



A comparison of story-recall metrics to predict hippocampal volume in older adults with and without cognitive impairment

Ainara Jauregi Zinkunegi, Davide Bruno, Tobey J. Betthausen, Rebecca Langhough, Sanjay Asthana, Nathaniel A. Chin, Bruce P. Hermann, Sterling C. Johnson & Kimberly D. Mueller

To cite this article: Ainara Jauregi Zinkunegi, Davide Bruno, Tobey J. Betthausen, Rebecca Langhough, Sanjay Asthana, Nathaniel A. Chin, Bruce P. Hermann, Sterling C. Johnson & Kimberly D. Mueller (2023): A comparison of story-recall metrics to predict hippocampal volume in older adults with and without cognitive impairment, *The Clinical Neuropsychologist*, DOI: [10.1080/13854046.2023.2223389](https://doi.org/10.1080/13854046.2023.2223389)

To link to this article: <https://doi.org/10.1080/13854046.2023.2223389>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 22 Jun 2023.



[Submit your article to this journal](#)



[View related articles](#)



[View Crossmark data](#)

A comparison of story-recall metrics to predict hippocampal volume in older adults with and without cognitive impairment

Ainara Jauregi Zinkunegi^a, Davide Bruno^a, Tobey J. Betthausen^{b,c}, Rebecca Langhough^{b,c,d}, Sanjay Asthana^{b,d,e}, Nathaniel A. Chin^{b,c}, Bruce P. Hermann^{d,f}, Sterling C. Johnson^{b,c,d,e} and Kimberly D. Mueller^{c,d,g}

^aSchool of Psychology, Liverpool John Moores University, Liverpool, UK; ^bDepartment of Medicine, School of Medicine and Public Health, University of Wisconsin, Madison, WI, USA; ^cWisconsin Alzheimer's Disease Research Center, School of Medicine and Public Health, University of Wisconsin, Madison, WI, USA; ^dWisconsin Alzheimer's Institute, School of Medicine and Public Health, University of Wisconsin, Madison, WI, USA; ^eGeriatric Research Education and Clinical Center, William S. Middleton Veterans Hospital, Madison, WI, USA; ^fDepartment of Neurology, University of Wisconsin, Madison, WI, USA; ^gDepartment of Communication Sciences and Disorders, University of Wisconsin-Madison, Madison, WI, USA

ABSTRACT

Objective: Process-based scores of episodic memory tests, such as the recency ratio (Rr), have been found to compare favourably to, or to be better than, most conventional or “traditional” scores employed to estimate memory ability in older individuals (Bock et al., 2021; Bruno et al., 2019). We explored the relationship between process-based scores and hippocampal volume in older adults, while comparing process-based to traditional story recall-derived scores, to examine potential differences in their predictive abilities. **Methods:** We analysed data from 355 participants extracted from the WRAP and WADRC databases, who were classified as cognitively unimpaired, or exhibited mild cognitive impairment (MCI) or dementia. Story Recall was measured with the Logical Memory Test (LMT) from the Wechsler Memory Scale Revised, collected within twelve months of the magnetic resonance imaging scan. Linear regression analyses were conducted with left or right hippocampal volume (HV) as outcomes separately, and with Rr, Total ratio, Immediate LMT, or Delayed LMT scores as predictors, along with covariates. **Results:** Higher Rr and Tr scores significantly predicted lower left and right HV, while Tr showed the best model fit of all, as indicated by AIC. Traditional scores, Immediate LMT and Delayed LMT, were significantly associated with left and right HV, but were outperformed by both process-based scores for left HV, and by Tr for right HV. **Conclusions:** Current findings show the direct relationship between hippocampal volume and all the LMT scores examined here, and that process-based scores outperform traditional scores as markers of hippocampal volume.

ARTICLE HISTORY

Received 6 February 2023

Accepted 5 June 2023

Published online 23 June 2023

KEYWORDS

Recency ratio; hippocampus; Alzheimer's disease; story recall; logical memory

CONTACT Ainara Jauregi Zinkunegi ✉ a.jauregizinkunegi@ljamu.ac.uk 📍 Tom Reilly Building, Liverpool John Moores University, Liverpool L3 3AF, UK.

📎 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/13854046.2023.2223389>.

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Introduction

In research studies and clinical settings, performance on common neuropsychological tests is often calculated by adding the number of correct responses into a total score, which is then interpreted using cut-off points from standard normative data. However, this quantitative or “traditional” scoring and interpretation method does not provide much information on the cognitive functions engaged during the test or the strategies used by patients to obtain the scores (Grant & Adams, 2009), or which of these factors may be contributing to an impaired score (Blanco-Campal et al., 2021). To overcome these limitations, a qualitative or “process-based” approach, also known as the Boston Process Approach (BPA), was first developed by Kaplan (1988). This approach focuses on the strategies used and errors committed during the test (Libon et al., 2022), and can be implemented with procedural modifications of the tasks, such as time limits or adding new components, or by modifying the calculation of their scores (for a review, see Milberg et al., 2009). An example of process approach applied to episodic memory tests (e.g. list-learning and story recall tasks) is analysis of serial position, where primacy, recency, or clustering effects can be analysed in addition to “traditional” scores (Bruno et al., 2013; Diaz-Orueta et al., 2018; Grant & Adams, 2009; Talamonti et al., 2020).

Serial position effects are found when early items (primacy) and late items (recency) on a list are remembered better than items in the middle (Murdock, 1962). Individuals with Alzheimer’s disease (AD) have been reported to show intact or exaggerated immediate memory recall of items learned at the end of a list (recency items, e.g. Foldi et al., 2003), while performing poorly when recalling the same items after a delay (Carlesimo et al., 1995). To leverage this discrepancy in recency performance, Bruno et al. (2016, 2018) proposed a process-based score, the *recency ratio* (Rr), which is calculated by dividing the recency scores in the immediate recall trial by the corresponding scores in the delayed recall trial. Higher Rr scores, showing disproportionate loss of recency recall from immediate to delayed testing, indicate more recency forgetting and, consequently, more overall risk of cognitive impairment (Bruno et al., 2022). Studies using Rr from list-learning tasks have shown that higher (i.e. worse) scores predict cognitive decline (Bruno et al., 2016), early mild cognitive impairment (MCI; Bruno et al., 2018; Egeland, 2021), and amyloid- β pathology in individuals with MCI (Bruno et al., 2019). Rr scores have also been found to accurately discriminate between individuals diagnosed with AD and other types of dementia, such as frontotemporal, Lewy body and vascular dementias (Turchetta et al., 2018), and to correctly classify MCI patients who are more likely to convert to AD (Turchetta et al., 2020). Bruno et al. (2021a) found Rr to be sensitive to the levels of cerebrospinal fluid (CSF) Neurogranin, a post-synaptic protein reported to reflect synaptic dysfunction in AD and MCI, in individuals with major depressive disorder. Furthermore, higher Rr scores have been found to be associated with increased CSF levels of p-tau and t-tau in older adults with and without cognitive impairment, while total and delayed recall scores were not linked with any of the AD biomarkers (Bruno et al., 2022). Overall, Rr tends to compare favourably to, or perform better than, most conventional or “traditional” scores employed to estimate memory ability in older individuals (Bock et al., 2021; Bruno et al., 2019).

In principle, serial position effects can be examined in any common neuropsychological test of recall, however, until recently, these have been studied primarily in list-learning tasks. In contrast to list-learning tasks, story recall tasks require participants to learn coherent stories comprised of items of varying semantic and lexical categories. These stories must also be recalled immediately and after a delay. Although story recall tasks, such as the Logical Memory test (LMT; Wechsler, 1987), are very commonly used cognitive tools in research and clinical practice, little is known about serial position effects within these tasks compared to list-learning tasks. Only recently, a study by Bruno et al. (2023) found Rr to be applicable to story recall and to be sensitive to CSF levels of amyloid beta 1-40 (A β 40), p-tau, t-tau, Neurogranin and alpha synuclein—overall outperforming traditional LMT scores, but more evidence that Rr works in story recall tasks is needed.

