

# Women have Farther to Fall: Gender Differences Between Normal Elderly and Alzheimer's Disease in Verbal Memory Engender Better Detection of Alzheimer's Disease in Women

Robert M. Chapman,<sup>1</sup> Mark Mapstone,<sup>2</sup> Margaret N. Gardner,<sup>3</sup> Tiffany C. Sandoval,<sup>3</sup> John W. McCrary,<sup>1</sup> Maria D. Guillily,<sup>3</sup> Lindsey A. Reilly,<sup>3</sup> AND Elizabeth DeGrush<sup>3</sup>

<sup>1</sup>Department of Brain and Cognitive Sciences and Center for Visual Science, University of Rochester, Rochester, New York

<sup>2</sup>Department of Neurology, University of Rochester Medical Center, Rochester, New York

<sup>3</sup>Department of Brain and Cognitive Sciences, University of Rochester, Rochester, New York

(RECEIVED September 19, 2010; FINAL REVISION March 3, 2011; ACCEPTED March 3, 2011)

## Abstract

We analyzed verbal episodic memory learning and recall using the Logical Memory (LM) subtest of the Wechsler Memory Scale-III to determine how gender differences in AD compare to those seen in normal elderly and whether or not these differences impact assessment of AD. We administered the LM to both an AD and a Control group, each comprised of 21 men and 21 women, and found a large drop in performance from normal elders to AD. Of interest was a gender interaction whereby the women's scores dropped 1.6 times more than the men's did. Control women on average outperformed Control men on every aspect of the test, including immediate recall, delayed recall, and learning. Conversely, AD women tended to perform worse than AD men. Additionally, the LM achieved perfect diagnostic accuracy in discriminant analysis of AD *versus* Control women, a statistically significantly higher result than for men. The results indicate the LM is a more powerful and reliable tool in detecting AD in women than in men. (*JINS*, 2011, 17, 654–662)

**Keywords:** Gender differences, Alzheimer's disease (AD), Normal elderly, Logical memory, Wechsler Memory Scale (WMS), Verbal episodic memory, Learning, Sex, Diagnosis

## INTRODUCTION

It remains a controversial issue whether there are gender differences in the cognition and behavior among individuals afflicted with Alzheimer's disease (AD) (Woodard et al., 2009), a degenerative neurological disorder with primary cognitive deficits in the domain of memory (McKhann et al., 1984). Some evidence suggests that men and women may be affected differently by AD (Azad, Al Bugami, & Loy-English, 2007; Beinhoff, Tumani, Bretschneider, Bittner, & Riepe, 2008; Heun & Kockler, 2002; Millet et al., 2009). Normal, cognitively healthy women tend to perform better than men on semantic tasks, including oral verbal fluency (Benton & Hamsher, 1978; Fuld, 1977; Kolb & Wishaw, 1985), written word fluency (Yeudall, Fromm, Reddon, & Stefanyk, 1986), naming (Denckla & Rudel, 1974), and verbal learning (Bleecker, Bolla-Wilson, Agnew, & Meyers, 1988; Brouwers,

Cox, Martin, Chase, & Fedio, 1984; Kramer, Delis, & Daniel, 1988). This advantage often seen in normal women over normal men on semantic tasks may reverse in AD (McPherson, Back, Buckwalter, & Cummings, 1999; Ripich, Pettrill, Whitehouse, & Ziol, 1995). The implications of gender effects on the assessment of memory impairments is not well known, but how men and women might disparately perform on cognitive tests could impact the diagnoses and conclusions reached by clinicians and researchers.

Of particular interest in the diagnosis of memory impairment are increased deficits seen in AD women on tasks of semantic memory and learning (McPherson et al., 1999), as difficulties with memory are the hallmark of the degeneration seen in AD. One task of semantic episodic memory commonly administered in the assessment of AD is the Logical Memory (LM) subtest of the Wechsler Memory Scale (Wechsler, 1945, 1987, 1997). The LM has been found to be a powerful tool in differentiating between AD and normal elderly (Chapman et al., 2010; Storandt & Hill, 1989) and in predicting conversion to AD in individuals with mild cognitive impairment (MCI) (Chapman et al., 2011). One previous gender study

Correspondence and reprint requests to: Robert M. Chapman, Center for Visual Science at the University of Rochester, 775 Library Road, Rochester, New York 14627-0270. E-mail: rmc@cvs.rochester.edu

using the LM focused on AD rather than the comparison between AD and normal aging (McPherson et al., 1999). Nonetheless, McPherson and colleagues using the LM showed gender differences in their sample of AD subjects.

Here, we aim to analyze the LM in depth, comparing the performance of men and women within and between groups of early-stage AD and normal elderly, to determine if the test could be a more useful tool in detecting early AD in women than in men. Because the LM is so commonly used by clinicians as part of a diagnostic neuropsychological battery, the impact of gender on performance in both normal elderly and patients with AD is important to recognize and take into account when assessing memory impairment. We examine each LM story's scores to determine if normal elderly women perform better on measures of verbal memory and recall than normal elderly men and if this trend disappears or reverses in AD. Using discriminant analysis, we provide evidence that the LM may prove to be a more powerful and reliable tool for detecting cognitive decline in women than in men.

