on the possible state of the individual quickly and cost-effectively.

Disclosure statement

MC is a full-time employee and shareholder of Roche Diagnostics International Ltd., GK is a full-time employee of Roche Diagnostics GmbH, and NW is a full-time employee of Roche Diagnostics GmbH. COBAS, COBAS E and ELECSYS are trademarks of Roche. The Elecsys® β-Amyloid (1-42) CSF assay, the Elecsys® Phospo-Tau (181P) CSF assay and the Elecsys® Total-Tau CSF assay are not approved for clinical use in the US. The NeuroToolKit robust prototype assays are for investigational purposes and are not approved for clinical use. HZ has served at scientific advisory boards and/or as a consultant for Abbvie, Acumen, Alector, Alzinova, ALZPath, Annexon, Apellis, Artery Therapeutics, AZTherapies, CogRx, Denali, Eisai, Nervgen, Novo Nordisk, Optoceutics, Passage Bio, Pinteon Therapeutics, Prothena, Red Abbey Labs, reMYND, Roche, Samumed, Siemens Healthineers, Triplet Therapeutics, and Wave, has given lectures in symposia sponsored by Cellectricon, Fujirebio, Alzecure, Biogen, and Roche, and is a co-founder of Brain Biomarker Solutions in Gothenburg AB (BBS), which is a part of the GU Ventures Incubator Program (outside submitted work). KB is supported by the Swedish Research Council (#2017-00915), the Alzheimer Drug Discovery Foundation (ADDF), USA (#RDAPB-201809-2016615), the Swedish Alzheimer Foundation (#AF-930351, #AF-939721 and #AF-968270), Hjärnfonden, Sweden (#FO2017-0243 and #ALZ2022-0006), the Swedish state under the agreement between the Swedish government and the County Councils, the ALF-agreement (#ALFGBG-715986 and #ALFGBG-965240), the European Union Joint Program for Neurodegenerative Disorders (JPND2019-466-236), the National Institute of Health (NIH), USA, (grant #1R01AG068398-01), and the Alzheimer's Association 2021 Zenith Award (ZEN-21-848495). No other author reports any conflicts of interests or disclosures.

Data availability statement

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. 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

Funding

This work was supported by the National Institute on Aging [R01 144 AAI8612]. We wish to thank all WRAP and ADRC participants. This secondary analysis of WRAP and ADRC data was funded by a NIH-NIA (R01 144 AAI8612) grant to KDM, in which DB and RL are co-investigators. The full results from the Bayesian regression analysis are reported in the Supplementary Materials. The R code used in JASP for these analyses is also included there. The ethical regulations that govern WRAP and ADRC prevent unrestricted public archiving of anonymised study data.

ORCID

Davide Bruno http://orcid.org/0000-0003-1943-9905
Kimberly D. Mueller http://orcid.org/0000-0001-8691-9517

References

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., Phelps, C. H., Holtzman, D. M., Jagust, W. J., Petersen, R. C., Snyder, P. J., Carrillo, M. C., Thies, B., & 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. https://doi.org/10.1016/j.jalz.2011.03.008

Baek, M. J., Kim, H. J., & Kim, S. (2012). Comparison between the story recall test and the word-list learning test in Korean patients with mild cognitive impairment and early stage of Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology*, *34*(4), 396–404. https://doi.org/10.1080/13803395.2011.645020

Bruno, D., Jauregi, A., Pomara, N., Zetterberg, H., Blennow, K., Koscik, R., & Mueller, K. (2022). Cross-sectional associations of CSF tau levels with Rey's AVLT: A recency ratio study. *Neuropsychology*. https://doi.org/10.1037/neu0000821

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. https://doi.org/10.1017/S1041610218000467

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. https://doi.org/10.1111/jnp.12235

Bruno, D., Reiss, P. T., Petkova, E., Sidtis, J. J., & Pomara, N. (2013). Decreased recall of primacy words predicts cognitive decline. *Archives of Clinical Neuropsychology*, *28*(2), 95–103. https://doi.org/10.1093/arclin/acs116

De Simone, M. S., Perri, R., Fadda, L., Caltagirone, C., & Carlesimo, G. A. (2019). Predicting progression to Alzheimer's disease in subjects with amnestic mild cognitive impairment using performance on recall and recognition tests. *Journal of Neurology*, *266*(1), 102–111. https://doi.org/10.1007/s00415-018-9108-0

De Simone, M. S., Perri, R., Fadda, L., De Tollis, M., Turchetta, C. S., Caltagirone, C., & Carlesimo, G. A. (2017). Different

deficit patterns on word lists and short stories predict conversion to Alzheimer's disease in patients with amnestic mild cognitive impairment. *Journal of Neurology*, *264*(11), 2258–2267. https://doi.org/10.1007/s00415-017-8623-8

De Tollis, M., De Simone, M. S., Perri, R., Fadda, L., Caltagirone, C., & Carlesimo, G. A. (2021). Verbal and spatial memory spans in mild cognitive impairment. *Acta Neurologica Scandinavica*, *144*(4), 383–393. https://doi.org/10.1111/ane.13470

Dubois, B., Feldman, H. H., Jacova, C., DeKosky, S. T., Barberger-Gateau, P., Cummings, J., Scheltens, P., Galasko, D., Gauthier, S., Jicha, G., Meguro, K., O'Brien, J., Pasquier, F., Robert, P., Rossor, M., Salloway, S., Stern, Y., Visser, P. J., & Scheltens, P. (2007). Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS–ADRDA criteria. *The Lancet Neurology*, *6*(8), 734–746. https://doi.org/10.1016/S1474-4422(07)70178-3