It has been suggested that higher Rr scores could be due to a combination of reduction in long-term retention, possibly as a result of diminished consolidation ability, paired with a compensatory increased reliance upon short-term memory processing (Bruno et al., 2018). The medial-temporal lobe (MTL) in general, and hippocampus in particular, are essential in the formation and consolidation of long-term memory (Wixted, 2004; Wixted & Cai, 2013). Considering that early neuropathological changes in AD occur within the MTL, especially in the entorhinal cortex and hippocampus (Weiner et al., 2015), it could be argued that higher Rr scores in AD might be associated with changes in these areas (Turchetta et al., 2020). Additionally, verbal episodic memory performance has been shown to be associated with left, rather than right, hippocampal volume (Ezzati et al., 2016; Shi et al., 2009) and greater left hippocampal reductions, compared to right hippocampal reductions, have been reported in individuals with AD (Dhikav et al., 2016; Ezzati et al., 2016; Li et al., 2016; Müller et al., 2005; Shi et al., 2009; Thompson et al., 2003, 2007; Wicking et al., 2014). Thus, Rr might be sensitive to hippocampal health, and more specifically, to that of the left hippocampus.

To be certain that any possible effects are due to recency forgetting and not total forgetting, we propose a relatively novel process-based score, the Total ratio (Tr) which indices forgetting independently of recency (Bruno et al., 2021b, 2023). Tr is obtained by dividing the immediate recall trial scores by the delayed recall trial scores, where higher Tr scores reflect more total forgetting. Bruno et al. (2023) found Rr to be more sensitive to CSF biomarkers of neurodegeneration than Tr, but we know little about how Tr fares, compared to Rr, with different AD biomarker outcomes.

The aim of the current study was to explore the relationship between process-based scores, Rr and Tr, and hippocampal volume in older adults with and without cognitive impairment. We examined this relationship using memory scores collected within twelve months of the MRI scan, while also comparing process-based to traditional story recall-derived memory scores, by analysing data from both WRAP and WADRC samples. We predicted that Rr and Tr from story recall would be negatively associated with hippocampal volume, and that Rr and Tr would be a better predictor of hippocampal volume than traditional LMT and composite scores.

Methods

Participants

Data were extracted from the Wisconsin Alzheimer's Disease Research Center (WADRC) and the Wisconsin Registry for Alzheimer's Prevention (WRAP) studies. Participants were selected based on having completed one T1-weighted structural magnetic resonance imaging (MRI) scan, and one cognitive screening visit within twelve months of the MRI scan, including complete Logical Memory Test (LMT) story recall data, in either WRAP or WADRC. Exclusion criteria for both studies included major neurologic disorder (e.g. head trauma with loss of consciousness, seizures, or neoplasms), current (within the previous 12 months) major psychiatric disorders, or any other significant medical illness. From the total pool of 2,498 participants, 355 participants fulfilled the above criteria: 351 were native English speakers; four participants (1.13%) reported their race as American Indian or Native American, one (0.28%) as Asian, nine (2.54%) as Black, African American, or mixed, 340 (95.78%) as non-Hispanic White or White, and one (0.28%) as Spanish or Hispanic.

Participants were classified as either cognitively unimpaired, with MCI due to presumed AD (MCI-AD), or with dementia due to presumed AD (Dementia-AD), *via* multi-disciplinary consensus conference review that was blind to AD biomarkers statuses. In WRAP, a two-tiered consensus conference approach was used (for details, see Johnson et al., 2018; Langhough Kosciak et al., 2021). For both WRAP and ADRC, cognitive statuses were determined by teams that included physicians, clinical neuropsychologists, and clinical nurse practitioners, and based on core clinical criteria developed by the National Institute on Aging and the Alzheimer's Association (Albert et al., 2011; McKhann et al., 2011). Among the 355 participants included in the study, 282 individuals were cognitively unimpaired, 39 had a diagnosis of MCI-AD, and 34 had a diagnosis of Dementia-AD. 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.

Structural MRI

MRI images were acquired using two identical GE 3.0 Tesla MR750 scanners (Waukesha, WI, USA) with an 8-channel head coil (Excite HD Brain Coil; GE Healthcare) in one scanning session. T1-weighted brain volumes were acquired in the axial plane with a 3-D inversion-recovery prepared fast spoiled gradient-echo sequence using the following parameters: inversion time (TI) = 450 ms; repetition time (TR) = 8.2 ms; echo time (TE) = 3.2 ms; flip angle = 12°; acquisition matrix = 256 × 256 × 156 mm; field of view (FOV) = 256 mm; slice thickness = 1.0 mm. Cushions inside the head coil helped reduce head movement during scanning. All MRI scans were read by an experienced clinical neuroradiologist, who excluded participants from further analyses due to structural abnormalities, if required (see other exclusion criteria above).

During pre-processing, T1-weighted volumes were segmented into tissue classes (grey matter, white matter, and cerebrospinal fluid) using Statistical Parametric Mapping (SPM), Version 12 (<https://www.fil.ion.ucl.ac.uk/spm>). Hippocampal volume was

estimated using FSL-FIRST (Patenaude et al., 2011) and total intracranial volume was determined using the reverse brain mask method in SPM, Version 12. See Table 1 for elapsed times between MRI scan and neuropsychological assessment in each diagnostic group.

Cognitive assessment

The Logical Memory Test

The LMT was used to assess learning and memory performance. The LMT is a subtest of the Weschler Memory Scale Revised (WMS-R; Wechsler, 1987) which comprises two stories, A and B, with 25 items each ("idea units"), of different semantic and lexical categories. Each story is read aloud to the participant and then the participant is asked to recall both stories immediately and again after a 25–30-minute delay. 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 such as numerical expressions or proper names.

In the ADRC sample, story recall was measured with story A of the LMT, as story B was not administered, while in the WRAP sample, story recall was measured with story A and B of the LMT. Regardless of the number of stories tested, because scores from story A and B are averaged, LMT scores from WRAP are computed by averaging story A and B, whereas LMT scores from WADRC are based on one story, resulting in the same number of scores across the WRAP and WADRC samples. Immediate LMT and Delayed LMT recall scores were calculated by adding all the correctly recalled items in the immediate recall trial and delayed recall trial, respectively, and were analysed as raw scores, as in previous studies examining traditional and process-based LMT measures (Bruno et al., 2021b, 2023). Possible scores for Immediate and Delayed Recall trials range from 0 to 25 for each, where higher scores reflect more items being recalled. Immediate LMT and Delayed LMT scores were calculated individually from the closest visit to MRI scan data, if collected within twelve months of the scan.

Process-based measures of the LMT: recency ratio and Total ratio

Recency was defined as the final eight items of the story (Bruno et al., 2021b) and immediate and delayed recency scores were calculated as the number of correctly recalled recency items in immediate and delayed recall trials, respectively. Rr was obtained by dividing the recency scores in the immediate recall trial by the corresponding scores 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 (Bruno et al., 2018), possible Rr scores range from 1 to 9, where higher scores reflect more recency forgetting.

To provide a non-recency-based Rr analogue that would account for memory loss, we also computed a ratio score with Immediate LMT and Delayed LMT ((Immediate LMT + 1)/(Delayed LMT + 1)), the Total ratio (Tr; see also Bruno et al., 2021b). Possible Tr scores range from 1 to 26, where higher scores reflect more total forgetting. As with traditional story recall-derived memory scores, Rr and Tr scores were calculated

individually from the closest visit to MRI scan data, if collected within twelve months of the scan.