## METHOD

### Study Sample

A total of 84 subjects participated in this study (Table 1). Both the AD group and the Control group had 42 subjects comprised of 21 men and 21 women. The AD subjects were from the Alzheimer's Disease Center at Strong Memorial Hospital and the Memory Disorders Clinic at Monroe Community Hospital. Each AD subject met standard criteria for AD (NINCDS-ADRDA criteria) (McKhann et al., 1984) and DSM-4TR criteria for Dementia of the Alzheimer's Type (American Psychiatric Association, 2000) and were early in the course of the disease (Table 1). The clinical diagnosis of AD was performed by memory-disorders physicians and was based on the patient history, cognitive testing, relevant laboratory findings, and imaging studies routinely performed as part of the clinical assessment of dementia (Morris et al., 1989; Petersen et al., 2001). The neuropsychological test data collected in this study did not contribute to the clinical diagnoses of the subjects.

Control subjects were elderly volunteers from the community, many of whom were diagnosed to have normal cognitive functioning by the same memory-disorders physicians. Exclusion criteria for all subjects included Parkinson's

disease, HIV/AIDS, clinical (or imaging) evidence of stroke, reversible dementias, and treatment with benzodiazepines, antipsychotic, or antiepileptic medications. The subjects spoke English as a primary language. Informed consent approved by the Research Subjects Review Board at the University of Rochester was obtained.

As inclusion criteria, each participant had a Mini Mental State Examination (MMSE) (Folstein, Folstein, & McHugh, 1975) score of 22 or greater (out of 30, where a higher score indicates better performance). The mean MMSE score was appropriate for each group's diagnosis (mean (*SD*) score for the AD group = 24.3 (3.6) and for the Control group = 28.3 (1.7)). There was no significant difference in MMSE score between AD men and AD women ( $t(40) = -.08; p = .93$ ), indicating that the severity of global cognitive impairment was approximately equal in both genders. There was a significant gender effect in the Control group ( $t(40) = 2.18; p = .04$ ) such that women outperformed men. On the North American National Adult Reading Test (AMNART) (Grober & Sliwinski, 1991), there was a significant group effect ( $F(1,80) = 6.20; p < .05$ ; Control > AD). However, there was no significant gender effect ( $F(1,80) = 1.93; p = .17$ ) or significant group by gender interaction ( $F(1,80) = 0.79; p = .38$ ). There was no significant main group effect for age ( $F(1,80) = 0.43; p = .52$ ) or education ( $F(1,80) = 1.98; p = .16$ ). Similarly, there was no significant main gender effect for either age ( $F(1,80) = 0.43; p = .52$ ) or education ( $F(1,80) = 3.27; p = .07$ ). Finally, there was no significant group by gender interaction for either age ( $F(1,80) = 0.02; p = .88$ ) or education ( $F(1,80) = 0.00; p = .97$ ). The number of right-handed individuals did not differ significantly among gender subgroups (Fisher's exact test,  $\chi^2(1, N = 84) = 0.01; p = .91$ ). At the time of testing, 39 of the 42 AD individuals were taking cholinesterase inhibitors and/or memantine (20 women and 19 men). The proportion of individuals taking these medications did not significantly differ between men and women within the AD group (Fisher's exact test,  $\chi^2(1, N = 42) = 0.36, p = .55$ ). Two Control women and one AD woman were also on hormone replacement therapy at the time of testing.

### Evaluation with the WMS-III Logical Memory Subtest

All subjects were administered the WMS-III Logical Memory (LM) subtest, which included both the immediate and delayed

**Table 1.** Demographic information for the Alzheimer's disease (AD) and control groups by gender

Group	Gender	<i>N</i>	Age	Education	MMSE <sup>a</sup>	Hand <sup>b</sup>	AMNART <sup>c</sup>
Control	Women	21	74.2 (6.0)	14.9 (2.3)	28.9 (1.4)	85%	41.1 (5.2)
Control	Men	21	76.6 (6.0)	15.9 (2.8)	27.8 (1.9)	85%	36.8 (10.4)
AD	Women	21	75.4 (8.4)	14.0 (2.6)	24.3 (3.4)	90%	34.7 (6.9)
AD	Men	21	77.3 (4.5)	15.1 (3.0)	24.4 (4.0)	85%	33.7 (11.0)

*Note.* Values appear as mean (*SD*). The age and education are number of years.

<sup>a</sup>The maximum score on the Mini Mental State Examination (MMSE) (Folstein, et al., 1975) is 30.

<sup>b</sup>Percent of individuals who were right-handed.

<sup>c</sup>North American National Adult Reading Test (Grober & Sliwinski, 1991).

recall portions (Wechsler, 1997). This test was administered as part of a neuropsychological battery that targeted areas of cognition affected by AD. In the LM, the examiner reads a short story and afterward asks the participant to recall as many of the details of the story as possible. There are two stories (Story A and Story B), each comprised of only a few sentences. The first, Story A, is read a single time. The second story, Story B, is read twice. After each reading of each story, the subject is asked to recall as many details as he or she can remember. The examiner marks each of the individual details the subject can spontaneously produce (as per the scoring criteria included with the test). Then, after a delay (approximately 20 min filled with other, unrelated cognitive tests), the participant is asked again to recall each story from memory but without an additional reading or prompting. The following neuropsychological tests were given during the delay as part of our battery: a clock-drawing test (Tuokko, Hadjistavropoulos, Miller, & Beattie, 1992), the North American National Adult Reading Test (Grober & Sliwinski, 1991), the Stroop test (Golden, 1978), the immediate recall portion of the Brief Visuospatial Memory Test-Revised (Benedict & Groninger, 1995), and the Controlled Oral Word Association Test and Category Fluency (Benton & Hamsher, 1978). We examined gender effects in AD on these tests and others in our battery that target other cognitive domains; a deep exploration of these results is not relevant to this article. However, no significant effects were obtained save on the Money Road-Map test (Money, 1976), which is an effect discussed by others (Lezak, Howieson, & Loring, 2004).

