



## Research Report

# Proper names from story recall are associated with beta-amyloid in cognitively unimpaired adults at risk for Alzheimer's disease

Kimberly D. Mueller <sup>a,b,c,\*</sup>, Rebecca L. Kosciak <sup>b</sup>, Lianlian Du <sup>b</sup>,  
Davide Bruno <sup>d</sup>, Erin M. Jonaitis <sup>b</sup>, Audra Z. Kosciak <sup>a,b</sup>,  
Bradley T. Christian <sup>c,e,f</sup>, Tobey J. Betthausen <sup>b,c,f</sup>, Nathaniel A. Chin <sup>b,c</sup>,  
Bruce P. Hermann <sup>b,g</sup> and Sterling C. Johnson <sup>b,c,h</sup>

<sup>a</sup> Department of Communication Sciences and Disorders, University of Wisconsin – Madison, Madison, WI, USA

<sup>b</sup> Wisconsin Alzheimer's Institute, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA

<sup>c</sup> Wisconsin Alzheimer's Disease Research Center, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA

<sup>d</sup> School of Natural Sciences and Psychology, Liverpool John Moores University, Liverpool, UK

<sup>e</sup> Waisman Laboratory for Brain Imaging and Behavior, University of Wisconsin – Madison, Madison, WI, USA

<sup>f</sup> Department of Medical Physics, University of Wisconsin – Madison, Madison, WI, USA

<sup>g</sup> Department of Neurology, University of Wisconsin School of Medicine and Public Health, University of Wisconsin – Madison, Madison, WI, USA

<sup>h</sup> Geriatric Research Education and Clinical Center, William S. Middleton Veterans Hospital, Madison, WI, USA

## ARTICLE INFO

## Article history:

Received 19 May 2020

Reviewed 25 June 2020

Revised 11 July 2020

Accepted 21 July 2020

Action editor Peter Garrard

Published online 31 July 2020

## Keywords:

Alzheimer's disease

Beta-amyloid

MCI

Semantic memory

Proper names

Story recall

Language

Early detection

Logical Memory

## ABSTRACT

Due to advances in the early detection of Alzheimer's disease (AD) biomarkers including beta-amyloid (A $\beta$ ), neuropsychological measures that are sensitive to concurrent, subtle changes in cognition are critically needed. Story recall tasks have shown sensitivity to early memory declines in persons with mild cognitive impairment (MCI) and early stage dementia, as well as in persons with autosomal dominantly inherited AD up to 10 years prior to a dementia diagnosis. However, the evidence is inconclusive regarding relationships between evidence of A $\beta$  and story recall measures. Because story recall tasks require the encoding and delayed retrieval of several lexical-semantic categories, such as proper names, verbs, and numerical expressions, and because lexical categories have been shown to be differentially impaired in persons with MCI, we focused on item-level analyses of lexical-semantic retrieval from a quintessential story recall task, Logical Memory from the Wechsler Memory Scale-Revised. Our objective was to investigate whether delayed recall of lexical categories (proper names, verbs and/or numerical expressions), as well as the traditional total score measure, was associated with “preclinical AD,” or cognitively unimpaired adults with positive A $\beta$  deposition on positron emission tomography (PET) neuroimaging using Pittsburgh Compound B (PiB). We developed an item-level scoring system, in which we parsed items into lexical categories and examined the immediate and delayed recall of these lexical categories from 217 cognitively unimpaired participants from the

\* Corresponding author. Department of Communication Sciences and Disorders, University of Wisconsin – Madison, Madison, WI, USA.  
E-mail address: [kdmueller@wisc.edu](mailto:kdmueller@wisc.edu) (K.D. Mueller).

<https://doi.org/10.1016/j.cortex.2020.07.008>

0010-9452/Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Wisconsin Registry for Alzheimer's Prevention (WRAP). We performed binary logistic regression models with story recall score as predictor and A $\beta$  status (positive/negative) as the outcome. Using baseline Logical Memory data, proper names from delayed story recall were significantly associated with A $\beta$  status, such that participants who recalled more proper names were less likely to be classified as PiB(+) (odds ratio = .58,  $p$  = .01). None of the other story recall variables, including total score, were associated with PiB status. Secondary analyses determined that immediate recall of proper names was not significantly associated with A $\beta$ , suggesting a retrieval deficit rather than that of encoding. The present findings suggest that lexical semantic retrieval measures from existing story recall tasks may be sensitive to A $\beta$  deposition, and may provide added utility to a widely-used, long-standing neuropsychological test for early detection of cognitive decline on the AD continuum.

Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Alzheimer's disease (AD) is a neurodegenerative process defined by the presence of beta amyloid plaques (A $\beta$ ) and neurofibrillary tau tangles, resulting in neuronal cell death, eventual cognitive decline, and dementia (Braak & Braak, 1991; Jack Jr et al., 2018). Although Alzheimer's dementia has a heterogeneous presentation, the gradual and insidious cognitive decline leading to dementia typically involves documentable changes to both episodic memory, especially the learning of new material, and semantic memory, particularly the timely and efficient word retrieval from a variety of lexical-semantic categories. These patterns of cognitive decline match the evidenced early distribution of pathologic tau, which typically begins in the medial-temporal lobe, particularly the entorhinal cortex which mediates episodic learning and memory, and the parahippocampal gyrus which mediates semantic memory. There is accruing evidence showing that tests of both episodic memory [such as the Rey Auditory Verbal Learning Test (R-AVLT) or Logical Memory story recall from the Wechsler Memory Scale (WMS)] (Mormino & Papp, 2018) and tests of semantic memory (such as category verbal fluency or face-name association tasks) (Rentz et al., 2011) are associated with very early amyloid plaque accumulation, even antecedent to the onset of clinical cognitive impairment.

Because the AD neuropathological processes are suspected to begin decades before clinically significant cognitive decline, interventions are expected to be most effective at the earliest stages of disease, before considerable neurodegeneration has occurred. As a result, sensitive measures of early cognitive change are needed to identify the individuals who will be most likely to benefit from interventions, and as a means of monitoring response to treatment in clinical trials. The National Institute on Aging-Alzheimer's Association (NIA-AA) research framework for AD indicates that cognitively unimpaired individuals with A $\beta$  in vivo biomarker positivity are characterized as in the early stage of Alzheimer's pathologic change (Jack Jr et al., 2018), and therefore considered to be an ideal group to study for early detection of cognitive decline due to AD and for response to treatment in intervention trials.

Studies evaluating relationships between A $\beta$  deposition using positron emission tomography (PET) imaging and cognition in at-risk, yet otherwise healthy individuals have shown mixed results. While some cross-sectional studies show an association between A $\beta$  and cognition (Donohue et al., 2017), other studies have found no such associations (Jansen et al., 2018). Longitudinal studies evaluating the change in cognition in cognitively unimpaired adults in relation to A $\beta$  deposition have shown positive relationships, such that A $\beta$  positive individuals tend to show faster prospective declines on cognitive composite scores and on tests of verbal learning and memory than individuals who were A $\beta$  negative (Betthausen et al., 2020; Clark, Racine, et al., 2016; Farrell et al., 2017; Rabin et al., 2018).

Although tests of cognitive function typically focus on episodic memory in AD longitudinal studies, some researchers argue that semantic memory, the long-term storage of conceptual knowledge including meanings and lexical (word) information, may be a more sensitive target for prodromal AD (Venneri et al., 2016). Because the earliest AD pathology (neurofibrillary tau tangles) and neurodegeneration tend to occur in the entorhinal cortex and parahippocampus (Braak et al., 1993), areas known to be neural correlates of semantic processing (Venneri et al., 2016), measurement of semantic processing may be a particularly sensitive target.

Semantic memory deficits have been well documented in AD dementia using a variety of neuropsychological tests, including confrontation naming (Williams et al., 1989) category fluency (Garrard et al., 1998; Rascovsky et al., 2007), and visual-verbal semantic matching (Hodges & Patterson, 1995). Importantly, advances in AD biomarker detection (detection of amyloid plaques and tau neurofibrillary tangles via imaging and cerebrospinal fluid analysis) have allowed researchers to evaluate semantic memory differences between healthy controls and adults in the preclinical phase of AD (i.e., before noticeable functional decline or noticeable symptoms of cognitive impairment). For example, Papp et al. (2016) showed that participants with evidence of elevated A $\beta$  deposition on PET using tracer Pittsburgh Compound B (PiB) showed steeper decline in category fluency than participants without amyloidosis, despite none of the participants in either group

meeting criteria for clinical impairment of cognitive function (Papp et al., 2016).

While category fluency tasks comprise one sensitive way of measuring semantic memory and retrieval, researchers have also found specific lexical categories (i.e., nouns, verbs, proper names) to be differentially affected in typical aging and dementia. For example, proper name retrieval (as compared to regular nouns) is particularly problematic in typical aging (Cohen, 1990) but accentuated in mild dementia due to probable AD (Semenza et al., 2000). Lesion studies and functional imaging research have shown that the anterior inferior temporal lobe may be more involved in the retrieval of proper names than other grammatical classes (Gainotti, 2007) and therefore may be sensitive to the early neuropathology associated with AD. Verb disadvantages over nouns have also been documented in AD dementia in tasks such as verb versus object fluency (Davis et al., 2010) or in action versus object picture-naming (Druks et al., 2006), but the neural substrates of verb processing are not fully understood. Activation of the dorsolateral prefrontal cortex in verb processing has been documented by several researchers (Cappa et al., 2002; Cotelli et al., 2011; Shapiro et al., 2006), but this region is often spared in the earliest stages of AD. It is possible that association pathways involving semantic memory and retrieval may partially explain these verb deficits (Beber et al., 2019). Perhaps more importantly, the association between verb processing in cognitively unimpaired adults with biomarker evidence of AD pathology is unknown.