Global cognitive composite score

A composite score analogous to the Preclinical Alzheimer's Cognitive Composite 4 (PACC4; for details, see Donohue et al., 2014), based on available tests, was included as a measure of global cognitive functioning, for comparison. In the WRAP study, this composite score was calculated using the average of standardised tests scores of the following tests: total scores for the Logical Memory II subtest (i.e. delayed recall of stories A and B) from the Wechsler Memory Scale—Revised (WMS-R; Wechsler, 1987), total scores from the Digit Substitution test of the Wechsler Abbreviated Intelligence Scale-Revised (WAIS-R; Wechsler, 1981), total recall learning trials 1–5 from the Rey Auditory Verbal Learning Test (RAVLT; Schmidt, 1996), and the total score from the Mini-Mental Status Examination (MMSE; Folstein et al., 1975); for more details on how the score was calculated, see Jonaitis et al. (2019). In the WRAP sample, PACC4-analogue scores were collected at the same cognitive screening visit as LMT scores and were also calculated individually; we did not have PACC4-analogue scores for the WADRC sample.

Genotyping

The *APOE* e4 allele is considered the most important genetic risk factor of AD (Coon et al., 2007; Hobel et al., 2019) and thus, genetic risk was accounted for by calculating an *APOE* risk score based on the odds ratios of the e2/e3/e4 genotype, as previously reported (Darst et al., 2017). DNA was extracted from whole blood and samples were aliquoted on 96-well plates for determination of *APOE* genotypes. The *APOE* risk score was included as a covariate in all the regression analyses.

Statistical analysis

An analysis of variance (ANOVA) was carried out to compare age, time elapsed between MRI scan and memory assessment (calculated as an absolute value of months), years of education, *APOE* risk score, total intracranial volume (TIV), left hippocampal volume, and right hippocampal volume among participants, classified by cognitive status (cognitively unimpaired, MCI or dementia). When any of these comparisons was significant, a post-hoc paired comparison was conducted, by using Tukey's honest significant difference (HSD) test, to account for multiple testing. Differences in sex and sample (WRAP or WADRC) were assessed with a Pearson Chi-square test ($p < 0.05$). To test for differences in LMT scores, an analysis of covariance (ANCOVA) was conducted by adding age, gender, sample, elapsed time between MRI scan and neuropsychological assessment, years of education and *APOE* risk score as covariates; post-hoc comparisons were also adjusted using the Tukey's HSD test. See Table 1 for sample details, reported for the whole sample and by cognitive status.

To understand how correlated the LMT scores were, we ran bivariate Spearman's rank-order correlations between R_r , T_r , Immediate LMT, and Delayed LMT, as these scores were not normally distributed. Bivariate Spearman's rank-order correlations

were also conducted between left and right hippocampal volume, to explore how associated their size was. Partial correlations, controlling for age, gender, sample, elapsed time between MRI scan and neuropsychological assessment, years of education and APOE risk score, were used to explore the relationship between the LMT scores and left or right hippocampal volume. Steiger's Z tests (Steiger, 1980) were conducted on partial correlation coefficients between any two significant predictors, to determine if the strength of one association between one memory score and left or right HV outcome was stronger than the association between another memory score and the same HV outcome, by using a calculator (<http://quantpsy.org>; Lee & Preacher, 2013).

Linear regression analyses were conducted with Rr, Tr, Immediate LMT, and Delayed LMT as predictors (in separate models); sex, age at MRI scan, elapsed time between MRI scan and memory assessment (calculated as an absolute value), years of education, sample (WADRC or WRAP), and APOE risk score were used as control variables (covariates). Right or left hippocampal volume (HV) divided by total intracranial volume (TIV) represented the outcomes in separate analyses. We adjusted for multiple testing using a false discovery rate-based approach (FDR; Benjamini & Hochberg, 1995) for the four predictors, corrected across left and right HV. To determine which LMT score is the best predictor of left and right HV, we compared AIC fit statistics (Aiken et al., 1991) across otherwise parallel models, lower AIC values indicate a better fit, and a model with a delta-AIC (i.e. the difference between the two AIC values being compared) greater than 2 is considered significantly better than the model it is being compared to (Burnham & Anderson, 2004).

As an additional analysis, to compare the predictive abilities of process-based scores with a standardised measure of global cognition, separate regression analyses were carried out with left or right HV (divided by TIV) as outcome, Rr, Tr, and PACC4-analogue scores as predictors. The same covariates as above were included, except for sample (WADRC or WRAP), since the analyses were conducted only in participants from WRAP ($N=232$; see Global cognitive composite score above). We adjusted for multiple testing using FDR, and AIC fit statistics were compared to determine which model showed the best model fit, as above. Analyses were performed with SPSS, Version 27 (IBM).

Results

In Table 1, means and standard deviations are described for all the variables included in the current study, for the whole sample and by cognitive status closest to MRI scan. All the variables tested showed significant differences between the cognitive status groups. Specifically, post-hoc pairwise analysis revealed that participants with a worse cognitive status were older, with less years of education, and had less left and right HV than participants who were cognitively unimpaired. Mean Rr, Tr, Immediate LMT and Delayed LMT scores were also significantly different between groups, post-hoc analysis indicating the worse the cognitive status was, the higher the Rr and Tr values, and the lower the Immediate LMT and Delayed LMT scores were. In [Supplementary materials](#), we report pairwise plots of left vs. right HV (Table S1), and of left or right HV vs. age by cognitive status (Tables S2.1 and S2.2, respectively).

Bivariate Spearman's rank-order correlations showed that Rr, Tr, Immediate LMT, and Delayed LMT scores significantly correlated between them, whilst left and right

Table 1. Demographics, hippocampal volumes, and memory tests scores for the study participants. Variables are reported either as N(%) for categorical variables or mean(sd) for continuous variables, and as adjusted mean(sd) from ANCOVA for LMT scores in each cognitive status.

Characteristic	Total (N=355)	CU (N=282)	MCI-AD (N=39)	Dementia-AD (N=34)	p Value	Post-hoc (p < .05)
Sex (female)	225 (71.8%)	200 (70.9%)	12 (30.8%)	13 (38.2%)	<.001	
Sample (WRAP)	230 (64.8%)	224 (79.4%)	6 (15.4%)	0 (0%)	<.001	
Age at MRI scan (years)	65.0 (9.4)	62.6 (7.8)	74.8 (7.0)	73.7 (11.7)	<.001	CU < MCI CU < Dem Dem < CU Dem < MCI
Education (years)	16.1 (2.8)	16.3 (2.7)	16.2 (3.1)	14.5 (2.9)	.002	Dem < CU MCI < CU
Elapsed time (months)	4.9 (3.7)	5.6 (3.7)	3.0 (2.9)	1.5 (1.1)	<.001	Dem < CU MCI < CU
APOE risk score	1.3 (0.8)	1.3 (0.8)	1.4 (0.8)	1.8 (1.0)	.001	CU < Dem
TIV (mL)	1471.9 (141.5)	1460.4 (139.3)	1542.9 (148.9)	1485.7 (129.3)	.002	CU < MCI
Left HV (mL)	3638.8 (528.2)	3770.7 (452.5)	3168.1 (479.8)	3084.1 (506.6)	<.001	MCI < CU Dem < CU
Right HV (mL)	3844.5 (542.0)	3955.4 (477.5)	3497.9 (522.3)	3322.1 (606.0)	<.001	MCI < CU Dem < CU
Recency ratio	1.1 (0.5)	1.0 (0.0)	1.4 (0.1)	2.0 (0.1)	<.001	CU < MCI CU < Dem MCI < Dem
Total ratio	1.4 (1.0)	1.1 (0.1)	2.4 (0.2)	2.7 (0.2)	<.001	CU < MCI CU < Dem
Immediate LMT	12.7 (4.8)	14.1 (0.2)	9.2 (0.6)	5.8 (0.7)	<.001	MCI < CU Dem < CU Dem < MCI
Delayed LMT	11.3 (5.3)	13.0 (0.2)	6.1 (0.6)	2.9 (0.7)	<.001	MCI < CU Dem < CU Dem < MCI

Note: CU=cognitively unimpaired; MCI-AD=MCI due to presumed AD; Dementia-AD or Dem=dementia due to presumed AD; LMT=logical memory test; HV=Hippocampal volume; TIV=total intracranial volume; Elapsed time=time elapsed between MRI scan and neuropsychological assessment. pvalue for the omnibus test or chi-square test.