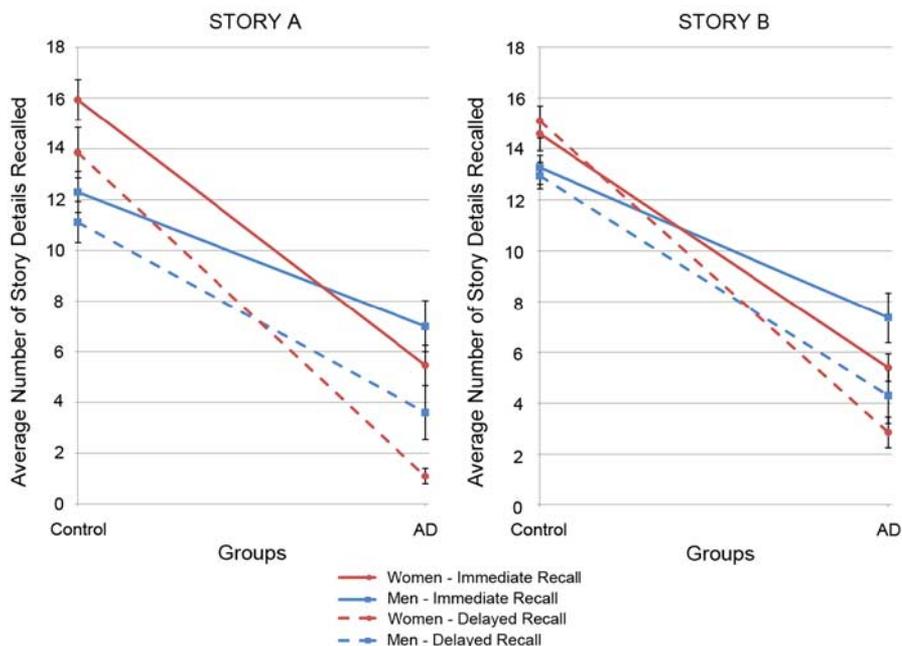
The LM was administered by trained laboratory staff. Each test was graded twice to ensure accuracy. A gist scoring method

was used where the story detail had to be reported verbatim or the basic idea of the detail had to be reported to receive a point (Abikoff et al., 1987). This is a slightly easier and more lenient assessment approach, and it was taken to help prevent a bottoming effect among the AD subjects. A perfect score requires remembrance of all 25 items in each story. The raw scores consisted of a sum of all the details remembered, resulting in a score from 0 to 25 for each story and recall. For the purposes of graphically comparing Story B to Story A in Figure 1, the first and second immediate recall scores of Story B were averaged together. These immediate recall scores of Story B remained separate for other analyses (Table 2; Figure 2).

### Statistical Analysis

Logical Memory measures of immediate and delayed recall were the dependent variables that were studied in relation to three independent variables with Gender (women, men) and diagnostic Group (AD, Control) as between-subjects factors, and Story (A, B) to be remembered as a within-subject factor. Three-way factorial analyses of variance (ANOVAs) of Gender, diagnostic Group, and Story provided a method of studying the effects of each of these factors as well as their interactions, the Gender  $\times$  Group interaction being a particularly important one in this study.

To examine the ability of the LM to detect early AD in women and men, linear discriminant analysis was applied entering all five measures (immediate and delayed recalls for Story A, and the two immediate and one delayed recalls for Story B). Separate discriminant analysis was computed for men and for women to optimize the outcomes for



**Fig. 1.** The average number of story details recalled by men and women of the Control and AD groups. The immediate (solid lines) and delayed recall (dashed lines) scores are presented for both Story A and Story B (the immediate recall for Story B in this figure is the average of the first and second recall scores). Error bars represent SEM. The gender by group interaction is significant ( $p < .001$ ).

**Table 2.** ANOVA results for logical memory immediate and delayed recall

	<i>df</i>	<i>F</i>	<i>p</i>	Partial $\eta$
<b>Story A</b>				
<i>Between-subjects effects</i>				
Group (AD, Control)	1, 80	147.34	<.0001	.79
Gender (Women, Men)	1, 80	0.64	.42	NS
Group $\times$ Gender	1, 80	12.48	.0007	.24
<i>Within-subject effects</i>				
Recall (Immediate, Delayed)	1, 80	96.18	<.0001	.74
Recall $\times$ Group	1, 80	15.79	.0002	.41
Recall $\times$ Gender	1, 80	2.86	.10	NS
Recall $\times$ Group $\times$ Gender	1, 80	0.01	.93	NS
<b>Story B</b>				
<i>Between-subjects effects</i>				
Group (AD, Control)	1, 80	129.30	<.0001	.79
Gender (Women, Men)	1, 80	0.01	.91	NS
Group $\times$ Gender	1, 80	5.10	.03	.24
<i>Within-subject effects</i>				
Recall (Immediate 1, Immediate 2, Delayed)	2, 80	68.15	<.0001	.68
Recall $\times$ Group	2, 80	20.61	<.0001	.45
Recall $\times$ Gender	2, 80	2.34	.10	NS
Recall $\times$ Group $\times$ Gender	2, 80	0.79	.46	NS

NS = not significant.

each gender. The subsequent classification accuracies were analyzed with one-tailed Fisher's exact tests. An alpha level of .05 was chosen for statistical significance.