In order to explore this question, we capitalized on a widely used memory test, the Logical Memory story recall task from the Wechsler Memory Scale-Revised (Wechsler, 1987). Logical Memory, although an episodic memory test, arguably provides a unique measure of discourse from which variations in semantic categories of spontaneous recall can be derived. In this novel method, we obtained item-level data from participants from the Wisconsin Registry for Alzheimer's Prevention (WRAP), a longitudinal cohort study of participants enriched for AD risk who are free of dementia and preclinical conditions (mild cognitive impairment – MCI). We separated correctly recalled items into lexical categories including proper names (names of people or places), verbs, and numerical expressions. We had two aims, and hypothesized that recall of either proper names, verbs or numerical expressions would be associated with both 1) progression from cognitively unimpaired to clinical MCI and/or 2) amyloid positivity as evidenced by PiB PET imaging. In sensitivity analyses designed to understand how our novel predictors compared to more traditional measures, we compared the relationship with A $\beta$  status and novel story recall variables to that of a commonly used cognitive composite score, the Preclinical Alzheimer's Cognitive Composite (PACC; Donohue et al., 2014).

## 2. Materials and methods

### 2.1. Participants

The study sample was drawn from the Wisconsin Registry for Alzheimer's Prevention (WRAP) study, an ongoing longitudinal cohort study examining risk factors, lifestyles, and

cognition in participants who are late-middle-aged and enriched for parental history of AD. A subset of WRAP participants also participates in AD biomarker studies (e.g., imaging and cerebrospinal fluid studies). The first follow-up visit occurred four years after baseline, and subsequent visits occurred every two years thereafter (see Johnson et al., 2018 for detailed information about the WRAP sample). Logical Memory was first added to the test battery in 2007, and this item-level analysis project was added in summer of 2018. Data entry of item-level information was prioritized to focus on those who had provided PET biomarker data and/or had progressed from an unimpaired to impaired status. Once those records were entered, data entry continued for all other charts. At the time of these analyses, participants were selected for these analyses if they were free of dementia at any visit, free of neurological diagnoses (stroke, Parkinson disease, multiple sclerosis, or epilepsy/seizure disorder), had English as their native language, and had their cognitive status reviewed via consensus conference ( $n = 696$ ). A second subset of individuals was selected for the second aim who had completed amyloid PET scans (completed at median visit 3 of WRAP study visit) and met the above-described inclusion criteria ( $n = 217$ ). All activities for this study were approved by the University of Wisconsin – Madison Institutional Review Board and completed in accordance with the Helsinki Declaration. The conditions of our ethics approval do not permit public archiving of anonymized study data. Readers seeking access to the data should contact the lead author ([kdmueller@wisc.edu](mailto:kdmueller@wisc.edu)), or the executive committee of the Wisconsin Registry for Alzheimer's Prevention by submitting a formal data request at the following URL: <https://wrap.wisc.edu/data-requests>. Access will be granted to named individuals in accordance with ethical procedures governing the reuse of sensitive data. Specifically, requestors must meet the following condition to obtain access: completion of a formal data sharing agreement.

No part of the study procedures or analyses for this research was preregistered prior to the research being conducted. We report how we determined our sample size, all data exclusions, all inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study.

### 2.2. Experimental variables from Logical Memory story recall

Logical Memory a story recall subtest of the Wechsler Memory Scale – Revised (WMS-R; Wechsler, 1987) is a standardized, norm-referenced neuropsychological test that evaluates both learning and episodic memory. Standardized procedures for test administration were followed in accordance with the WMS-R manual. In this task, a first short story (Story A) consisting of a few lines was read aloud to the participant, and the participant was asked to retell the story immediately with the following instructions: “tell the story back to me, using as close to the same words as you can remember; you should tell me all you can, even if you are not sure.” The same procedure was repeated with a second story (Story B). The participant was then asked to recall both stories again after a 25–35 minute delay. Standardized scoring procedures per the

WMS-R manual were followed. Each story contains 25 “idea units”, consisting of a target word, phrase, or idea; as the participant recalled the story, the examiner was instructed to underline correctly expressed idea units, and to notate alternative wordings for scoring later. The scoring criteria allows for some alterations of idea units; for example, in a hypothetical example, the phrase “raise” is acceptable for the idea unit “bring up (her children).” Some responses must be verbatim in order to be accepted, including proper names (names of people or places) and numerical expressions (e.g., “twenty-five cents” or “three dogs”).

In order to create the lexical category variables, we first set up a database such that each idea unit was a separate variable and was coded as 1 (correctly recalled) or 0 (not recalled). Next, we parsed each idea unit into a lexical category, by running a transcript of the Logical Memory stories through a part-of-speech tagger (Computerized Language Analysis Program “CLAN”, MacWhinney, 2014). This allowed an objective measure of part-of-speech; for example, “Beth” was coded as “proper noun” and “worked” as “verb|past tense.” Finally, we assigned each idea unit into one of three lexical categories, based on our theory-driven hypotheses: proper names, verbs, and numerical expressions. See Fig. 1 for a schematic of the lexical categories and the idea units they contain (we used a hypothetical story in order to protect copyright and test integrity). For primary analyses, we used data from the first administration of the story recall task as the predictors (median = visit 2, range = visits 1–3). Specifically, we summed the semantic categories from stories A and B for each of the immediate and delayed recall conditions resulting in a semantic category total immediate recall, and semantic category total delayed recall. We then converted all variables to standardized z-scores, so that Logical Memory total score and the lexical categories scores were comparable among models.

### 2.3. Secondary predictor: the Preclinical Cognitive Composite Score (PACC)

As first described by Donohue et al. (2014), the PACC is an average of commonly used tests of episodic memory, executive function, and global cognition, including total recall from the Free and Cued Selective Reminding Test, delayed recall score from Logical Memory IIa, Digit Symbol Substitution Test,

and the Mini-Mental Status Examination (MMSE; Folstein, Robins, & Helzer, 1983) total score. Recently our group compared several versions of the PACC, as well as theoretically derived cognitive composite scores, and found that within the WRAP group a version referred to as “PACC-3” was sensitive to cognitive decline. This version omits the MMSE and includes an average of standardized scores from total recall from Rey-Auditory Verbal Learning Test (R-AVLT), delayed recall from Logical Memory A & B, and Digit Symbol Substitution (Jonaitis et al., 2019). In secondary analyses, we examined whether the PACC-3 from the first available visit (median visit = 2) would predict more variability in PiB positivity than the lexical categories variables.

### 2.4. Cognitive diagnosis

Because the WRAP cohort is relatively young (mean age at baseline = 54; mean age at most recent visit = 67), and the majority of participants are cognitively unimpaired, one priority for the WRAP study has been to develop methods that detect subtle declines in cognition, despite the fact that participants’ scores may still fall within the norm-referenced range of normal. Researchers from our group have therefore developed internal “robust” norms, in which the normative group consists of non-declining WRAP participants over time (Clark, Kosciak, et al., 2016; Kosciak et al., 2014). Algorithms were developed to “flag” participants whose cognitive performance falls outside the range of the robust norms at any particular visit; flagged participants’ cases are then brought to a consensus review committee consisting of dementia specialists, including neuropsychologists, physicians, nurse practitioners, and physician assistants. The details of this consensus review are described elsewhere (Johnson et al., 2018; Kosciak et al., 2016); in brief, the committee reviews cognitive data, medical and social histories, mental health status, and self- and informant-reported assessments of mood, cognitive and functional status across all available visits. Clinicians and researchers are blinded to biomarker status prior to making their cognitive status diagnoses. A diagnosis of “clinical MCI” is assigned if the consensus panel agrees that the participant met the criteria for MCI from the National Institute on Aging-Alzheimer’s Association (NIA-AA; Albert et al., 2011). The research category “cognitively

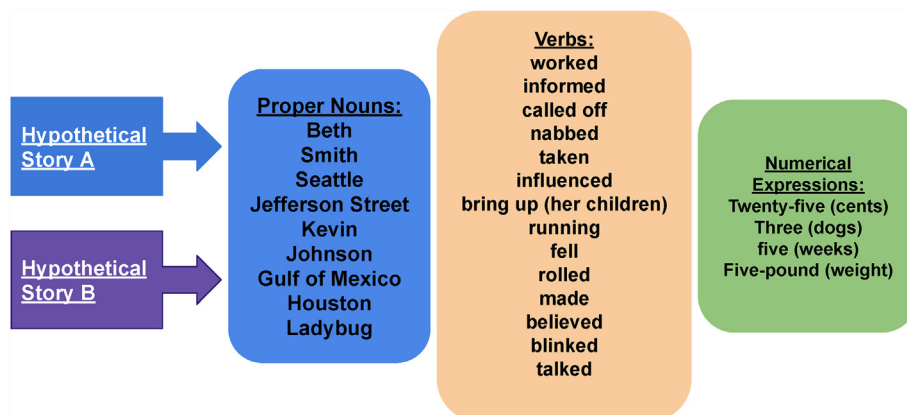


Fig. 1 – Schematic showing assignment of idea units to one of three lexical categories, based on two hypothetical stories.



unimpaired-declining” is assigned to participants who show lower than expected performance ( $>1.5$  standard deviations below internal robust norms), but few or no subjective complaints or functional deficits. This category is similar to clinical stage 2 from the 2018 diagnostic AD framework (Jack Jr et al., 2018), but without consideration of biomarker status, and was renamed from “early MCI” (e.g., Mueller et al., 2018) in order to align with the research definitions outlined in the NIA-AA framework. Participants who do not meet the above-noted criteria or are not flagged by the algorithm are classified as “cognitively unimpaired-stable.” For the purposes of these analyses, we were interested in those participants who had a diagnosis of “cognitively unimpaired-stable” (CU-S) at the first visit at which story recall was administered, and whether or not they progressed to clinical MCI at their most recent visit.