Table 2. Bivariate and partial correlations between memory scores, between left and right hippocampal volume, and between memory score and left or right hippocampal volume. Rr=recency ratio; Tr=Total ratio; LMT=logical memory test; HV=hippocampal volume; TIV=total intracranial volume.

	Rr ^a	Tr ^a	Immediate LMT ^a	Delayed LMT ^a	Left HV by TIV ^b	Right HV by TIV ^{b c}
Rr	1	0.503**	−0.162*	−0.310**	−0.270**	−0.186**
Tr		1	−0.290**	−0.572**	−0.297**	−0.215**
Immediate LMT			1	0.929**	0.168*	0.156*
Delayed LMT				1	0.251**	0.208**
Left HV					1	0.733**
Right HV						1

Note: N=355. ^aBivariate Spearman's rank-order correlations between Rr, Tr, Immediate LMT and Delayed LMT. ^bPartial correlations, controlling for age, gender, sample, elapsed time between MRI scan and neuropsychological assessment, years of education and APOE risk score, between memory scores and left or right HV. ^cBivariate Spearman's rank-order correlations between left and right HV corrected for TIV. *p < .05; **p < .01; ***p < .001. Bold: indicates the stronger partial coefficients in the column, when comparing the strength of the associations between each memory score and left or right HV, as per Steiger's Z-test.

HV were also significantly and positively associated. Partial correlations indicated all memory scores significantly correlated with left and right HV, see Table 2 for bivariate and partial correlation coefficients. Steiger's Z-tests showed that for left HV, the association with Rr was significantly stronger than with Immediate LMT (Z=−5.42; p < .001) or Delayed LMT (Z=−5.99; p < .001), but not than with Tr (Z=0.62; p = .267),

whereas the association with Tr was significantly stronger than with Immediate LMT ($Z = -5.75$; $p < .001$) or Delayed LMT ($Z = -6.19$; $p < .001$); the association with Delayed LMT was stronger than with Immediate LMT ($Z = 3.47$; $p < .001$). For right HV, the association with Rr was significantly stronger than with Immediate LMT ($Z = -4.18$; $p < .001$), but not than with Delayed LMT ($Z = -4.46$; $p < .001$) or Tr ($Z = 0.65$; $p = .258$), whereas the association with Tr was significantly stronger than with Immediate LMT ($Z = -4.53$; $p < .001$) or Delayed LMT ($Z = -4.71$; $p < .001$); the association with Delayed LMT was stronger than with Immediate LMT ($Z = 2.16$; $p = 0.015$).

Linear regression analyses with right HV divided by TIV as outcome, showed that the separate model fits with either Rr, Total ratio, Immediate LMT, or Delayed LMT were significant; as were their coefficients, see Table 3 for details. The AIC fit statistics across the four models with right HV as outcome were compared and showed that the Total ratio and Delayed LMT models had the lowest AIC, as shown in Table 3. Delta-AIC between the Total ratio model and either the Rr or Immediate LMT models were greater than two, indicating the Total ratio model was significantly better. The model with Delayed LMT also showed a delta-AIC difference greater than two with either the Rr or Immediate LMT models, revealing the Delayed LMT model was significantly better; yet the difference was less than two AIC values between the Total ratio and Delayed LMT models.

With left HV divided by TIV as outcome, separate model fits with either Rr, Total ratio, Immediate LMT, or Delayed LMT were significant; as were their coefficients,

Table 3. Linear regression models predicting left or right hippocampal volume, corrected for Total intracranial volume.

Predictor	Outcome	Recency ratio Model1		Total ratio Model2		Immediate LMT Model3		Delayed LMT Model4	
		β	p	β	p	β	p	β	p
Age at MRI scan	Left HV	-0.34	<.001	-0.33	<.001	-0.31	<.001	-0.29	<.001
	Right HV	-0.31	<.001	-0.30	<.001	-0.28	<.001	-0.26	<.001
Gender	Left HV	0.30	<.001	0.31	<.001	0.30	<.001	0.28	<.001
	Right HV	0.26	<.001	0.27	<.001	0.25	<.001	0.24	<.001
Education	Left HV	0.07	.114	0.07	.105	0.05	.264	0.04	.343
	Right HV	0.02	.613	0.02	.598	0.01	.923	0.00	.993
APOE risk score	Left HV	-0.08	.063	-0.08	.078	-0.06	.162	-0.05	.241
	Right HV	-0.04	.363	-0.04	.407	-0.02	.620	-0.02	.730
Elapsed time	Left HV	-0.05	.318	-0.07	.188	-0.05	.305	-0.06	.256
	Right HV	-0.06	.258	-0.07	.174	-0.06	.240	-0.07	.210
Sample	Left HV	0.07	.226	0.08	.128	0.05	.379	0.03	.580
	Right HV	0.13	.022	0.14	.012	0.11	.085	0.10	.111
LMT score	Left HV	-0.23	<.001*	-0.25	<.001*	0.18	.002*	0.26	<.001*
	Right HV	-0.17	<.001*	-0.19	<.001*	0.17	.004*	0.23	<.001*
AIC	Left HV	-837.50		-843.34		-820.72		-833.68	
	Right HV	-818.49		-822.83		-814.68		-821.69	
Adjusted R²	Left HV	.370		.380		.339		.363	
	Right HV	.293		.302		.286		.300	

Note: $N = 355$. β = Standardised regression coefficient; LMT = logical memory test; HV = Hippocampal volume; Elapsed time = time elapsed between MRI scan and neuropsychological assessment; p = p value; * adjusted p value for multiple comparisons. In all models, variables showed a Variation Inflation Factor (VIF) < 2; AIC = Akaike Information Criterion; Adjusted R^2 = adjusted proportion of explained variance. Bold: model with significantly best model fit, as per delta-AIC. Models with left HV by TIV as outcome: ¹Model with Rr: $F(7,347) = 30.644$ $p < .001$; ²Model with Tr: $F(7,347) = 31.973$ $p < .001$; ³Model with Immediate LMT: $F(7,347) = 26.940$ $p < .001$; ⁴Model with Delayed LMT: $F(7,347) = 29.784$ $p < .001$. With right HV by TIV as outcome: ¹Model with Rr: $F(7,347) = 22.006$ $p < .001$; ²Model with Tr: $F(7,347) = 22.888$ $p < .001$; ³Model with Immediate LMT: $F(7,347) = 21.244$ $p < .001$; ⁴Model with Delayed LMT: $F(7,347) = 22.655$ $p < .001$.

see Table 3 for details. The AIC fit statistics across the four models with left HV by TIV as outcome were compared and showed that the Total ratio model had the lowest AIC, followed by the Rr model, as shown in Table 3. Delta-AIC between the Total ratio model and either the Rr, Immediate LMT, or Delayed LMT models were greater than two, indicating the Total ratio model was significantly better. Delta-AICs between the Rr model and the models with either of the traditional scores were also greater than two, showing that the Rr model fit was significantly better than that of the Immediate LMT or Delayed LMT models for left HV.