SAS 9.1.3 (the GLM, DISCRIM, and NPAR1WAY procedures) was used to compute statistical analyses in this study (SAS Institute Inc., 2002).

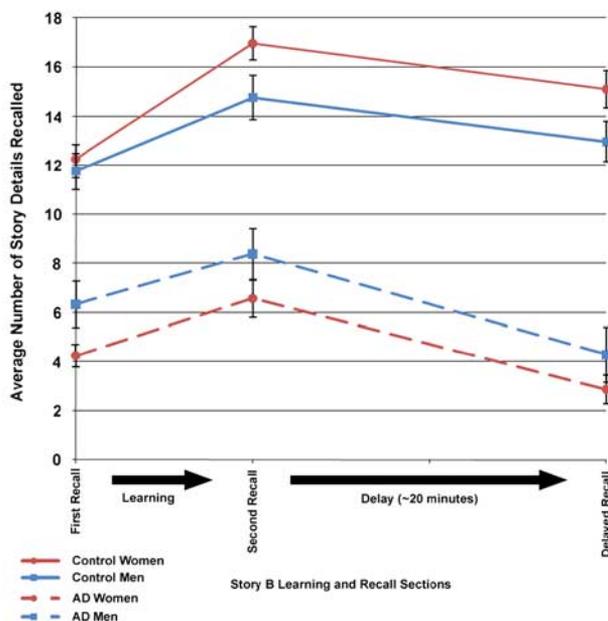
## RESULTS

### Gender and Group Effects on Verbal Episodic Memory Recall

#### Story A

An examination of the mean recall scores for each story (Figure 1; Table 2) revealed that the Control group consistently and significantly outperformed the AD group in remembering the details of Story A during immediate recall [overall Control mean score (*SD*) = 14.1 (3.8) versus overall AD score = 6.2 (3.9)] and delayed recall [overall Control mean score (*SD*) = 12.5 (4.0) versus overall AD score = 2.4 (3.7)]. Also, Control women generally recalled more details than Control men. However, the opposite was true for the AD group as on average the AD men remembered more of Story A than the AD women did.

A mixed factorial ANOVA with Group (AD vs. Control) and Gender (women vs. men) as between-subjects variables and Recall Trial (Immediate vs. Delayed) as the within-subjects variable was performed on Recall performance for Story A. The group effect (between subjects) was significant ( $F(1,80) = 147.34$ ;  $p < .0001$ ) and in the expected direction (Controls performing much better than ADs). In addition, there was a significant interaction of gender with group ( $F(1,80) = 12.48$ ;  $p = .0007$ ) represented by the gender lines crossing in Figure 1. The main between subjects effect of gender was not significant ( $F(1,80) = 0.64$ ;  $p = .42$ ), because the higher performance of women over men for the Controls was nearly counterbalanced by the relative higher performance of men over women for the ADs.



**Fig. 2.** Story B learning curves for women and men of the Control (solid line) and AD (dashed line) groups. The first and second recall scores represent learning scores (where the story was presented immediately prior to testing). Error bars represent SEM.

There was also a significant main effect of recall (Immediate vs. Delayed) ( $F(1,80) = 96.18$ ;  $p < .0001$ ) and a significant interaction between recall and group ( $F(1,80) = 15.79$ ;  $p = .0002$ ). These recall effects were expected, as the subjects tended to perform worse on the delayed recall than the immediate recall, and this drop in score is greatly amplified in AD. However, there was no significant interaction between recall and gender, and the three-way interaction of recall, group, and gender was not significant.

### Story B

The results for Story B were quite similar to those for Story A. The mean scores for the immediate and delayed recalls (Figure 1) of Story B reveal that the differences among the groups and genders seen in the immediate recall scores persisted. The Control Group performed significantly better than the AD group on the immediate recalls [averaged overall Control mean score ( $SD$ ) = 13.9 (3.4) versus averaged overall AD score = 6.4 (3.6)] and for the delayed recall [overall Control mean score ( $SD$ ) = 14.0 (3.8) versus overall AD score = 3.6 (4.3)], as expected.

Again, Control women on average scored higher than Control men; whereas, conversely, AD men performed better than AD women (Story B crossing lines in Figure 1). Another mixed factorial ANOVA with diagnostic Group (AD vs. Control) and Gender (women vs. men) as between-subjects variables and Recall Trial as the within-subjects variable (Immediate 1, Immediate 2, and Delayed) was performed on Recall performance for Story B (Table 2). A main group (AD vs. Control) effect was found ( $F(1,80) = 129.30$ ;  $p < .0001$ ), as well as a group by gender interaction ( $F(1,80) = 5.10$ ;  $p < .05$ ). The main gender effect was, as for Story A, not significant. There was also a main recall (Immediate 1, Immediate 2, Delayed) effect ( $F(2,80) = 68.15$ ;  $p < .0001$ ), and the interaction of recall with group was significant as well ( $F(2,80) = 20.61$ ;  $p < .0001$ ). As for Story A, neither the interaction of recall with gender nor the three-way interaction reached significance.