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269–299. https://doi.org/10.1006/jmps.2001.1388

Howard, M. W., Kahana, M. J., & Wingfield, A. (2006). Aging and contextual binding: Modeling recency and lag recency effects with the temporal context model. *Psychonomic Bulletin & Review*, *13*(3), 439–445. https://doi.org/10.3758/BF03193867

Illán-Gala, I., Pegueroles, J., Montal, V., Vilaplana, E., Carmona-Iragui, M., & Alcolea, D., & Alzheimer's Disease Neuroimaging Initiative. (2018). Challenges associated with biomarker-based classification systems for Alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *10*, 346–357. https://doi.org/10.1016/j.dadm.2018.03.004

Lemos, R., Duro, D., Simões, M. R., & Santana, I. (2014). The free and cued selective reminding test distinguishes frontotemporal dementia from Alzheimer's disease. *Archives of Clinical Neuropsychology*, *29*(7), 670–679. https://doi.org/10.1093/arclin/acu031

Libon, D. J., Swenson, R., Ashendorf, L., Bauer, R. M., & Bowers, D. (2013). Edith Kaplan and the Boston process approach. *The Clinical Neuropsychologist*, *27*(8), 1223–1233. https://doi.org/10.1080/13854046.2013.833295

Manera, V., Rovini, E., & Wais, P. (2023). Early detection of neurodegenerative disorders using behavioral markers and new technologies: New methods and perspectives. *Frontiers in Aging Neuroscience*, 15. https://doi.org/10.3389/fnagi.2023.1149886

Mansbach, W. E., Mace, R. A., & Clark, K. M. (2014). Differentiating levels of cognitive functioning: A comparison of the brief interview for mental status (BIMS) and the brief cognitive assessment tool (BCAT) in a nursing home sample. *Aging & Mental Health*, *18*(7), 921–928. https://doi.org/10.1080/13607863.2014.899971

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Jr, Kawas, C. H., Klunk, W. E., Koroshetz, W. J., Manly, J. J., Mayeux, R., Mohs, R. C., Morris, J. C., Rossor, M. N., Scheltens, P., Carrillo, M. C., Thies, B., Weintraub, S., & 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. https://doi.org/10.1016/j.jalz.2011.03.005

Milberg, W. P., Hebben, N., Kaplan, E., Grant, I., & Adams, K. (2009). The Boston process approach to neuropsychological assessment. *Neuropsychological Assessment of Neuropsychiatric and Neuromedical Disorders*, *3*, 42–65.

Park, J. H., Park, H., Sohn, S. W., Kim, S., & Park, K. W. (2017). Memory performance on the story recall test and prediction of cognitive dysfunction progression in mild cognitive impairment and Alzheimer's dementia. *Geriatrics & Gerontology International*, *17*(10), 1603–1609. https://doi.org/10.1111/ggi.12940

Perri, R., Fadda, L., Caltagirone, C., & Carlesimo, G. A. (2013). Word list and story recall elicit different patterns of memory deficit in patients with Alzheimer's disease, frontotemporal dementia, subcortical ischemic vascular disease, and *Lewy* body dementia. *Journal of Alzheimer's Disease*, *37*(1), 99–107. https://doi.org/10.3233/JAD-130347

Porteri, C., Albanese, E., Scerri, C., Carrillo, M. C., Snyder, H. M., Martensson, B., & for the Roadmap, G. T. F. (2017). The biomarker-based diagnosis of Alzheimer's disease. 1—ethical and societal issues. *Neurobiology of Aging*, *52*, 132–140. https://doi.org/10.1016/j.neurobiolaging.2016.07.011

Rey, A. (1958). *L'examen clinique en psychologie*. Presses Universitaries De France.

Salvadó, G., Larsson, V., Cody, K. A., Cullen, N. C., Jonaitis, E. M., Stomrud, E., & Hansson, O. (2022). Optimal combinations of CSF biomarkers for predicting cognitive decline and clinical conversion in cognitively unimpaired participants and mild cognitive impairment patients: A multicohort study. *medRxiv*. https://doi.org/10.1002/alz.12907

Talamonti, D., Koscik, R., Johnson, S., & Bruno, D. (2019). Predicting early mild cognitive impairment with free recall: The primacy of primacy. *Archives of Clinical Neuropsychology*, *35*(2), 133–142. https://doi.org/10.1093/arclin/acz013

Teichmann, M., Epelbaum, S., Samri, D., Nogueira, M. L., Michon, A., Hampel, H., Dubois, B. (2017). Free and Cued Selective Reminding Test–accuracy for the differential diagnosis of Alzheimer's and neurodegenerative diseases: A large-scale biomarker-characterized monocenter cohort study (ClinAD). *Alzheimer's & Dementia*, *13*(8), 913–923. https://doi.org/10.1016/j.jalz.2016.12.014

Trevethan, R. (2017). Sensitivity, specificity, and predictive values: Foundations, pliabilities, and pitfalls in research and practice. *Frontiers in Public Health*, *5*, 307. https://doi.org/10.3389/fpubh.2017.00307

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. https://doi.org/10.3233/JAD-180690

Van Hulle, C., Jonaitis, E. M., Betthauser, T. J., Batrla, R., Wild, N., Kollmorgen, G., … Blennow, K. (2021). An examination of a novel multipanel of CSF biomarkers in the Alzheimer's disease clinical and pathological continuum. *Alzheimer's & Dementia*, *17*(3), 431–445. https://doi.org/10.1002/alz.12204

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

World Health Organization. *Dementia*. Geneva: World Health Organization, 15 March 2023. https://www.who.int/news-room/fact-sheets/detail/dementia