As a secondary analysis, we carried out linear regression analyses to compare the model fits of process-based measures of the LMT, Rr and Tr, to that of a standardised composite score of global cognition. With PACC4-analogue as predictor, linear regression analyses showed that the model fit for left HV divided by TIV as outcome was significant ($F(6,225) = 7.077, p < .001$), as was for right HV divided by TIV ($F(6,225) = 5.342, p < .001$), yet PACC4-analogue coefficients were not significant for left HV ($t=1.082$, adjusted- $p = .337, \beta = .073$) or right HV ($t=0.914$, adjusted- $p = .362, \beta = .063$). The AIC values of the PACC4-analogue models were 85.457 for left HV and 87.719 for right HV.

With process-based scores of story recall, Rr or Tr, as predictors in WRAP participants only, the model fits were significant for both left HV by TIV (Rr, $F(6,225) = 8.382, p < .001$; Tr, $F(6,225) = 8.994, p < .001$) and right HV by TIV (Rr, $F(6,225) = 5.541, p < .001$; Tr, $F(6,225) = 6.550, p < .001$). The Rr coefficient was significant for left HV ($t=-2.791$, adjusted- $p = .016, \beta = -0.169$), but not for right HV ($t=-1.373$, adjusted- $p = .257, \beta = -0.086$), while the Tr coefficient was significant for both left HV ($t=-3.301$, adjusted- $p = .006, \beta = -0.199$) and right HV ($t=-2.684$, adjusted- $p = .016, \beta = -0.166$). The AIC values of the Rr model for left HV was 78.762 and 86.664 for right HV, the Tr model had an AIC value of 75.686 for left HV, 81.268 for right HV. Delta-AIC between the Total ratio model and either the Recency ratio or PACC4-analogue models were greater than two, for left and right HV, indicating the Total ratio model was significantly better, while delta-AIC was only significantly between Rr and PACC4-analogue for left HV.

Discussion

In this study, we explored the relationship between process-based measures of story recall, i.e. Rr and Total ratio, and hippocampal volume in older adults with and without cognitive impairment. We also compared these measures to traditional scoring procedures for the same neuropsychological test. We hypothesised that process-based scores from story recall would be negatively associated with hippocampal volume, and that these scores would be a better predictor of hippocampal volume than traditional LMT and standardised composite scores, while controlling for demographics, elapsed time between MRI scan and neuropsychological assessment, sample, and APOE risk. As hypothesised, significant associations were found between story recall-derived memory scores and hippocampal volume. Although traditional and process-based scores were significantly associated with hippocampal volume, Steiger's Z-tests indicated Rr and Total ratio outperformed traditional scores for left HV, while Total ratio outperformed traditional scores for right HV. Furthermore, the model with Total ratio had significantly better fit than the models with Rr, Immediate LMT, or

Delayed LMT, when predicting left HV; whereas the models with either Total ratio or Delayed LMT showed the best fit for right HV, as indicated by AIC values. Altogether, current findings showed that tracking forgetting in LMT, as opposed to simply measuring total recall performance, is a better option when predicting hippocampal volume.

Higher R_r scores have been suggested to be the consequence of a reduction in long-term memory, due to a loss of consolidation ability, with increased reliance on a more intact ability to retain verbal information in the short-term memory (Bruno et al., 2018; Turchetta et al., 2020). The hippocampus is one of the first structures to atrophy in AD (Weiner et al., 2015). It is essential for the formation of long-term memory and for synaptic consolidation. Specifically, hippocampal volume reductions are known to occur before converting to AD and are closely related to cognitive decline (Weiner et al., 2015). Previous studies reported hippocampal atrophy to be a useful biomarker for early detection of amnesic MCI and AD (Chincarini et al., 2011; Devanand et al., 2012; Mondragon et al., 2016), for distinguishing patients, such as AD or MCI from controls (Chupin et al., 2009; Karow et al., 2010), or those who will convert from MCI to AD (Wolz et al., 2010). Therefore, it could be argued that finding predictive markers of hippocampal atrophy is crucial for the early detection of cognitive decline (Weiner et al., 2015). Our findings show that R_r applied to story recall is sensitive to hippocampal integrity, as higher R_r scores predicted lower left HV, outperforming Immediate and Delayed LMT scores, as shown by Steiger's Z-tests, while also having a significantly better model fit than traditional scores for left HV, as indicated by AIC fit statistics.

To be certain that the effects observed here were due to recency forgetting and not to total forgetting, a non-recency based R_r analogue was computed that accounted for memory loss using a ratio score with Immediate LMT and Delayed LMT, the Total ratio. Our findings demonstrated that this novel process-based measure was sensitive to hippocampal volume, as higher Total ratio scores predicted lower left HV, outperforming Immediate and Delayed LMT scores, as shown by Steiger's Z-tests. Furthermore, the model with Total ratio not only showed a significantly better fit than models with traditional scores, but it was also significantly better than the R_r model in predicting right and left HV. These findings suggest that the effects observed here, especially for left HV, could be due to total forgetting as opposed to recency forgetting. Current findings are in contrast with what Bruno et al. (2023) observed when comparing R_r to Tr as predictors of CSF biomarkers, who reported that R_r outperformed Tr in predicting biomarker outcomes associated with neurodegeneration. It is possible therefore that Tr and R_r might each be better under different circumstances, but more research is needed to determine whether they represent complementary measures. Process-based scores from story recall alone are not intended to serve a diagnostic purpose per se, yet we believe these can be useful tools for clinicians, considering their association with volumetric changes of the hippocampus, and thus, its implications of cognitive decline, in a simple and accessible way.

Against expectations, when evaluating right HV, the models with either Total ratio or Delayed LMT showed the best model fits, with no significant difference between them, as indicated by their AIC values. Typically, verbal episodic memory performance, which was assessed in this analysis, is associated with left, rather than right,

hippocampal volume (see Ezzati et al., 2016; Hardcastle et al., 2020). Additionally, greater left hippocampal reductions, compared to right hippocampal reductions, have been reported in individuals with AD (Dhikav et al., 2016; Ezzati et al., 2016; Li et al., 2016; Müller et al., 2005; Shi et al., 2009; Thompson et al., 2003, 2007; Wicking et al., 2014). Therefore, it could be argued that story-recall scores displaying sensitivity to left hippocampal volume may be more helpful, for screening and diagnostic purposes, than scores more sensitive to right hippocampal volume. It is also possible that story recall involves other cognitive abilities, such as narrative coherence, which has been reported to engage the right hippocampus (Cohn-Sheehy et al., 2021). Nevertheless, this empirical finding should be investigated in future research.

In the current study, all LMT scores were analysed as raw scores, as in previous studies examining traditional and process-based LMT measures (Bruno et al., 2021b, 2023). However, composite scores from averaged standardised tests scores, such as the Preclinical Alzheimer Cognitive Composite 4 (PACC4; for details, see Methods section and Donohue et al., 2014), are more likely to be used in clinical settings. Although composite scores have shown reduced variability and stronger associations with amyloid burden-related cognitive changes than raw scores in cognitively unimpaired individuals (Bransby et al., 2019; Jonaitis et al., 2019), the regression models with raw process-based scores showed a significantly better fit than the PACC4-analogue model when predicting left and right hippocampal volume. Specifically, the model with Total ratio was significantly better than the model with PACC4-analogue scores for both left and right HV, while the Rr model was significantly better for left HV. Current findings suggest that process-based scores derived from story recall offer an advantage over standardised composite scores when predicting hippocampal atrophy.