## 2.5. [ $^{11}\text{C}$ ]PiB PET

For the second research question, participants were selected who underwent a 70-min dynamic [ $^{11}\text{C}$ ]PiB scan on a Siemens EXACT HR+ scanner [and prior to 2015, a T1-weighted magnetic resonance imaging (MRI) scan on a GE 3.0 T MR750 using an 8-channel head coil]. Neuroimaging was completed, on average, 1.4 years (SD = 1.4) after the first visit at which story recall was administered (median visit 2). [ $^{11}\text{C}$ ]PiB radiosynthesis, acquisition and reconstruction parameters, image processing and quantification have been described previously (Johnson et al., 2014). Briefly, the PET time series was motion corrected, denoised, and co-registered to T1-w MRI. Time-activity curves were extracted from the co-registered PET data in subject MRI space from gray matter restricted Anatomic Labeling atlas (Tzourio-Mazoyer et al., 2002) regions of interest (ROIs) warped to MRI space and used to estimate ROI-level distribution volume ratios (DVRs); Logan Graphical Analysis, cerebellum GM reference region,  $k_2 = .149 \text{ min}^{-1}$  (Logan et al., 1996; Lopresti et al., 2005). PiB positivity was ascertained by applying a threshold to the mean DVR (global PiB) across eight bilateral ROIs (global DVR  $\geq 1.19$ ) (Racine et al., 2016; Racine et al., 2016).

## 2.6. Statistical analyses

All statistical analyses were performed in R (Team, 2019) or SPSS version 25. R code for analyses is available for viewing at the following URL: [https://osf.io/vjmy9/?view\\_only=472fb3b2d4784ee9a10c6d794d0295a0](https://osf.io/vjmy9/?view_only=472fb3b2d4784ee9a10c6d794d0295a0). Linear mixed effects models were run with the R package ‘lme4’ (Bates et al., 2015). Significance level was set at  $p < .05$ . To account for multiple comparisons, a false discovery rate approach was applied, which calculates the proportion of false positives among those tests for which the null hypothesis is rejected. This aids in controlling error while preserving power, by adjusting the  $p$ -value criterion for significance based on the number of tests performed (Curran-Everett, 2000).

### 2.6.1. Relationship between lexical categories from story recall and progression to clinical MCI

A series of binomial logistic regression analyses were performed to model the relationship between progression to clinical MCI at the most recent visit (progression = 1; no progression = 0) and each of four story recall scores from the

first available visit, considered in separate models: the number of words recalled in each lexical category, and the total number recalled. Covariates included age, sex, and Wide-Range Achievement Test – 3 (WRAT-3; Wilkinson, 1993) reading subtest were included in each model. We use the WRAT-3 reading test as a proxy for educational attainment as described elsewhere (Manly et al., 2002). Within each set of models, we used the Akaike information criterion (AIC) to assess whether any of the Logical Memory lexical sub-scores provided superior fit to that obtained using the traditional total score. The AIC is an estimator of the relative quality of statistical models, with the preferred model containing the minimum AIC (Bozdogan, 1987). In order to determine if the model fits were comparable, we followed guidance from Burnham and Anderson (2002), such that models were deemed comparable when the difference in AIC relative to the smaller AIC model was  $<2$  (Burnham & Anderson, 2002).

### 2.6.2. Relationship between lexical categories and A $\beta$ positivity

Similar binomial logistic regression analyses were performed to model the relationship between amyloid status at the most recent PET scan (PiB(+) = 1; PiB(–) = 0) and the lexical category scores and total scores. Covariates included age, sex, WRAT-3 and APOE- $\epsilon 4$  status (APOE- $\epsilon 4$  positive = at least one allele). Primary analyses focused on delayed recall scores in order to capture retrieval from semantic memory versus encoding; immediate recall scores were secondary. Again, AICs were compared between models to identify the best fitting model(s). In sensitivity analyses, we repeated the analysis substituting the Logical Memory variables from the most recent visit for the baseline Logical Memory variables. In additional exploratory analyses, we examined 1) whether models including multiple lexical variables or a lexical variable and remaining total score predicted more variability than just single lexical variables and covariates; and 2) whether there was a relationship between longitudinal story recall variables and A $\beta$  at most recent PET scan.

### 2.6.3. Bootstrapping to determine the significance of association and the stability of findings

We applied the following bootstrapping technique to approximate a distribution and test the stability of regression coefficient estimates for each of the eight logistic regression models in 2.5.2. First, we resampled the first available visit data with replacement to obtain 1000 samples the same size as our analysis sample. We then ran the batch of logistic regression models and stored the coefficients from each. We repeated these two steps 200 times. We used the resulting distributions of parameter estimates to obtain 95% confidence intervals.

## 3. Results

### 3.1. The relationship between lexical categories and progression to clinical MCI

Participant demographics and clinical characteristics for the first study, the relationship between lexical categories and progression to clinical MCI, are presented in Table 1. 696

**Table 1 – Demographics and clinical characteristics of participants with cognitive status and story recall.**

	Total	Cognitively unimpaired-stable	Progressed to clinical MCI at most recent visit	p
n (with cognitive status)	696	678	18	
Age at visit 2	58.2 (6)	58.1 (6)	63.8 (5)	<.001
Age at most recent visit (median = vis. 5)	65 (6.8)	64.9 (7)	70.5 (5)	.001
Sex (% female)	485 (69.7)	477 (70)	8 (44)	.04
Race (%)				.66
African-American	51 (7.3)	50 (7)	1 (6)	
Non-Hispanic white	628 (90.4)	612 (90)	16 (89)	
Other	17 (.02)	16 (2.1)	1 (5.6)	
Parental history AD dementia (%)	532 (76.4)	517 (76)	15 (83)	.68
WRAT-3 Reading Standard Score	107.5 (9)	107.6 (9)	105.1(11)	.25
Total years education	16 (2)	16 (3)	16 (2.8)	.57
APOE-ε4 carriers (%)	285 (40.9)	279 (41)	6 (33)	.67
MMSE	29.4 (.9)	29.4 (.9)	29.6 (.9)	.45
R-AVLT total	50 (8.9)	50.3 (9)	36.7 (8)	<.001
Logical Memory Total delayed recall (range 0–50)	25.4 (7.5)	25.7 (7)	15.6 (7)	<.001
Logical Memory Proper Names (range 0–9)	5 (2)	4.8 (2)	2.5 (2)	<.001
Logical Memory Verbs (range 0–14)	6 (2)	6.5 (2)	4.2 (2)	<.001
Logical Memory Numerical Expressions (range 0–4)	2.4 (1)	2.4 (1)	1.2 (.8)	<.001
Abbreviations: WRAT-3 = Wide Range Achievement Test-3 Reading Subtest (Wilkinson, 1993); MMSE = Mini-Mental Status Examination (Folstein et al., 1983); R-AVLT = Rey Auditory Verbal Learning Test (Schmidt, 1996); Logical Memory = subtest from the Wechsler Memory Scale-Revised (WMS-R; Wechsler, 1987). Items in <b>boldface</b> indicate statistical significance at $p < .05$ .				

participants had a mean age at visit 2 of 58 (sd = 6) and mean age at most recent visit was 65 (sd = 6.8). 70% were female, 90% were non-Hispanic white, and 76% had a parental family history of AD dementia. Eighteen of the 696 participants (2.6%) progressed from cognitively unimpaired-stable at the first story recall visit (median = visit 2) to clinical MCI at the most recent visit (median = visit 5). Those who progressed to MCI did not differ from those who were unimpaired at their last visit by race, parental history of AD, WRAT-3 Reading, years of education, APOE-ε4 status, or MMSE scores from visit 2. The participants who progressed to MCI were significantly older than the

CU-S participants (64 vs 58 at first available story recall,  $p < .001$ ), and had a significantly higher percentage of males (56% of participants with MCI were male vs 30% of CU-S participants,  $p = .04$ ). The MCI participants performed significantly worse on R-AVLT, Logical Memory total score, and all Logical Memory experimental lexical categories at visit 2, adjusting for age, sex and WRAT-3 reading (when all participants in this sample were deemed cognitively unimpaired – stable).