Overall, the three Story B recall measures (moving along the x-axis in Figure 2) had a characteristic pattern of increasing from the first to the second immediate recall and then decreasing on the delayed recall. However, this recall pattern was somewhat different for Controls than for AD, for example, AD showing more of a drop during the delay period. Moreover, while Control women learned at a higher rate than men [learning slope between first and second immediate recall: ( $t(40) = +2.21$ ;  $p = .03$ )], they forgot details at roughly the same rate in the time between learning and delayed recall ( $t(40) = -0.07$ ;  $p = .94$ ). The rate of forgetting was also roughly equivalent between AD women and AD men ( $t(40) = +0.38$ ;  $p = .71$ ), as was their difference in rate of learning ( $t(40) = +0.35$ ;  $p = .73$ ).

### Gender Trends

Figure 1 depicts the gender effects found for Story A and Story B in both immediate recall and delayed recall (the two

immediate recall scores of Story B are averaged in this graph to produce a single mean immediate recall score for each gender subgroup). First, it is apparent that the general trend of the gender effect is the same for both Story A and Story B. Second, whereas Control women outperformed Control men for both the immediate and delayed recalls of Story A and Story B, the opposite gender effect is true for the AD group. In AD, women fell below men for both the immediate and delayed recalls (thus producing the “swapping” effect seen in Figure 1 where the lines cross when moving along the x-axis from Control to AD). The gender effect was not disturbed by the time delay between the immediate and delayed recalls.

### Classifying AD in men and women using the LM subtest

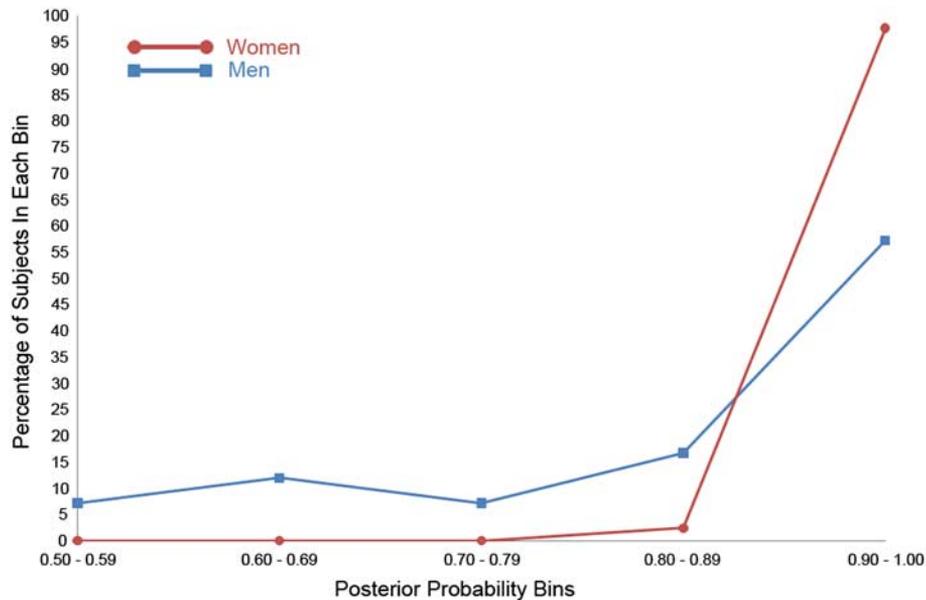
Because we hypothesized that the gender disparity on the LM would better diagnose AD in women than in men, we performed separate discriminant analyses on the two genders. Two discriminant functions were separately (for the men and the women) created and applied using the set of five recall measures of the LM subtest. The function for women classified each woman as either AD or Control and produced perfect 100% accuracy [Fisher's exact test,  $\chi^2(1, N = 42) = 42.0$ ;  $p < .0001$ ]. The function for men discriminated AD with 88% accuracy [Fisher's exact test,  $\chi^2(1, N = 42) = 24.4$ ;  $p < .0001$ ]. Thus, the LM performed better at detecting AD in women than in men.

The posterior probabilities of group membership that are supplied in discriminant analysis (Chapman et al., 2010) were compared for men and women (Figure 3). Individuals were placed into one of five posterior probability bins. Not only did the LM achieve greater classification accuracy for women than for men, but many more women had the highest probability of classification. Nearly 98% of the women lie in the highest probability bin (.9–1.0), while only 57% of the men do.

We analyzed whether there was a significant difference between the posterior probabilities of women as opposed to those of men. A median two-sample nonparametric test was chosen because the distributions are heavy-tailed. The sums of scores statistics were 36.0 for women and 6.0 for men, resulting in a one-sided  $p < .0001$ . Clearly the LM identified AD vs. Control group with greater accuracy and greater diagnostic confidence for women than it did for men.

## DISCUSSION

The results of this study indicate that the advantage normal elderly women have over men in learning and recalling verbal episodic memory disappears or reverses in AD. Control women outperformed Control men in every measure of the Logical Memory subtest (Figure 1). This was true for both Stories A and B and despite a time delay between the immediate recall and the delayed recall. Additionally, Control women had a higher rate of learning between the first and second immediate recalls of Story B (Figure 2).



**Fig. 3.** Percentage of women and men in each posterior probability bin when classified as AD or Control using the LM in discriminant analyses. Each woman and man was classified as either a member of the AD or Control group with associated posterior probability of group membership. The discriminant function was performed separately on women and men using the LM substest.

Men and women appeared to forget story details at a similar rate during a time delay filled with other activity (as depicted by loss of accuracy in detail recollection between the second immediate recall of Story B and the delayed recall in Figure 2).