Considering that, as described in Table 1, most cognitively unimpaired participants were female (71%), whilst among the cognitively impaired, most were males (69% of MCI-AD; Dementia-AD, 62%), differences in sex distribution across consensus diagnoses were explored *post hoc*. Even though a detailed examination of sex-related differences is beyond the scope of the current study, potential differences in LMT scores across sexes were examined with both parametric and non-parametric tests. The *post hoc* analyses showed that in cognitive unimpaired participants, Delayed LMT scores were significantly higher in females than in males, yet no significant differences were observed in Immediate LMT, Rr or Tr scores, while in cognitively impaired participants, none of the scores significantly differed across sexes. Possible interactions between each LMT score and sex were also examined in regression analyses. We found that none of the interaction terms significantly predicted left or right HV, indicating that sex did not moderate the associations between LMT scores and hippocampal volume in participants with or without cognitive impairment.

A potential limitation of the present paper is the difference in the number of stories to be recalled, as in the WADRC sample, only story A of the LMT was tested, while stories A and B were tested in the WRAP sample. To examine if results differed in each sample, we carried out separate analyses in WADRC ($N=123$) and WRAP ($N=232$) participants, controlling for the same covariates except sample, and correcting for multiple comparisons across left and right HV (for details, see Table S3 in Supplementary materials). In WADRC only, all LMT scores were significantly associated with left and right HV, except for Immediate LMT, whereas in WRAP only, all LMT

scores were significantly associated with left and right HV, except for Rr, which was no longer a significant predictor of right HV. It appears that using a single story varied the predictive ability of Immediate LMT scores, favouring the use of both stories, while Rr scores varied for right HV only, when using both stories. However, considering that the WRAP sample had almost twice the number of participants than WADRC, these results should be taken with caution.

Another limitation is that the sample size for cognitively unimpaired participants is significantly larger than for MCI and AD, this, however, was determined by availability; yet future studies would benefit from larger sample sizes of individuals with MCI, AD, and other dementia pathologies. The samples also consisted mostly of non-Hispanic White participants. Evidence from previous studies indicates that 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 need to be replicated in a more ethnically diverse sample to confirm its generalizability.

Another possible issue is that all the evidence collected so far in support of the utility of Rr and Tr as process scores in story recall derives from a single database (WADRC; Bruno et al., 2023). While this is a drawback, in great part borne out of the novelty of these examinations, we believe it emphasises the need for further research, including in clinical settings, to explore whether these metrics are effective also beyond WRAP and ADRC.

Finally, it should be noted that practice effects were not addressed in the regression models, which could result in skewed scores for some participants. To check whether results differed in participants for whom the LMT score closest to MRI scan was collected in their first cognitive assessment visit, and thus, were unaffected by practice, the same regression analyses were carried out in these participants only ($N=124$). Results indicated that all LMT scores, except Immediate LMT, were significantly associated with left and right hippocampal volume, suggesting that only Immediate recall might be affected by practice, while the model with Total ratio showed the best fit of all (for details, see [Table S4](#) in [Supplementary materials](#)).

To summarise, the current results demonstrated the direct relationship between story recall-derived memory scores and hippocampal volume. Higher Rr scores, reflecting disproportionate recency recall in the immediate test, were associated with lower left hippocampal volume when controlling for covariates (demographics, elapsed time between MRI scan and neuropsychological assessment, sample, APOE risk score). Higher Total ratio scores, which accounted for total memory loss, significantly predicted lower left hippocampal volume, when controlling for the same covariates, and showed significantly better model fit than any other model, as per AIC fit statistics; suggesting the effects observed here could be due to total forgetting as opposed to recency forgetting. Immediate LMT and Delayed LMT scores, when adjusting for the same covariates, were significantly associated with hippocampal volume, specifically, the model with Delayed LMT showed the best fit along with the model with Total ratio for right HV; yet traditional scores were outperformed by Rr and Total ratio for left HV, and by Total ratio for right HV. Altogether, these findings showed that tracking forgetting in LMT, as opposed to simply measuring total recall performance, is a better option when predicting hippocampal volume in older adults with and without cognitive impairment.

Acknowledgements

We would like to thank the participants, the staff, and researchers at the University of Wisconsin ADRC, it is because of them that this work is possible.

Disclosure statement

TB received an honorarium to present grand rounds for Intermountain Healthcare and has and/or currently has received research funding from the NIH/NIA and the Alzheimer's Association. No other author reports any conflicts of interests or disclosures.

Funding

WADRC is funded by P50 AG033514. This secondary analysis of ADRC data was funded by a NIH-NIA (R01 144 AAI8612) grant to KDM, in which DB and RL are co-investigators. Additional funding sources that made this work possible include: R01 AG054059; R01 AG027161; R01 AG021155; R01 AG054047.

Data availability statement

WRAP and WADRC data can be requested here: <https://wrap.wisc.edu/data-requests/>.

References

- Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., Gamst, A., 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: The Journal of the Alzheimer's Association*, 7(3), 270–279. <https://doi.org/10.1016/j.jalz.2011.03.008>
- Apostolova, L. G., Morra, J. H., Green, A. E., Hwang, K. S., Avedissian, C., Woo, E., Cummings, J. L., Toga, A. W., Jack, C. R., Weiner, M. W., & Thompson, P. M. (2010). Automated 3D mapping of baseline and 12-month associations between three verbal memory measures and hippocampal atrophy in 490 ADNI subjects. *NeuroImage*, 51(1), 488–499. <https://doi.org/10.1016/j.neuroimage.2009.12.125>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Blanco-Campal, A., Diaz-Orueta, U., Navarro-Prados, A. B., Burke, T., Libon, D. J., & Lamar, M. (2021). Features and psychometric properties of the Montreal Cognitive Assessment: Review and proposal of a process-based approach version (MoCA-PA). *Applied Neuropsychology. Adult*, 28(6), 658–672. <https://doi.org/10.1080/23279095.2019.1681996>
- Bock, J. R., Russell, J., Hara, J., & Fortier, D. (2021). Optimizing cognitive assessment outcome measures for Alzheimer's disease by matching wordlist memory test features to scoring methodology. *Frontiers in Digital Health*, 3. <https://doi.org/10.3389/fdgth.2021.750549>
- Bransby, L., Lim, Y. Y., Ames, D., Fowler, C., Roberston, J., Harrington, K., Snyder, P. J., Villemagne, V. L., Salvado, O., Masters, C. L., & Maruff, P. (2019). Sensitivity of a Preclinical Alzheimer's Cognitive Composite (PACC) to amyloid β load in preclinical Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology*, 41(6), 591–600. <https://doi.org/10.1080/13803395.2019.1593949>
- Brickman, A. M., Schupf, N., Manly, J. J., Luchsinger, J. A., Andrews, H., Tang, M. X., Reitz, C., Small, S. A., Mayeux, R., DeCarli, C., & Brown, T. R. (2008). Brain morphology in older African