Results of the logistic regression analyses using story recall delayed total score and lexical categories as predictors for conversion to clinical MCI are presented in Table 2. All

**Table 2 – Binomial logistic regression results of story recall variables at visit 2 predicting most recent cognitive status (cognitively unimpaired versus mild cognitive impairment).**

Predictors	Proper names			Verbs			Numbers			Total score		
	Odds ratios	CI	p	Odds ratios	CI	p	Odds ratios	CI	p	Odds ratios	CI	p
(Intercept)	.00	.00–.10	<b>.011</b>	.00	.00–.07	<b>.009</b>	.00	.00–.03	<b>.005</b>	.00	.00–.03	<b>.005</b>
Age	1.17	1.05–1.29	<b>.004</b>	1.19	1.07–1.33	<b>.001</b>	1.19	1.07–1.32	<b>.001</b>	1.17	1.05–1.30	<b>.005</b>
Sex	.40	.15–1.08	.069	.31	.11–.84	<b>.022</b>	.46	.17–1.26	.129	.35	.12–.98	<b>.046</b>
WRAT-Reading	.96	.87–1.06	.416	.95	.86–1.05	.308	.96	.87–1.06	.387	.99	.89–1.10	.845
Proper Names	.38	.22–.67	<b>.001</b>									
Verbs				.40	.23–.68	<b>.001</b>						
Numbers							.31	.17–.57	<b>&lt;.001*</b>			
Total score										.29	.16–.51	<b>&lt;.001</b>
AIC	39.08			39.75			40.26			38.36†		
AICΔ	+ .72			+ 1.39			+ 1.9			N/A		
Cox & Snell's R <sup>2</sup>	.055/.248			.054/.244			.061/.277			.068/.307		
Nagelkerke's R <sup>2</sup>												

Visit 2 = first available story recall data. Most recent cognitive status was at median visit 5. “Cognitively unimpaired” versus “MCI” was determined via robust normative data review and consensus conference (Clark, Racine, et al., 2016; Johnson et al., 2018; Kosciak et al., 2014, 2016). Abbreviations: WRAT-3 = Wide Range Achievement Test-3 Reading Subtest (Wilkinson, 1993); CI = 95% confidence interval; AIC = Akaike information criterion. † = model with the lowest AIC, against which change in AIC (AICΔ) is calculated. Items in **boldface** = statistically significant at  $p < .05$ . Cox & Snell's R<sup>2</sup> and Nagelkerke's R<sup>2</sup> = the proportion of the explained variance in the random effect of the full (conditional model) compared to the null (unconditional model).

variables from story recall (proper names, verbs, numerical expressions and total score) were significant predictors of conversion to clinical MCI at the most recent visit. When comparing the AICs among all four models, the model using total score as predictor was the better fitting model (AIC = 38.36); however, the four models' AICs did not differ by more than 2, indicating similar fits overall (Burnham & Anderson, 2002). Similarly, Cox & Snell's  $R^2$  and Nagelkerke's  $R^2$  indicate the proportion of the explained variance in the random effect of the full (conditional model) compared to the null (unconditional) model; the smaller the ratio, the greater the improvement of the model. As with the AIC, the variance is explained slightly better by the total score model.

### 3.2. The relationship between lexical categories and A $\beta$ status

Participant demographics and clinical characteristics of participants with PET-PiB imaging data are presented in Table 3. At the time of these analyses, a total of 217 participants had both PiB data and coded story recall variables. 172 participants

were classified as PiB(–) and 45 were PiB(+) at the most recent imaging visit. The PiB(+) group was significantly older (mean age = 61, sd = 5) than the PiB(–) group (mean age = 58, sd = 6) ( $p = .005$ ), and had a higher percentage of APOE-e4 carriers (72% in the PiB(+) group vs 31% in the PiB(–) group). The two groups did not differ in any other variables, including education, WRAT-3 reading scores, Mini-Mental State Examination scores, or R-AVLT total scores.

Results from the logistic regression models are presented in Table 4. Using baseline Logical Memory data, proper names from delayed story recall were significantly associated with PiB status, such that participants who recalled more proper names from the stories were less likely to be classified as PiB(+) (odds ratio = .58,  $p = .01$ ). None of the other story recall variables, including total score, were associated with PiB status, and the proper names model was the best fitting model when the AICs were compared; specifically, the proper names model's AIC was 5.63 units lower than the next best fitting model (verbs), which surpasses the criteria of >2 units indicating better fit (Burnham & Anderson, 2002).

Fig. 2 provides a supporting visual illustration of %PiB(+) by low, medium, and high z-scores for delayed recall of proper nouns (left) and delayed total recall (right).

In order to test the robustness of the observed relationship between baseline proper names and most recent PiB status, we ran sensitivity analyses using the logistic regression with most recent proper names recall as predictor (median = visit 4, range = 2–6). We found the same pattern, with proper names significantly predicting PiB status (odds ratio = .64, confidence interval = .45–.92,  $p = .02$ ), while total score was not significantly associated with PiB status. These results were also confirmed when we performed bootstrapping analyses (Table 5). Furthermore, we ran the models with immediate recall of proper names and there was no significant relationship with PiB status.

In exploratory analyses using the baseline delayed Logical Memory data, we examined whether including multiple lexical variables simultaneously explained additional variance in PiB status. After adjusting for proper names, higher number of verbs and total score were associated with increased risk of amyloid (odds ratio 2.04,  $p = .0004$  and 2.40,  $p = .008$  respectively). Follow-up analyses to understand these patterns indicated that low number of proper names ( $z < 0$ ) and higher number of verbs or higher total score ( $z > 0$ ) had higher risk of amyloid positivity than those who were higher on both proper names and verbs or proper names and total (Table 6). We created a 4-level variable via combinations of low versus high number of proper names and verbs based on median split of each, with the reference group being those with higher counts of proper names and higher counts of verbs. These sensitivity analyses sought to examine whether the results observed using continuous proper name and verb data persisted when simplified to categorical levels.

### 3.3. Sensitivity analysis: comparing the relationship with A $\beta$ status and proper names to the PACC

In additional sensitivity analyses, we also examined whether the proper names model would explain more variability in PiB(+) status than the WRAP version (Jonaitis et al., 2019) of a

**Table 3 – Demographics and clinical characteristics of subset of WRAP participants with PiB-PET imaging.**

	PiB(–)	PiB(+)	<i>p</i>
N	172	45	
Age at visit 2 (mean (sd))	58.11 (6.22)	60.96 (4.86)	.005
Age at PET scan	65.3 (6.8)	69 (5.3)	<.001
Age at PET – Age at visit 2	7.2 (3)	8.2 (3)	.04
Female (%)	121 (70.3)	30 (66.7)	.767
Progressed to clinical MCI = 1 (%)	3 (1.7)	2 (4.4)	.605
Race (%)			.813
Native American	3 (1.7)	0 (.0)	
Asian	1 (.6)	0 (.0)	
African-American	5 (2.9)	2 (4.4)	
Non-Hispanic White	162 (94.2)	43 (95.6)	
Other	1 (.6)	0 (.0)	
Family History positive (%)	123 (71.5)	36 (80.0)	.339
WRAT-3 Reading (mean (sd))	108.56 (8.50)	109.25 (7.63)	.626
Total years of education (mean (sd))	16.38 (2.77)	17.22 (2.91)	.073
APOE carrier (%)	56 (32.6)	32 (71.1)	<.001
Baseline MMSE (mean (sd)) <sup>a</sup>	29.37 (.96)	29.40 (.84)	.858
Baseline R-AVLT total score (mean (sd)) <sup>a</sup>	55.6 (7.8)	54.6 (7.9)	.416
Baseline Logical Memory delayed total (mean (sd)) <sup>a</sup>	26.34 (7.13)	26.91 (6.66)	.630
Proper Name delayed total	5.3 (2.0)	4.4 (1.9)	.010
Verb delayed total	6.6 (2.2)	6.2 (1.3)	.214
Numerical Expressions delayed total	2.4 (1.1)	2.6 (1.1)	.520

Items in **bold** are statistically significant at  $p < .05$  using t-tests or chi-square tests (unadjusted means). Abbreviations: WRAT-3 = Wide Range Achievement Test-3 Reading Subtest (Wilkinson, 1993); MMSE = Mini-Mental Status Examination (Folstein et al., 1983); Rey-AVLT = Rey Auditory Verbal Learning Test (Schmidt, 1996); Logical Memory = subtest from the Wechsler Memory Scale-Revised (WMS-R; Wechsler, 1987).

<sup>a</sup> “Baseline” = collected at the first visit at which Logical Memory was administered, median = visit 2.

Table 4 – Binomial logistic regression results of story recall variables at visit 2 predicting most recent beta-amyloid status.

Predictors	Proper names			Verbs			Numbers			Total score		
	Odds ratios	CI	p	Odds ratios	CI	p	Odds ratios	CI	p	Odds ratios	CI	p
(Intercept)	.00	.00–.00	<.001	.00	.00–.00	<.001	.00	.00–.00	<.001	.00	.00–.00	<.001
Age at PiB	1.13	1.05–1.22	.001	1.15	1.07–1.23	<.001	1.14	1.06–1.23	<.001	1.14	1.06–1.23	<.001
PiB Age-Story Recall Age	1.08	.94–1.25	.289	1.02	.89–1.18	.754	1.03	.90–1.19	.653	1.03	.90–1.19	.646
Sex	1.26	.55–2.89	.590	.94	.42–2.10	.875	.95	.42–2.14	.895	.97	.43–2.18	.945
APOE-ε4	9.12	3.88–21.43	<.001	8.73	3.83–19.91	<.001	8.08	3.58–18.22	<.001	7.91	3.53–17.73	<.001
Proper names	.58	.38–.88	.004									
Verbs			1.49	.97–2.29	.070	1.20	.80–1.80	.371	1.15	.75–1.76	.529	
Numerical expressions												
Total score												
Cox & Snell's R <sup>2</sup> /Nagelkerke's R <sup>2</sup>	.206/.322			.194/.303			.184/.288			.182/.285		
AIC	175.26†			180.89			183.44			183.8		
AICA	N/A			+5.63			+8.18			8.54		

Abbreviations: WRAT-3 = Wide Range Achievement Test-3 Reading Subtest (Wilkinson, 1993); CI = 95% confidence interval; AIC = Akaike information criterion. † = model with the lowest AIC, against which change in AIC (AICA) is calculated; AICA = Change in AIC against the model with the lowest AIC, in this case the Proper Names model. Items in **bold** are statistically significant at False Discovery Rate critical value of  $p < .0125$ . Cox & Snell's R<sup>2</sup> and Nagelkerke's R<sup>2</sup> = the proportion of the explained variance in the random effect of the full (conditional model) compared to the null (unconditional) model.

commonly used cognitive composite score, the PACC (Donohue et al., 2014). The PACC-3 composite score from visit 2 was not associated with the most recent PiB status. The AIC for the Proper Name model was 175.26, while the AIC for the PACC-3 model was 189.73, indicating the proper name model was a better fitting model.