In the AD group, however, it is men who outperform women in tasks of verbal episodic memory. On average, AD men scored higher than women on every aspect of the Logical Memory (across both stories and across the time delay). Clearly the better performance in women over men that was striking and statistically significant in normal elderly Controls has disappeared or reversed in AD, a conclusion supported by statistically significant Group  $\times$  Gender interactions (Figure 1; Table 2). This finding is most apparent in Figure 1 where the solid lines representing immediate recalls, as well as the dashed lines for delayed recalls, cross for each gender when moving from Control to AD. Control women scored higher on average than Control men for both stories and both immediate and delayed recalls, and this trend was reversed in AD (such that AD women scored lower on average than AD men). Additionally, the improved rate of learning seen in Control women disappears in AD women, although the rate of forgetting between AD men and women appears to be the same (Figure 2).

These findings are consistent with results reported by others that gender differences are present in AD and relevant to the study of the disease (Buckwalter, Sobel, Dunn, Diz, & Henderson, 1993; McPherson et al., 1999). Significant gender effects may be easier to detect when AD subjects are more advanced in the disease (Beinhoff et al., 2008). To focus on very mild AD, our inclusion criteria required a higher score on the MMSE ( $\geq 22$  in this study and in

Beinhoff et al. [2008] vs.  $\geq 15$  in the study by McPherson et al. [1999]). It is possible that the gender differences favoring men become more pronounced as AD progresses, as others have suggested (Ripich et al., 1995). In some studies of gender differences in AD, their samples of AD subjects were moderately to greatly impaired at the time of analysis as indicated by low mean MMSE scores (Buckwalter et al., 1996; Buckwalter, et al., 1993) and significant differences on the MMSE between AD men and women were found (Buckwalter et al., 1993). We had no such gender difference on the MMSE in our AD group (Table 1).

There have been some inconsistencies in the literature concerning whether or not AD women perform significantly worse than AD men on semantic tasks (Beinhoff et al., 2008). We believe it is important to consider the performance of normal elderly when assessing gender differences in AD. For instance, the drop in performance from Control to AD for the five measures we studied in the LM was approximately constant for women ( $M = -10.8$ ;  $SD = 1.9$ ;  $N = 5$ ) and for men ( $M = -6.7$ ;  $SD = 1.4$ ;  $N = 5$ ). These drops were statistically greater for women than for men ( $t(8) = 3.91$ ;  $p = .005$ ); on average they were 1.6 times greater for women. Moreover, these are sizeable drops considering the full range of scores is 0–25. In general, statistically significant differences between AD women and AD men may not appear on every measure because women have to fall so far from their normal control performance to be significantly below men as ADs. Therefore, where women and men start (normal cognition) is important in evaluating the gender difference related to AD.

The comparison of normal elderly to AD in terms of gender differences reveals two important findings. First, when other

demographic variables (including age, education, handedness) are controlled, gender differences still impact performance on the LM in normal elderly and in an opposite direction in AD. Second, not only do normal, healthy older women outperform healthy older men, this advantage strikingly disappears in AD. This suggests that AD pathology may affect memory systems differently in women than in men (Cushman & Duffy, 2007).

However, it has also been suggested that estrogen replacement therapy may provide protective effects on cognition and verbal memory in women (Henderson, Watt, & Buckwalter, 1996; Kampen & Sherwin, 1994; Korol & Manning, 2001); see also Baum (Baum, 2005). Because of a limited sample size of only three women in our subject pool actively taking hormone replacement therapy and limited knowledge of our subjects' previous medications, we could not explore the effects of this treatment on cognitive performance. It is possible that our AD women perform poorly on the LM because by chance the vast majority of them did not receive hormone replacement therapy at the onset of menopause. It has been shown that women taking estrogen replacement therapy during menopause have a reduced risk of cognitive impairment and dementia later in life (Maki, 2006; Tang et al., 1996; Yaffe, Sawaya, Lieberburg, & Grady, 1998; Zandi et al., 2002), and many of our Control women have perhaps benefited from this effect. Though the cause of improved verbal memory in normal elderly women and diminished verbal memory in AD women in this study cannot be definitely stated, the fact that women with AD tended to exhibit poorer verbal memory skills than AD men must be considered when using the LM as a diagnostic tool.

It is also possible that AD affects verbal IQ differently in women, and this IQ gender difference could have influenced our LM results. If this were so, we would expect significant gender effects on the AMNART estimated verbal IQ similar to our main LM interaction. We found no significant gender effect or significant group by gender interaction in our analysis of the AMNART (see Methods), so we cannot infer that the interaction we report on the LM is due to overall lower estimated verbal IQ in women.

### Effect of Gender Differences on AD Diagnosis

Our findings suggest an interesting and possibly diagnostically important fact about the Logical Memory test. This test has proven to be particularly useful in differentiating normal elderly from AD with high sensitivity and specificity (Chapman et al., 2010; Storandt & Hill, 1989). The drop in performance from healthy cognition to AD is much larger in women than in men. For immediate recall, women drop on average over both stories 9.6 recall items whereas men drop only 5.7. Similarly, for delayed recall, women drop on average 12.5 and men drop only 8.1. Due to this larger drop in performance in women, it is possible the LM may more reliably detect AD in women. To examine this, we applied discriminant analysis to the LM measures of men and women separately to classify each as a member of the AD or Control

group. We found that the LM was capable of diagnosing AD in women with 100% accuracy, and this was 12% more accurate than our discrimination results with men. In addition, the posterior probabilities were significantly higher for women (Figure 3) than for men, which suggests that these correct diagnoses for women were made with extremely high confidence (higher than the classifications for many of the men). This concept of the LM providing greater diagnostic power for women than men requires further study, but such a gender difference in diagnostic specificity has been reported for other measures of the WMS-III (Heaton, Taylor, & Manly, 2003). Taking gender into account when assessing cognitive decline with the LM may increase its power in early detection of AD.