- Americans, Caribbean Hispanics, and whites from northern Manhattan. *Archives of Neurology*, 65(8), 1053–1061. <https://doi.org/10.1001/archneur.65.8.1053>
- 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: The Official Journal of the National Academy of Neuropsychologists*, 28(2), 95–103. <https://doi.org/10.1093/arclin/acs116>
- 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. <https://doi.org/10.1080/13803395.2016.1179721>
- Bruno, D., Kosciak, 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., Gleason, C. E., Kosciak, 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. <https://doi.org/10.1002/gps.5029>
- Bruno, D., Reichert Plaska, C., Clark, D. P., Zetterberg, H., Blennow, K., Verbeek, M. M., & Pomara, N. (2021a). CSF α -synuclein correlates with CSF neurogranin in late-life depression. *The International Journal of Neuroscience*, 131(4), 357–361. <https://doi.org/10.1080/00207454.2020.1744596>
- Bruno, D., Mueller, K. D., Betthausen, T., Chin, N., Engelman, C. D., Christian, B., Kosciak, R. L., & Johnson, S. C. (2021b). 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., Jauregi Zinkunegi, A., Pomara, N., Zetterberg, H., Blennow, K., Kosciak, R. L., Carlsson, C., Bendlin, B., Okonkwo, O., Hermann, B. P., Johnson, S. C., & Mueller, K. D. (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., Zinkunegi, A. J., Kollmorgen, G., Suridjan, I., Wild, N., Carlsson, C., Bendlin, B., Okonkwo, O., Chin, N., Hermann, B. P., Asthana, S., Zetterberg, H., Blennow, K., Langhough, R., Johnson, S. C., & Mueller, K. D. (2023). The recency ratio assessed by story recall is associated with cerebrospinal fluid levels of neurodegeneration biomarkers. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 159, 167–174. <https://doi.org/10.1016/j.cortex.2022.12.004>
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological methods & research*, 33(2), 261–304.
- 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; a Journal Devoted to the Study of the Nervous System and Behavior*, 31(4), 735–745. [https://doi.org/10.1016/s0010-9452\(13\)80024-x](https://doi.org/10.1016/s0010-9452(13)80024-x)
- Carmichael, O., Xie, J., Fletcher, E., Singh, B., & DeCarli, C. (2012). Localized hippocampus measures are associated with Alzheimer pathology and cognition independent of total hippocampal volume. *Neurobiology of Aging*, 33(6), 1124.e31–1124.e41. <https://doi.org/10.1016/j.neurobiolaging.2011.08.016>
- Chincarini, A., Bosco, P., Calvini, P., Gemme, G., Esposito, M., Olivieri, C., Rei, L., Squarcia, S., Rodriguez, G., Bellotti, R., Cerello, P., De Mitri, I., Retico, A., & Nobili, F. (2011). Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer's disease. *NeuroImage*, 58(2), 469–480. <https://doi.org/10.1016/j.neuroimage.2011.05.083>
- Chupin, M., G rardin, E., Cuingnet, R., Boutet, C., Lemieux, L., Le ricy, S., Benali, H., Garnero, L., & Colliot, O. (2009). Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI. *Hippocampus*, 19(6), 579–587. <https://doi.org/10.1002/hipo.20626>
- Cohn-Sheehy, B. I., Delarazan, A. I., Reagh, Z. M., Crivelli-Decker, J. E., Kim, K., Barnett, A. J., Zacks, J. M., & Ranganath, C. (2021). The hippocampus constructs narrative memories across distant events. *Current Biology: CB*, 31(22), 4935–4945.e7. <https://doi.org/10.1016/j.cub.2021.09.013>
- Coon, K. D., Myers, A. J., Craig, D. W., Webster, J. A., Pearson, J. V., Lince, D. H., Zismann, V. L., Beach, T. G., Leung, D., Bryden, L., Halperin, R. F., Marlowe, L., Kaleem, M., Walker, D. G., Ravid,

- R., Heward, C. B., Rogers, J., Papassotiropoulos, A., Reiman, E. M., Hardy, J., & Stephan, D. A. (2007). A high-density whole-genome association study reveals that APOE is the major susceptibility gene for sporadic late-onset Alzheimer's disease. *The Journal of Clinical Psychiatry*, 68(04), 613–618. <https://doi.org/10.4088/JCP.v68n0419>
- Darst, B. F., Kosciak, R. L., Racine, A. M., Oh, J. M., Krause, R. A., Carlsson, C. M., Zetterberg, H., Blennow, K., Christian, B. T., Bendlin, B. B., Okonkwo, O. C., Hogan, K. J., Hermann, B. P., Sager, M. A., Asthana, S., Johnson, S. C., & Engelman, C. D. (2017). Pathway-specific polygenic risk scores as predictors of amyloid- β deposition and cognitive function in a sample at increased risk for Alzheimer's disease. *Journal of Alzheimer's Disease: JAD*, 55(2), 473–484. <https://doi.org/10.3233/JAD-160195>
- 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), 382–391. <https://doi.org/10.1097/wad.0b013e318185e7fe>
- Devanand, D. P., Bansal, R., Liu, J., Hao, X., Pradhaban, G., & Peterson, B. S. (2012). MRI hippocampal and entorhinal cortex mapping in predicting conversion to Alzheimer's disease. *NeuroImage*, 60(3), 1622–1629. <https://doi.org/10.1016/j.neuroimage.2012.01.075>
- Dhikav, V., Duraisamy, S., Anand, K. S., & Garga, U. C. (2016). Hippocampal volumes among older Indian adults: Comparison with Alzheimer's disease and mild cognitive impairment. *Annals of Indian Academy of Neurology*, 19(2), 195–200. <https://doi.org/10.4103/0972-2327.176863>
- Díaz-Orueta, U., Blanco-Campal, A., & Burke, T. (2018). Rapid review of cognitive screening instruments in MCI: Proposal for a process-based approach modification of overlapping tasks in select widely used instruments. *International Psychogeriatrics*, 30(5), 663–672. <https://doi.org/10.1017/S1041610217002253>
- Divers, J., Hugenschmidt, C., Sink, K. M., Williamson, J. D., Ge, Y., Smith, S. C., Bowden, D. W., Whitlow, C. T., Lyders, E., Maldjian, J. A., & Freedman, B. I. (2013). Cerebral white matter hyperintensity in African Americans and European Americans with type 2 diabetes. *Journal of Stroke and Cerebrovascular Diseases: The Official Journal of National Stroke Association*, 22(7), e46–e52. <https://doi.org/10.1016/j.jstrokecerebrovasdis.2012.03.019>
- Donohue, M. C., Sperling, R. A., Salmon, D. P., Rentz, D. M., Raman, R., Thomas, R. G., Weiner, M., & Aisen, P. S. (2014). The preclinical Alzheimer cognitive composite: Measuring amyloid-related decline. *JAMA Neurology*, 71(8), 961–970. <https://doi.org/10.1001/jamaneurol.2014.803>
- 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. <https://doi.org/10.1080/13554794.2021.1893339>
- Ezzati, A., Katz, M. J., Zammit, A. R., Lipton, M. L., Zimmerman, M. E., Sliwinski, M. J., & Lipton, R. B. (2016). Differential association of left and right hippocampal volumes with verbal episodic and spatial memory in older adults. *Neuropsychologia*, 93(Pt B), 380–385. <https://doi.org/10.1016/j.neuropsychologia.2016.08.016>
- 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. [https://doi.org/10.1016/s0165-1781\(03\)00163-x](https://doi.org/10.1016/s0165-1781(03)00163-x)
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3), 189–198. [https://doi.org/10.1016/0022-3956\(75\)90026-6](https://doi.org/10.1016/0022-3956(75)90026-6)
- Grant, I. E., & Adams, K. M. (2009). *Neuropsychological assessment of neuropsychiatric and neuromedical disorders*(3rd revised ed.). Oxford University Press.
- Hardcastle, C., O'Shea, A., Kraft, J. N., Albizu, A., Evangelista, N. D., Hausman, H. K., Boutzoukas, E. M., Van Etten, E. J., Bharadwaj, P. K., Song, H., Smith, S. G., Porges, E. C., Dekosky, S., Hishaw, G. A., Wu, S. S., Marsiske, M., Cohen, R., Alexander, G. E., & Woods, A. J. (2020). Contributions of hippocampal volume to cognition in healthy older adults. *Frontiers in Aging Neuroscience*, 12, 593833. <https://doi.org/10.3389/fnagi.2020.593833>
- Hobel, Z., Isenberg, A. L., Raghupathy, D., Mack, W., & Pa, J. (2019). APOE ϵ 4 gene dose and sex effects on Alzheimer's disease MRI biomarkers in older adults with mild cognitive im-