### 3.4. Secondary analysis: evaluating longitudinal change in proper names expression and Aβ status

Because of the longitudinal relationships between PiB status and cognitive decline evidenced in the literature, we used linear mixed effects models with either total score or proper names expression as the outcome, and the interaction between PiB status and time as the predictor of interest. Results for these models are presented in Table 7. The interaction between PiB status and age was not significantly associated with proper names recall over time, but this interaction term was significantly associated with total score ( $p = .02$ ). As noted in Table 3, participants who were PiB(+) were already lower on proper names at the baseline story recall visit than those who were PiB(–).

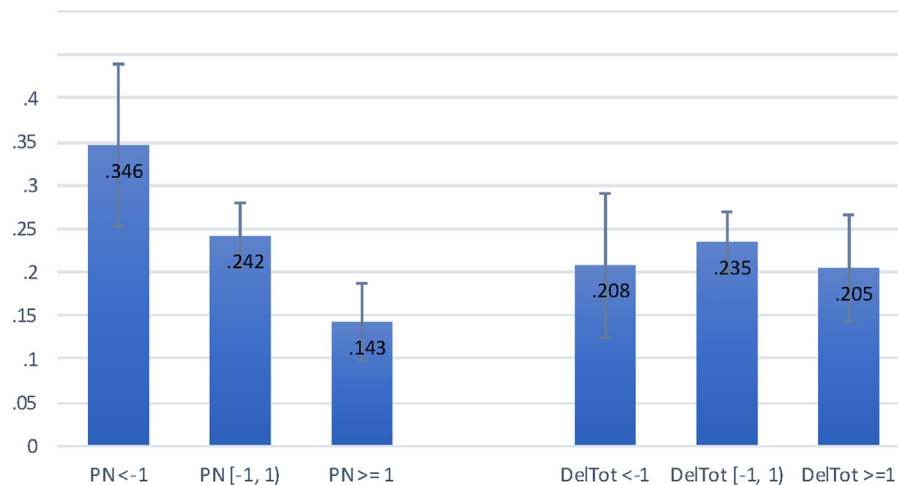
The predicted values of the simple age slopes for participants who were PiB (+) versus those who were PiB(–) for each model are depicted in Fig. 3.

## 4. Discussion

At the time of this study, the ability to detect AD neuropathological processes *in vivo*, particularly both Aβ plaque accumulation and tau neurofibrillary tangles, is a relatively new advancement in the field of AD research. As a result, the temporal patterns of cognitive decline, or the presence of cognitive differences in those individuals with accumulating AD biomarkers, is largely unknown. In this study, we show a novel finding that cognitively unimpaired participants who were amyloid positive using the PiB amyloid PET tracer were less likely to recall proper names from a story recall task at their baseline visit—an average of 7 years earlier—than those without increased PiB accumulation. We did not find such cross-sectional associations with PiB and story recall total score, or with the PACC. We replicated this finding in two follow-up analyses: one using participants' most recent visit (an average of 2 years before PiB scan) and the other using bootstrapped analyses from multiple visits. Moreover, this relationship between proper names and PiB positivity was only significant for delayed recall, not immediate recall, thus indicating that it is *retrieval* of proper names, not encoding, that is associated with PiB status.

Proper name retrieval has been shown to be particularly difficult not only in the face of cognitive impairment, but also in typical aging, and even throughout adulthood (Semenza et al., 1996). Several theories of proper name retrieval converge on the basic idea that proper names do not have the network of semantically related attributes to aid in retrieval as regular nouns do (Cohen & Burke, 1993; Semenza et al., 2000, 2003), and retrieval of proper names may not benefit from cognitive reserve particularly for this reason (Montemurro et al., 2018). It is this limited bank of cognitive resources that





**Fig. 2** – Percent amyloid positive by low, medium, and high z-scores for delayed recall of proper nouns (left) and delayed total recall (right). Proportion PiB + by low ( $z < -1$ ), medium ( $-1 \leq z < 1$ ), or high ( $z \geq 1$ ) delayed recall proper nouns (left set) and delayed total score.

**Table 5** – Follow up: Binomial logistic regression bootstrapped results of story recall variables at visit 2 predicting most recent beta-amyloid status.

Predictors	Proper noun		Verbs		Number		Total score	
	Odds ratios	CI	Odds ratios	CI	Odds ratios	CI	Odds ratios	CI
Age at PiB	1.19	1.16–1.23	1.18	1.15–1.22	1.18	1.15–1.22	1.18	1.15–1.22
WRAT-3 Reading	1.05	.98–1.11	.99	.93–1.06	1.01	.94–1.07	1.01	.94–1.07
Sex (female)	1.69	1.17–2.50	1.20	.78–1.77	1.21	.83–1.78	1.26	.87–1.82
APOE-ε4	11.63	7.86–19.54	9.70	6.46–15.56	9.25	6.12–14.73	9.09	5.98–14.51
Total years of education	1.17	1.04–1.33	1.15	1.03–1.32	1.15	1.03–1.31	1.15	1.02–1.32
Proper names	<b>.52</b>	.43–.64						
Verbs			1.47	1.18–1.99				
Numerical expressions					1.16	.96–1.42		
Total score							1.05	.85–1.42

Abbreviations: WRAT-3 = Wide Range Achievement Test-3 Reading Subtest (Wilkinson, 1993); CI = 95% confidence interval. Items in **bold** indicate statistical significance at  $p < .05$ .

retrieval of proper names has to draw upon that may have particular sensitivity to either focal or diffuse brain pathology, with disrupted phonological representations a contributing factor (Papagno et al., 1998). A study by Delazer et al. (2003) examined proper name anomia in a group of persons with dementia due to probable AD, a group with MCI, and a group of healthy controls. The participants with dementia named significantly fewer famous faces than the two other groups, but did not differ from the other groups regarding Tip of the Tongue (TOT) responses: they were able to provide semantic information about the person or respond to recognition cues. However, the participants with dementia did not respond as readily to phonological cues as did the other groups, providing support for the hypothesis that a proper name retrieval deficit may be due in part to degradation of phonological representations. Juncos-Rabadán et al. (2013) found that TOT responses for proper names significantly predicted persons with amnesic MCI versus healthy controls, and in another study, found that the MCI group differed in phonological access but not semantic access to proper names (Juncos-Rabadán et al., 2011).

Multiple lesion studies and fMRI studies suggest that the anterior temporal lobe and extended neural networks

bilaterally seem more involved in the retrieval of proper names than other lexical categories (Gorno-Tempini et al., 1998; Semenza, 2011; Tsukiura et al., 2002, 2003, 2011). Lesion studies have shown a specific proper name deficit associated with removal of the left uncinate fasciculus (Papagno et al., 2011, 2016) and severe atrophy of the left temporal pole (Papagno et al., 1998). Increased activation of the left temporal pole during proper name tasks has been extensively documented (Semenza, 2011). That the temporal lobe is the typical location of early pathologic tau accumulation (Braak & Braak, 1991) supports a connection between early proper name deficits and early AD pathology. Our findings here regarding amyloid pathology are supported by other studies that include formal tests of proper name retrieval, such as face-name association tasks or recall of recent and remote famous names. For example, Orlovsky et al. (2018) examined the remote and recent famous face-name recall in a group of unimpaired older adults (mean age = 78) who were either Aβ+ or Aβ-. The Aβ+ group recalled fewer recent proper names than the Aβ-group, even when provided with phonological (letter) cues (Orlovsky et al., 2018). Rentz et al. (2011) found that a demanding face-name association task

**Table 6 – Follow-up multinomial logistic regression results comparing risk of amyloid positivity in proper name X verb recall groups.**

Response*	Predictors	Proper name/Verb group		
		Odds ratios	CI	p
2	Age at PiB	.97	.91–1.04	.433
	WRAT-3 Reading	.89	.80–.99	<b>.039</b>
	Sex (female)	.28	.12–.67	<b>.004</b>
	APOE-ε4	.70	.27–1.80	.465
	PiB status (positive)	3.81	1.32–11.02	<b>.013</b>
3	Age at PiB	.98	.92–1.04	.504
	WRAT-3 Reading	.88	.80–.97	<b>.010</b>
	Sex (female)	.68	.28–1.64	.390
	APOE-ε4	2.18	.97–4.90	.058
	PiB status (positive)	.57	.18–1.78	.330
4	Age at PiB	1.01	.95–1.07	.853
	WRAT-3 Reading	.83	.75–.91	<b>&lt;.001</b>
	Sex (female)	.35	.15–.80	<b>.013</b>
	APOE-ε4	1.31	.56–3.03	.536
	PiB status (positive)	1.11	.38–3.23	.841
	Observations	216		
	R <sup>2</sup> Nagelkerke	.219		

4-level variable created via combinations of low versus high proper names and verbs based on median split of each. \*Response 1 (Reference Group): Noun positive & verb positive (n = 81); Response 2: Noun negative, verb positive (n = 38); Response 3: Noun positive, verb negative (n = 49); Response 4: Noun negative, verb negative (n = 49). Items in **bold** are statistically significant at p < .05.

was more sensitive to Aβ pathology than the 6-Trial Selective Reminding Test in a group of cognitively unimpaired older adults with a mean age of 71 (Rentz et al., 2011).