We are unaware of any published, validated normative data for the LM that take gender into account. However, standardizing the data to account for gender differences produces essentially the same discriminant results. We included every normal elderly subject we had available to produce LM normative data separately for men and women and standardized each gender accordingly (there were an additional 49 Control women and 41 Control men, totaling 70 Control women and 62 Control men. These individuals were not included in the main analysis to maintain balance with the AD group). The normalized scores for the LM entered into discriminant analysis for women and men separately and achieved essentially the same results as the non-normalized data: 99% of the women and 87% of the men were correctly classified.

Although cognitively healthy women and men exhibited a statistically significant gender difference on the MMSE, the MMSE is not as capable of detecting AD in women as the LM. The MMSE has been shown to not be a powerful tool at detecting AD (Chapman et al., 2010; Costa et al., 1996; Heun, Papassotiropoulos, & Jennessen, 1988). Using the same subjects whose LM data are reported here, we found the MMSE could detect AD in women with 83% accuracy and in men with 74% accuracy. While the LM detected AD significantly better in women than in men, the MMSE did not show such a gender difference (Fisher's exact test,  $\chi^2(1, N = 42) = 1.13; p = .29$ ). In addition, the LM achieved significantly higher accuracy in detecting AD in women over what the MMSE achieved (Fisher's exact test,  $\chi^2(1, N = 42) = 7.63; p = .001$ ). The MMSE and LM did not perform significantly differently in diagnosing AD in men (Fisher's exact test,  $\chi^2(1, N = 42) = 2.79; p = .10$ ). The MMSE is not as useful a tool in detecting AD, and the gender effect that boosts the power the LM has in assessing women is not present.

Our LM results require further examination with the WMS-IV, although it seems likely that the gender effects would persist given the minor changes to the LM in the newer version of the test. Additionally, we chose to focus on early-stage AD as identification of those individuals in the earliest stage of the disease is increasingly important, as well as understanding the nascent cognitive changes associated with early symptoms. These gender differences should be studied

both (1) in individuals with more advanced AD to better understand their relationship to the progression of the disease and (2) in individuals with Mild Cognitive Impairment and preclinical AD.

## ACKNOWLEDGMENTS

Maria Guillily is now at the Department of Pharmacology and Experimental Therapeutics at Boston University. Tiffany Sandoval is now at the San Diego State University/University of California at San Diego Joint Doctoral Program in Clinical Psychology. Elizabeth DeGrush is now at the Chicago College of Osteopathic Medicine at Midwestern University. Lindsey Reilly is now at the Springer Publishing Company. We thank: the Memory Disorders Clinic, University of Rochester Medical Center, Monroe Community Hospital, the Alzheimer's Disease Center, especially Anton Porsteinsson, Paul Coleman, Charles Duffy, and Roger Kurlan, for their strong support of our research; Robert Emerson and William Vaughn for their technical contributions; Rafael Klorman for critical discussions; Susan E. Chapman for help in writing; Cendrine Robinson for initial data analysis, research, and critical thinking; Dustina Holt, Jonathan DeRight, Kristen Morie, Anna Fagan, Michael Garber-Barron, Leon Tsao, and Brittany Huber for technical help; and the many voluntary subjects in this research. This research was supported by the National Institute of Health grants P30-AG08665, R01-AG018880, and P30-EY01319. The authors of this manuscript have no financial or other relationships that could be interpreted as a conflict of interest.