- pairment. *Journal of Alzheimer's Disease: JAD*, 71(2), 647–658. <https://doi.org/10.3233/JAD-180859>
- Johnson, S. C., Kosciak, R. L., Jonaitis, E. M., Clark, L. R., Mueller, K. D., Berman, S. E., Bendlin, B. B., Engelman, C. D., Okonkwo, O. C., Hogan, K. J., Asthana, S., Carlsson, C. M., Hermann, B. P., & Sager, M. A. (2018). The Wisconsin Registry for Alzheimer's Prevention: A review of findings and current directions. *Alzheimer's & Dementia (Amsterdam, Netherlands)*, 10, 130–142. <https://doi.org/10.1016/j.dadm.2017.11.007>
- Jonaitis, E. M., Kosciak, R. L., Clark, L. R., Ma, Y., Betthausen, T. J., Berman, S. E., Allison, S. L., Mueller, K. D., Hermann, B. P., Van Hulle, C. A., Christian, B. T., Bendlin, B. B., Blennow, K., Zetterberg, H., Carlsson, C. M., Asthana, S., & Johnson, S. C. (2019). Measuring longitudinal cognition: Individual tests versus composites. *Alzheimer's & Dementia (Amsterdam, Netherlands)*, 11, 74–84. <https://doi.org/10.1016/j.dadm.2018.11.006>
- Karow, D. S., McEvoy, L. K., Fennema-Notestine, C., Hagler, D. J., Jennings, R. G., Brewer, J. B., Hoh, C. K., & Dale, A. M. (2010). Relative capability of MR imaging and FDG PET to depict changes associated with prodromal and early Alzheimer disease. *Radiology*, 256(3), 932–942. <https://doi.org/10.1148/radiol.10091402>
- Langhough Kosciak, R., Hermann, B. P., Allison, S., Clark, L. R., Jonaitis, E. M., Mueller, K. D., Betthausen, T. J., Christian, B. T., Du, L., Okonkwo, O., Birdsill, A., Chin, N., Gleason, C., & 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*, 13, Article 688478. <https://doi.org/10.3389/fnagi.2021.688478>
- Lee, I. A., & Preacher, K. J. (2013, September). Calculation for the test of the difference between two dependent correlations with one variable in common [Computer software]. <http://quantpsy.org>
- Li, X., Li, D., Li, Q., Li, Y., Li, K., Li, S., & Han, Y. (2016). Hippocampal subfield volumetry in patients with subcortical vascular mild cognitive impairment. *Scientific Reports*, 6(1), 20873. <https://doi.org/10.1038/srep20873>
- Libon, D. J., Swenson, R., Lamar, M., Price, C. C., Baliga, G., Pascual-Leone, A., Au, R., Cosentino, S., & Andersen, S. L. (2022). The Boston process approach and digital neuropsychological assessment: past research and future directions. *Journal of Alzheimer's Disease: JAD*, 87(4), 1419–1432. <https://doi.org/10.3233/JAD-220096>
- McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., 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: The Journal of the Alzheimer's Association*, 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.
- Mondragón, J. D., Celada-Borja, C., Barinagarrementeria-Aldatz, F., Burgos-Jaramillo, M., & Barragán-Campos, H. M. (2016). Hippocampal volumetry as a biomarker for dementia in people with low education. *Dementia and Geriatric Cognitive Disorders Extra*, 6(3), 486–499. <https://doi.org/10.1159/000449424>
- Müller, M. J., Greverus, D., Dellani, P. R., Weibrich, C., Wille, P. R., Scheurich, A., Stoeter, P., & Fellgiebel, A. (2005). Functional implications of hippocampal volume and diffusivity in mild cognitive impairment. *NeuroImage*, 28(4), 1033–1042. <https://doi.org/10.1016/j.neuroimage.2005.06.029>
- Murdock, B. B. Jr. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, 64(5), 482–488. <https://doi.org/10.1037/h0045106>
- Patenaude, B., Smith, S. M., Kennedy, D. N., & Jenkinson, M. (2011). A Bayesian model of shape and appearance for subcortical brain segmentation. *NeuroImage*, 56(3), 907–922. <https://doi.org/10.1016/j.neuroimage.2011.02.046>
- Schmidt, M. (1996). *Rey auditory verbal learning test: A handbook* (Vol. 17). Western Psychological Services.

- Shi, F., Liu, B., Zhou, Y., Yu, C., & Jiang, T. (2009). Hippocampal volume and asymmetry in mild cognitive impairment and Alzheimer's disease: Meta-analyses of MRI studies. *Hippocampus*, 19(11), 1055–1064. <https://doi.org/10.1002/hipo.20573>
- Stavitsky, K., Du, Y., Seichepine, D., Laudate, T. M., Beiser, A., Seshadri, S., Decarli, C., Wolf, P. A., & 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. <https://doi.org/10.1159/000313443>
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87(2), 245–251. <https://doi.org/10.1037/0033-2909.87.2.245>
- Talamonti, D., Kosciak, R., Johnson, S., & Bruno, D. (2020). Predicting early mild cognitive impairment with free recall: The primacy of primacy. *Archives of Clinical Neuropsychology : The Official Journal of the National Academy of Neuropsychologists*, 35(2), 133–142. <https://doi.org/10.1093/arclin/acz013>
- Thompson, P. M., Hayashi, K. M., de Zubicaray, G., Janke, A. L., Rose, S. E., Semple, J., Herman, D., Hong, M. S., Dittmer, S. S., Doddrell, D. M., & Toga, A. W. (2003). Dynamics of gray matter loss in Alzheimer's disease. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 23(3), 994–1005. <https://doi.org/10.1523/JNEUROSCI.23-03-00994.2003>
- Thompson, P. M., Hayashi, K. M., Dutton, R. A., Chiang, M.-C., Leow, A. D., Sowell, E. R., De Zubicaray, G., Becker, J. T., Lopez, O. L., Aizenstein, H. J., & Toga, A. W. (2007). Tracking Alzheimer's disease. *Annals of the New York Academy of Sciences*, 1097(1), 183–214. <https://doi.org/10.1196/annals.1379.017>
- 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 dementia. *Journal of Alzheimer's Disease: JAD*, 66(2), 461–470. <https://doi.org/10.3233/JAD-180690>
- Turchetta, C. S., De Simone, M. S., Perri, R., Fadda, L., Caruso, G., De Tollis, M., Caltagirone, C., & 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: JAD*, 73(4), 1295–1304. <https://doi.org/10.3233/JAD-190509>
- Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Cedarbaum, J., Green, R. C., Harvey, D., Jack, C. R., Jagust, W., Luthman, J., Morris, J. C., Petersen, R. C., Saykin, A. J., Shaw, L., Shen, L., Schwarz, A., Toga, A. W., & Trojanowski, J. Q. (2015). 2014 Update of the Alzheimer's Disease Neuroimaging Initiative: A review of papers published since its inception. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, 11(6), e1–e120. <https://doi.org/10.1016/j.jalz.2014.11.001>
- Wechsler, D. (1981). *Wechsler adult intelligence scale-revised (WAIS-R)*. Psychological Corporation.
- Wechsler, D. (1987). *WMS-R: Wechsler memory scale-revised*. Psychological Corporation.
- Wicking, M., Nees, F., & Steiger, F. (2014). Neuropsychological measures of hippocampal function. *Frontiers of Neurology and Neuroscience*, 34, 60–70. <https://doi.org/10.1159/000356425>
- Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology*, 55, 235–269. <https://doi.org/10.1146/annurev.psych.55.090902.141555>
- Wixted, J. T., & Cai, D. J. (2013). Memory consolidation. In *The Oxford handbook of cognitive neuroscience* (vol. 1, pp. 436–456).
- Wolz, R., Heckemann, R. A., Aljabar, P., Hajnal, J. V., Hammers, A., Lötjönen, J., & Rueckert, D. (2010). Measurement of hippocampal atrophy using 4D graph-cut segmentation: Application to ADNI. *NeuroImage*, 52(1), 109–118. <https://doi.org/10.1016/j.neuroimage.2010.04.006>