Our study is different from these previous findings about proper names and early amyloid accumulation in several ways. First, the WRAP subset of participants with amyloid imaging and longitudinal neuropsychological testing is younger than most of the groups studied previously, with a mean age of 66 at PiB scan, and a mean age of 58 at the baseline story recall visit. Second, our proper name findings stem from a novel sub-component of a commonly used test of episodic memory: story

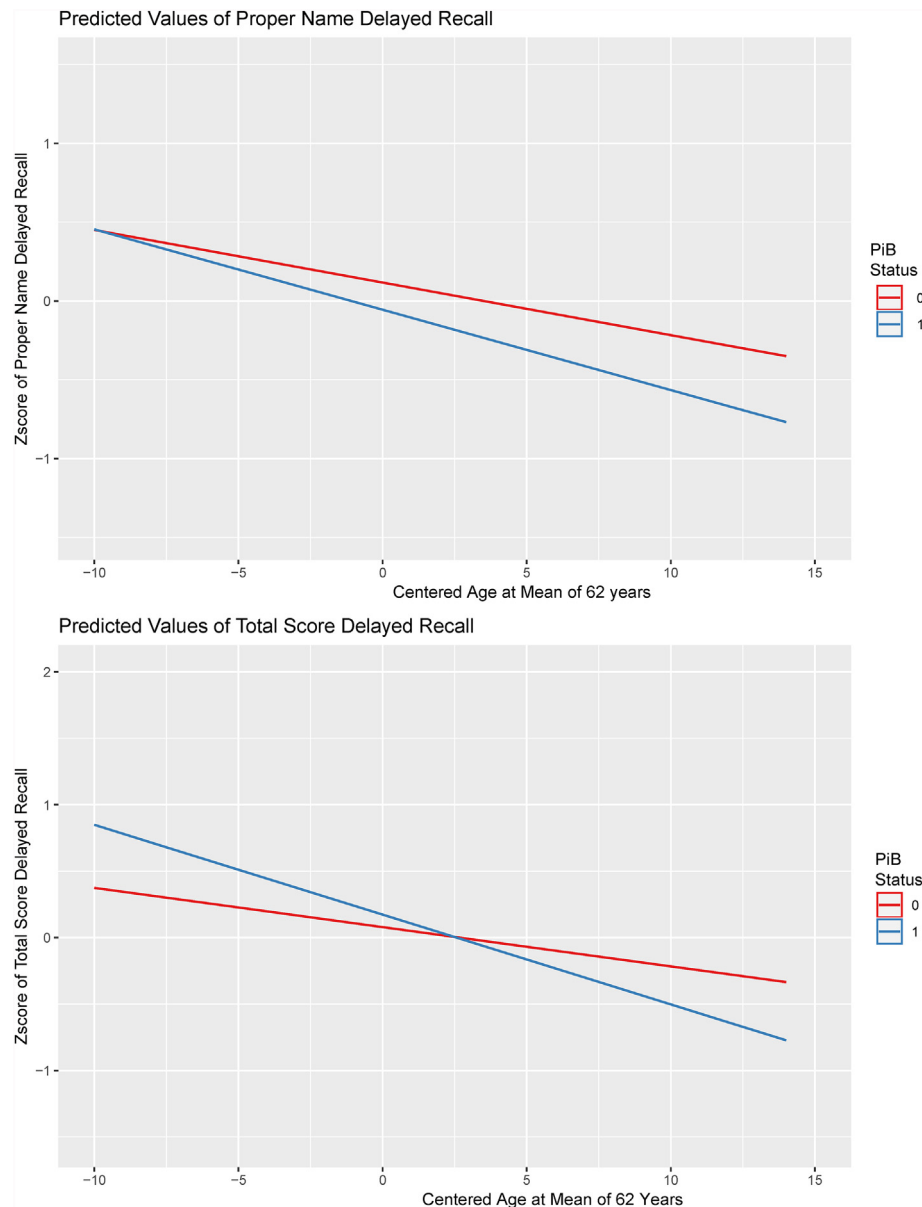
recall. To our knowledge, this is the first study to examine lexical categories from story recall and to show that there may be added sensitivity to Aβ accumulation above and beyond total score. This fact carries special importance in the field today, as there are multiple, ongoing and worldwide longitudinal studies of at-risk cohorts administering story recall, including studies in the Alzheimer's Disease Neuroimaging Initiative (ADNI), studies in the National Alzheimer's Coordinating Center (NACC), and the European Alzheimer's Disease Consortium. A search using the term "Logical Memory" in the Global Alzheimer's Association Interactive Network (GAIN) yields well over 36,000 participants who have completed this test for an AD research study. Because adding new tests to existing longitudinal studies is a concern due to time constraints, understandable reluctance to add to participant burden, and the need for increased staff time and resources, the prospect of using existing data in new ways is invaluable.

Although cross-sectional associations with proper name recall and Aβ status were present across multiple time points, significantly worse longitudinal *change* in proper names recall over time was not associated with Aβ status. There are several possibilities that may explain this disparity. First, delayed recall of proper names at baseline differed between PiB+ and PiB- while delayed total recall did not. Furthermore, the range of proper names from stories A and B from Logical Memory is small, from 0 to 9 (versus total score which ranges from 0 to 50). Thus, it is possible that participants showing subtle cognitive impairment associated with Aβ were already low at baseline logical memory and had less room to decline. That is, participants may recall very few proper names very early on, and perhaps an even earlier starting point – before amyloid accumulation reaches positivity – is required to see such change in proper name retrieval. Recent work by our group and others has shown that by the time a person has reached a PET amyloid positivity threshold, brain amyloid has already been accumulating for many years (Koscik et al., 2020). Future analyses will examine whether estimated duration of PiB positivity at cognitive baseline predicts story recall components or decline in story recall variables.

**Table 7 – Results from linear mixed effects model with amyloid status predicting proper names versus total score over time.**

Predictors	Proper names			Total score		
	Estimates	CI	p	Estimates	CI	p
Age	-.07	-.10 to -.04	<b>&lt;.001</b>	-.20	-.33 to -.08	<b>.001</b>
Sex (female)	.56	.13–.99	<b>.011</b>	1.94	.34–3.53	<b>.017</b>
Practice	.06	-.05 to .17	.263	.67	.30–1.03	<b>&lt;.001</b>
WRAT-3	.12	.08–.17	<b>&lt;.001</b>	.55	.39–.72	<b>&lt;.001</b>
PiB status (positive)	2.43	–1.28 to 6.15	.199	12.89	1.68–24.10	<b>.024</b>
PiB*Age	-.04	-.10 to .01	.138	-.21	-.38 to -.03	<b>.020</b>
<b>Random effects</b>						
σ <sup>2</sup>	2.08			15.82		
τ <sub>00</sub>	1.81	WRAPNo		28.10	WRAPNo	
ICC	.46			.64		
N	232	WRAPNo		232	WRAPNo	
Observations	925			925		
Marginal R <sup>2</sup> /Conditional R <sup>2</sup>	.153/.547			.174/.702		

Linear mixed effects model with centered age as the time variable, and the interaction between amyloid status (PiB) and centered age as the predictor of interest. Items in **bold** are statistically significant at p < .05.



**Fig. 3 – Simple slopes of linear mixed effects models of proper names and total recall as outcomes, and amyloid status (PiB) as the predictor of interest. PiB Status: 0 = amyloid negative, 1 = amyloid positive. Age is centered at a mean of 62.1 for ease of interpretation.**

Secondary analyses from our study showed that when verbs and total score were included in the proper name model, participants who recalled fewer proper names but higher numbers of verbs or total score, were more likely to be classified as PiB positive than those who recalled higher proper names and higher verbs or total score. This finding may represent an overcompensation, such that subtle word retrieval problems may result in circumlocution and result in more talking overall. Because the Logical Memory scoring manual allows alternate wording for certain items in verbs and other lexical categories that make up the total score (i.e., using our hypothetical example, the word “cancelled” would be awarded a point for the words “called off,” or “spoke” would be an allowed alternate response for “talked”), such circumlocution and non-verbatim responses might result in a greater

total score. Conversely, the proper names in each of the stories require verbatim or near-verbatim responses (e.g., “Sue” would be awarded a point for “Suzy”). Future analyses for understanding these dissociations (low proper names but high total scores) may include examining proportions of total score, as well as performing discourse analyses on digitally recorded story recall in order to quantify behaviors such as circumlocutions, order of recall, and errors and error monitoring.

This study also examined whether proper name recall explained additional variability above and beyond total score in terms of which participants progressed to clinical MCI from cognitively unimpaired. Although all story recall variables significantly contributed to this prediction, proper names did not explain additional variability. A limitation of

the clinical progression analysis was its circularity: story recall total score is a factor used in determining cognitive impairment in the WRAP consensus process. In future analyses, we can use actuarial definitions of impairment (Bondi et al., 2014; Jak et al., 2016) based on tests not including Logical Memory and our internally derived norms (Clark, Kosciak, et al., 2016; Kosciak et al., 2019, 2014) to compare how the lexical categories predict progression relative to the traditionally-used total scores.

#### 4.1. Strengths and limitations

This study presents with strengths and limitations which should be acknowledged. First, as noted previously, a significant strength is the use of existing data in a novel way, especially due to the fact that there is a wealth of this data worldwide that can be used to attempt to replicate our findings. Second, the nature of the story recall task itself is a strength; that is, the act of retelling a story that includes proper names, events and other details is a closely related snapshot of an everyday activity in which people engage. This ecological validity can help to uncover the real-world problems that people with subtle cognitive impairments face in communicating, and can inform both pharmacological and nonpharmacological intervention studies. Last but not least, the inclusion of individuals with PET imaging data confirming the presence of amyloid pathology allowed us to determine the relationship between lexical category retrieval and pathology, as opposed to clinical or sub-clinical findings, which are often variable from visit to visit, and may have myriad underlying causes. This study allows the interpretation that lexical retrieval may be related to preclinical AD, although replication in other cohorts and with additional longitudinal study visits and biomarker data is necessary.

A limitation of this study is the sample itself. WRAP is a self-selected, family history cohort made up of participants largely residing in the Upper Midwest. Participants generally are highly educated, and the sample is predominantly non-Hispanic White, thus not representative of the general population. This underscores the need for replication in other cohorts to confirm these associations. As noted above, an additional limitation exists in the ability to use these measures to understand progression to clinical status due to the circular nature of using subcomponents of a test widely used to determine cognitive decline.

#### 4.2. Conclusion

Our data suggest an early association between delayed recall of proper names from a story recall task and A $\beta$  accumulation. The wealth of story recall data that exists in longitudinal at-risk cohort studies will allow for possible replication of these findings, without adding additional participant burden. Future directions for this work include examining these associations with evidence of other AD biomarkers from imaging and cerebrospinal fluid, including markers of phosphorylated tau and neurodegeneration, and in other cohorts. Examining correlations between regional atrophy and proper name retrieval will also contribute to the understanding of these results.

## Acknowledgments

We would like to thank all participants and their families, and the faculty and staff at the University of Wisconsin's Wisconsin Alzheimer's Institute and Wisconsin Alzheimer's Disease Research Center, whose contributions make this work possible. The authors would like to acknowledge the National Institutes of Health, the National Institute on Aging, the National Institute of Child Health and Human Development, and the Alzheimer's Association for financial support of this work under the following grants: NIH R01 AG021155, R01AG027161, R01-AG054059, Wisconsin Alzheimer's Disease Research Center Project Development Grant-AG6394, NIH P50 AG033514, NIH U54 HD090256, AARF-19-614533.

## REFERENCES

- Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., ... Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, 7(3), 270–279.
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... Fox, J. (2015). Package "lme4". *Convergence*, 12(1), 2.
- Beber, B. C., Mandelli, M. L., Santos, M. A. S., Binney, R. J., Miller, B., Chaves, M. L., ... Shapiro, K. A. (2019). A behavioral study of the nature of verb–noun dissociation in the nonfluent variant of primary progressive aphasia. *Aphasiology*, 33(2), 200–215.
- Bethausen, T. J., Kosciak, R. L., Jonaitis, E. M., Allison, S. L., Cody, K. A., Erickson, C. M., ... Johnson, S. C. (2020). Amyloid and tau imaging biomarkers explain cognitive decline from late middle-age. *Brain*, 143(1), 320–335. <https://doi.org/10.1093/brain/awz378>
- Bondi, M. W., Edmonds, E. C., Jak, A. J., Clark, L. R., Delano-Wood, L., McDonald, C. R., ... Salmon, D. P. (2014). Neuropsychological criteria for mild cognitive impairment improves diagnostic precision, biomarker associations, and progression rates. *Journal of Alzheimer's Disease*, 42(1), 275–289.
- Boldogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52(3), 345–370.
- Braak, H., & Braak, E. (1991). Neuropathological staging of Alzheimer-related changes. *Acta Neuropathologica*, 82(4), 239–259.
- Braak, H., Braak, E., & Bohl, J. (1993). Staging of Alzheimer-related cortical destruction. *European Neurology*, 33(6), 403–408.
- Burnham, K. P., & Anderson, D. R. (2002). *A practical information-theoretic approach. Model selection and multimodel inference* (2nd ed.). New York: Springer.
- Cappa, S. F., Sandrini, M., Rossini, P. M., Sosta, K., & Miniussi, C. (2002). The role of the left frontal lobe in action naming: rTMS evidence. *Neurology*, 59(5), 720–723.
- Clark, L. R., Kosciak, R. L., Nicholas, C. R., Okonkwo, O. C., Engelman, C. D., Bratzke, L. C., ... Johnson, S. C. (2016). Mild cognitive impairment in late middle age in the Wisconsin registry for Alzheimer's prevention study: Prevalence and characteristics using robust and standard neuropsychological normative data. *Archives of Clinical Neuropsychology*, 31(7), 675–688.
- Clark, L. R., Racine, A. M., Kosciak, R. L., Okonkwo, O. C., Engelman, C. D., Carlsson, C. M., ... Johnson, S. C. (2016). Beta-



- amyloid and cognitive decline in late middle age: Findings from the Wisconsin Registry for Alzheimer's prevention study. *Alzheimer's & Dementia*, 12(7), 805–814.
- Cohen, G. (1990). Recognition and retrieval of proper names: Age differences in the fan effect. *European Journal of Cognitive Psychology*, 2(3), 193–204.
- Cohen, G., & Burke, D. M. (1993). Memory for proper names: A review. *Memory*, 1(4), 249–263.
- Cotelli, M., Calabria, M., Manenti, R., Rosini, S., Zanetti, O., Cappa, S. F., & Miniussi, C. (2011). Improved language performance in Alzheimer disease following brain stimulation. *Journal of Neurology, Neurosurgery, and Psychiatry*, 82(7), 794–797. <https://doi.org/10.1136/jnnp.2009.197848>
- Curran-Everett, D. (2000). Multiple comparisons: Philosophies and illustrations. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 279(1), R1–R8.
- Davis, C., Heidler-Gary, J., Gottesman, R. F., Crinion, J., Newhart, M., Moghekar, A., ... Hillis, A. E. (2010). Action versus animal naming fluency in subcortical dementia, frontal dementias, and Alzheimer's disease. *Neurocase*, 16(3), 259–266.
- Delazer, M., Semenza, C., Reiner, M., Hofer, R., & Benke, T. (2003). Anomia for people names in DAT—evidence for semantic and post-semantic impairments. *Neuropsychologia*, 41(12), 1593–1598.
- Donohue, M. C., Sperling, R. A., Petersen, R., Sun, C.-K., Weiner, M. W., & Aisen, P. S. (2017). Association between elevated brain amyloid and subsequent cognitive decline among cognitively normal persons. *Jama*, 317(22), 2305–2316.
- Donohue, M. C., Sperling, R. A., Salmon, D. P., Rentz, D. M., Raman, R., Thomas, R. G., ... Aisen, P. S. (2014). The preclinical Alzheimer cognitive composite: Measuring amyloid-related decline. *JAMA Neurology*, 71(8), 961–970.
- Druks, J., Masterson, J., Kopelman, M., Clare, L., Rose, A., & Rai, G. (2006). Is action naming better preserved (than object naming) in Alzheimer's disease and why should we ask? *Brain and Language*, 98(3), 332–340.
- Farrell, M. E., Kennedy, K. M., Rodrigue, K. M., Wig, G., Bischof, G. N., Rieck, J. R., ... Park, D. C. (2017). Association of longitudinal cognitive decline with amyloid burden in middle-aged and older adults: Evidence for a dose-response relationship. *JAMA Neurology*, 74(7), 830–838.
- Folstein, M. F., Robins, L. N., & Helzer, J. E. (1983). The mini-mental state examination. *Arch Gen Psychiatry*, 40(7), 812.
- Gainotti, G. (2007). Different patterns of famous people recognition disorders in patients with right and left anterior temporal lesions: A systematic review. *Neuropsychologia*, 45(8), 1591–1607.
- Garrard, P., Patterson, K., Watson, P. C., & Hodges, J. R. (1998). Category specific semantic loss in dementia of Alzheimer's type. Functional-anatomical correlations from cross-sectional analyses. *Brain*, 121(Pt 4), 633–646. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9577390>.
- Gorno-Tempini, M. L., Price, C. J., Josephs, O., Vandenberghe, R., Cappa, S. F., Kapur, N., ... Tempini, M. L. (1998). The neural systems sustaining face and proper-name processing. *Brain: a Journal of Neurology*, 121(11), 2103–2118.
- Hodges, J. R., & Patterson, K. (1995). Is semantic memory consistently impaired early in the course of Alzheimer's disease? Neuroanatomical and diagnostic implications. *Neuropsychologia*, 33(4), 441–459.
- Jack, C. R., Jr., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., ... Silverberg, N. (2018). NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia*, 14(4), 535–562.
- Jak, A. J., Preis, S. B., Beiser, A. S., Seshadri, S., Wolf, P. A., Bondi, M. W., & Au, R. (2016). Neuropsychological criteria for mild cognitive impairment and dementia risk in the Framingham Heart Study. *Journal of the International Neuropsychological Society*, 22(9), 937–943.
- Jansen, W. J., Ossenkoppele, R., Tijms, B. M., Fagan, A. M., Hansson, O., Klunk, W. E., ... Visser, P. J. (2018). Association of cerebral amyloid- $\beta$  aggregation with cognitive functioning in persons without dementia. *JAMA Psychiatry*, 75(1), 84–95.
- Johnson, S. C., Christian, B. T., Okonkwo, O. C., Oh, J. M., Harding, S., Xu, G., ... Sager, M. A. (2014). Amyloid burden and neural function in people at risk for Alzheimer's Disease. *Neurobiology of Aging*, 35(3), 576–584. <https://doi.org/10.1016/j.neurobiolaging.2013.09.028>
- Johnson, S. C., Kosciak, R. L., Jonaitis, E. M., Clark, L. R., Mueller, K. D., Berman, S. E., ... Sager, M. A. (2018). The Wisconsin Registry for Alzheimer's Prevention: A review of findings and current directions. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 10, 130–142.
- Jonaitis, E. M., Kosciak, R. L., Clark, L. R., Ma, Y., Betthauser, T. J., Berman, S. E., ... Johnson, S. C. (2019). Measuring longitudinal cognition: Individual tests versus composites. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 11, 74–84.
- Juncos-Rabadán, O., Facal, D., Lojo-Seoane, C., & Pereiro, A. X. (2013). Does tip-of-the-tongue for proper names discriminate amnesic mild cognitive impairment? *International Psychogeriatrics*, 25(4), 627–634.
- Juncos-Rabadán, O., Rodríguez, N., Facal, D., Cuba, J., & Pereiro, A. X. (2011). Tip-of-the-tongue for proper names in mild cognitive impairment. Semantic or post-semantic impairments? *Journal of Neurolinguistics*, 24(6), 636–651.
- Kosciak, R. L., Berman, S. E., Clark, L. R., Mueller, K. D., Okonkwo, O. C., Gleason, C. E., ... Johnson, S. C. (2016). Intraindividual cognitive variability in middle age predicts cognitive impairment 8–10 years later: Results from the Wisconsin Registry for Alzheimer's Prevention. *Journal of the International Neuropsychological Society*, 22(10), 1016–1025.
- Kosciak, R. L., Betthauser, T. J., Jonaitis, E. M., Allison, S. L., Clark, L. R., Hermann, B. P., ... Johnson, S. C. (2020). Amyloid duration is associated with preclinical cognitive decline and tau PET. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 12(1), 63–72.
- Kosciak, R. L., Jonaitis, E. M., Clark, L. R., Mueller, K. D., Allison, S. L., Gleason, C. E., ... Johnson, S. C. (2019). Longitudinal standards for mid-life cognitive performance: Identifying abnormal within-person changes in the Wisconsin Registry for Alzheimer's Prevention. *Journal of the International Neuropsychological Society*, 25(1), 1–14.
- Kosciak, R. L., La Rue, A., Jonaitis, E. M., Okonkwo, O. C., Johnson, S. C., Bendlin, B. B., ... Sager, M. A. (2014). Emergence of mild cognitive impairment in late middle-aged adults in the Wisconsin registry for Alzheimer's prevention. *Dementia and Geriatric Cognitive Disorders*, 38(1–2), 16–30.
- Logan, J., Fowler, J. S., Volkow, N. D., Wang, G.-J., Ding, Y.-S., & Alexoff, D. L. (1996). Distribution volume ratios without blood sampling from graphical analysis of PET data. *Journal of Cerebral Blood Flow & Metabolism*, 16(5), 834–840.
- Lopresti, B. J., Klunk, W. E., Mathis, C. A., Hoge, J. A., Ziolkowski, S. K., Lu, X., ... Price, J. C. (2005). Simplified quantification of Pittsburgh compound B amyloid imaging PET studies: A comparative analysis. *Journal of Nuclear Medicine*, 46(12), 1959–1972.
- MacWhinney, B. (2014). *The CHILDES project: Tools for analyzing talk, Volume II: The database*. Psychology Press.
- Manly, J. J., Jacobs, D. M., Touradj, P., Small, S. A., & Stern, Y. (2002). Reading level attenuates differences in neuropsychological test performance between African American and White elders. *Journal of the International Neuropsychological Society: JINS*, 8(3), 341–348.
- Montemurro, S., Mondini, S., Nucci, M., & Semenza, C. (2018). Proper name retrieval in cognitive decline: The role of cognitive reserve. *The Mental Lexicon*, 13(2), 215–229.
- Mormino, E. C., & Papp, K. V. (2018). Amyloid accumulation and cognitive decline in clinically normal older individuals:

- Implications for aging and early Alzheimer's disease. *Journal of Alzheimer's Disease*, 64(s1), S633–S646.
- Mueller, K. D., Kosciak, R. L., Hermann, B. P., Johnson, S. C., & Turkstra, L. S. (2018). Declines in connected language are associated with very early mild cognitive impairment: Results from the Wisconsin registry for Alzheimer's prevention. *Frontiers in Aging Neuroscience*, 9, 437.
- Orlovsky, I., Huijbers, W., Hanseeuw, B. J., Mormino, E. C., Hedden, T., Buckley, R. F., ... Papp, K. V. (2018). The relationship between recall of recently versus remotely encoded famous faces and amyloidosis in clinically normal older adults. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 10, 121–129.
- Papagno, C., & Capitani, E. (1998). Proper name anomia: A case with sparing of the first-letter knowledge. *Neuropsychologia*, 36(7), 669–679.
- Papagno, C., Casarotti, A., Comi, A., Pisoni, A., Lucchelli, F., Bizzi, A., ... Bello, L. (2016). Long-term proper name anomia after removal of the uncinate fasciculus. *Brain Structure & Function*, 221(1), 687–694.
- Papagno, C., Miracapillo, C., Casarotti, A., Romero Lauro, L. J., Castellano, A., Falini, A., ... Bello, L. (2011). What is the role of the uncinate fasciculus? Surgical removal and proper name retrieval. *Brain*, 134(2), 405–414.
- Papp, K. V., Mormino, E. C., Amariglio, R. E., Munro, C., Dagley, A., Schultz, A. P., ... Rentz, D. M. (2016). Biomarker validation of a decline in semantic processing in preclinical Alzheimer's disease. *Neuropsychology*, 30(5), 624.
- Rabin, J. S., Schultz, A. P., Hedden, T., Viswanathan, A., Marshall, G. A., Kilpatrick, E., ... Chhatwal, J. P. (2018). Interactive associations of vascular risk and  $\beta$ -amyloid burden with cognitive decline in clinically normal elderly individuals: Findings from the Harvard Aging Brain Study. *JAMA Neurology*, 75(9), 1124–1131.
- Racine, A. M., Clark, L. R., Berman, S. E., Kosciak, R. L., Mueller, K. D., Norton, D., ... Johnson, S. C. (2016). Associations between performance on an abbreviated CogState battery, other measures of cognitive function, and biomarkers in people at risk for Alzheimer's disease. *Journal of Alzheimer's Disease*, 54(4), 1395–1408.
- Rascovsky, K., Salmon, D. P., Hansen, L. A., Thal, L. J., & Galasko, D. (2007). Disparate letter and semantic category fluency deficits in autopsy-confirmed frontotemporal dementia and Alzheimer's disease. *Neuropsychology*, 21(1), 20–30. <https://doi.org/10.1037/0894-4105.21.1.20>
- Rentz, D. M., Amariglio, R. E., Becker, J. A., Frey, M., Olson, L. E., Frishe, K., ... Sperling, R. A. (2011). Face-name associative memory performance is related to amyloid burden in normal elderly. *Neuropsychologia*, 49(9), 2776–2783.
- Schmidt, M. (1996). In *Rey Auditory Verbal Learning Test: A handbook*. Western Psychological Services.
- Semenza, C. (2011). Naming with proper names: The left temporal pole theory. *Behavioural Neurology*, 24(4), 277–284.
- Semenza, C., Borgo, F., Mondini, S., Pasini, M., & Sgaramella, T. (2000). Proper names in the early stages of Alzheimer's disease. *Brain and Cognition*, 43(1–3), 384–387.
- Semenza, C., Mondini, S., Borgo, F., Pasini, M., & Sgaramella, M. (2003). Proper names in patients with early Alzheimer's disease. *Neurocase*, 9(1), 63–69.
- Semenza, C., Nichelli, F., & Gamboz, N. (1996). The primacy effect in free recall of lists of common and proper names: A study on young, elderly, and Alzheimer's disease subjects. In *Academy of Aphasia 34th Annual meeting* (Vol. 55, pp. 45–47).
- Shapiro, K. A., Moo, L. R., & Caramazza, A. (2006). Cortical signatures of noun and verb production. *Proceedings of the National Academy of Sciences*, 103(5), 1644–1649.
- Team, R. C.. (2019). *R: A language and environment for statistical computing*. Retrieved from <https://www.R-project.org>.
- Tsukiura, T., Fujii, T., Fukatsu, R., Otsuki, T., Okuda, J., Umetsu, A., ... Yamadori, A. (2002). Neural basis of the retrieval of people's names: Evidence from brain-damaged patients and fMRI. *Journal of Cognitive Neuroscience*, 14(6), 922–937.
- Tsukiura, T., Namiki, M., Fujii, T., & Iijima, T. (2003). Time-dependent neural activations related to recognition of people's names in emotional and neutral face-name associative learning: An fMRI study. *NeuroImage*, 20(2), 784–794.
- Tsukiura, T., Sekiguchi, A., Yomogida, Y., Nakagawa, S., Shigemune, Y., Kambara, T., ... Kawashima, R. (2011). Effects of aging on hippocampal and anterior temporal activations during successful retrieval of memory for face–name associations. *Journal of Cognitive Neuroscience*, 23(1), 200–213.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., ... Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, 15(1), 273–289.
- Venneri, A., Mitolo, M., & De Marco, M. (2016). Paradigm shift: Semantic memory decline as a biomarker of preclinical Alzheimer's disease. In *Future medicine*.
- Wechsler, D. (1987). *Manual for the Wechsler Memory Scale-Revised*. San Antonio, TX: Psychological Corporation.
- Wilkinson, G. S. (1993). In D. E. Wilmington (Ed.), *The Wide Range Achievement test: Manual* (3rd ed.). Wide Range, Inc.
- Williams, B. W., Mack, W., & Henderson, V. W. (1989). Boston naming test in Alzheimer's disease. *Neuropsychologia*, 27(8), 1073–1079.