## REFERENCES

- Abikoff, H., Alvir, J., Hong, G., Sukoff, R., Orazio, J., Solomon, S., & Saravay, S. (1987). Logical memory subtest of the Wechsler Memory Scale: Age and education norms and alternate-form reliability of two scoring systems. *Journal of Clinical and Experimental Neuropsychology*, 9(4), 435–448.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of the mental disorders* (4th, text rev. ed.). Washington, DC: Author.
- Azad, N.A., Al Bugami, M., & Loy-English, I. (2007). Gender differences in dementia risk factors. *Gender Medicine*, 4(2), 120–129.
- Baum, L.W. (2005). Sex, hormones, and Alzheimer's disease. *Journal of Gerontology: Medical Sciences*, 60A(6), 736–743.
- Beinhoff, U., Tumani, H., Bretschneider, J., Bittner, D., & Riepe, M.W. (2008). Gender-specificities in Alzheimer's disease and mild cognitive impairment. *Journal of Neurology*, 255(1), 117–122.
- Benedict, R.H.B., & Groninger, L. (1995). Preliminary standardization of a new visuospatial memory tests with six alternate forms. *The Clinical Neuropsychologist*, 9, 11–16.
- Benton, A.L., & Hamsher, K. (1978). *Multilingual aphasia examination*. Iowa City: University of Iowa.
- Bleecker, M.L., Bolla-Wilson, K., Agnew, J., & Meyers, D.A. (1988). Age-related sex differences in verbal memory. *Journal of Clinical Psychology*, 44(3), 403–411.
- Brouwers, P., Cox, C., Martin, A., Chase, T., & Fedio, P. (1984). Differential perceptual-spatial impairment in Huntington's and Alzheimer's dementias. *Archives of Neurology*, 41, 1073–1076.
- Buckwalter, J.G., Rizzo, A.A., McCleary, R., Shankle, R., Dick, M., & Henderson, V.W. (1996). Gender comparisons of cognitive performance among vascular dementia, Alzheimer disease, and older adults without dementia. *Archives of Neurology*, 53, 436–439.
- Buckwalter, J.G., Sobel, E., Dunn, M.E., Diz, M.M., & Henderson, V.W. (1993). Gender differences on a brief measure of cognitive functioning in Alzheimer's disease. *Archives of Neurology*, 50, 757–760.
- Chapman, R.M., Mapstone, M., McCrary, J.W., Gardner, M.N., Porsteinsson, A., DeGrush, E., ... Guillily, M.D. (2011). Predicting conversion from mild cognitive impairment to Alzheimer's disease using neuropsychological tests and multivariate methods. *Journal of Clinical and Experimental Neuropsychology*, 33(2), 187–199. doi:10.1080/13803395.2010.499356
- Chapman, R.M., Mapstone, M., Porsteinsson, A.P., Gardner, M.N., McCrary, J.W., DeGrush, E., ... Guillily, M.D. (2010). Diagnosis of Alzheimer's disease using neuropsychological testing improved by multivariate analyses. *Journal of Clinical and Experimental Neuropsychology*, 32(8), 793–808. doi:10.1080/13803390903540315
- Costa, P.T., Albert, M.S., Butters, N.M., Folstein, M.F., Gilman, S., Gurland, B.J., ... Williams, T.F. (1996). *Early identification of Alzheimer's disease and related dementias. Clinical practice guideline, quick reference guide for clinicians, No. 19*. Publication No. 97-0703. Rockville, MD: AHCPR.
- Cushman, L.A., & Duffy, C.J. (2007). The sex specificity of navigational strategies in Alzheimer disease. *Alzheimer Disease and Associated Disorders*, 21(2), 122–129.
- Denckla, M.B., & Rudel, R. (1974). Rapid "automatized" naming of pictured objects, colors, letters, and numbers by normal children. *Cortex*, 10, 186–202.
- Folstein, M.F., Folstein, S.E., & McHugh, P.R. (1975). "Minimal state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3), 189–198.
- Fuld, P. (1977). *Full object-memory evaluation: Instruction manual*. Chicago: Stoelting.
- Golden, C.J. (1978). *Stroop color and word test: A manual for clinical and experimental uses*. Chicago: Stoelting.
- Grober, E., & Sliwinski, M. (1991). Development and validation of a model for estimating premorbid verbal intelligence in the elderly. *Journal of Clinical and Experimental Neuropsychology*, 13, 933–949.
- Heaton, R.K., Taylor, M.J., & Manly, J. (2003). Demographic effects and use of demographically corrected norms with the WAIS-III and WMS-III. In D.S. Tulskey, D.H. Saklofske, R.K. Heaton, R. Bornstein, M.F. Ledbetter, G.J. Chelune, R.J. Ivnik & A. Prifitera (Eds.), *Clinical Interpretation of the WAIS-III and WMS-III* (pp. 183–210). San Diego: Academic Press.
- Henderson, V.W., Watt, L., & Buckwalter, J.G. (1996). Cognitive skills associated with estrogen replacement in women with Alzheimer's disease. *Psychoneuroendocrinology*, 21(4), 421–430.
- Heun, R., & Kockler, M. (2002). Gender differences in the cognitive impairment in Alzheimer's disease. *Archives of Women's Mental Health*, 4, 129–137.
- Heun, R., Papassotiropoulos, A., & Jennssen, F. (1988). The validity of psychometric instruments for detection of dementia in the elderly general population. *International Journal of Geriatric Psychiatry*, 13(6), 368–380.
- Kampen, D.L., & Sherwin, B.B. (1994). Estrogen use and verbal memory in healthy post-menopausal women. *Obstetrics and Gynecology*, 83, 979–983.
- Kolb, B., & Wishaw, Q. (1985). *Fundamentals of human neuropsychology* (2nd ed.). San Francisco: W.H. Freeman.

- Korol, D.L., & Manning, C.A. (2001). Effects of estrogen on cognition: Implications for menopause. In M.E. Carroll, & J.B. Overmier (Eds.), *Animal research and human health: Advancing human welfare through behavioral science* (pp. 305–322). Washington, DC: American Psychological Association Book Publishers.
- Kramer, J.H., Delis, D.C., & Daniel, M. (1988). Sex differences in verbal learning. *Journal of Clinical Psychology, 44*(6), 907–915.
- Lezak, M.D., Howieson, D.B., & Loring, D.W. (2004). *Neuropsychological Assessment* (4th ed.). Oxford: Oxford University Press.
- Maki, P.M. (2006). Hormone therapy and cognitive function: Is there a critical period for benefit? *Neuroscience, 138*, 1027–1030.
- McKhann, G., Drachman, D., Folstein, M.F., Katzman, R., Price, D., & Stadlan, E.M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's disease. *Neurology, 34*, 939–944.
- McPherson, S., Back, C., Buckwalter, J.G., & Cummings, J.L. (1999). Gender-related cognitive deficits in Alzheimer's disease. *International Psychogeriatrics, 11*(2), 117–122.
- Millet, X., Raoux, N., Le Carret, N., Bouisson, J., Dartigues, J.F., & Amieva, H. (2009). Gender-related differences in visuospatial memory persist in Alzheimer's disease. *Archives of Clinical Neuropsychology, 24*(8), 783–789.
- Money, J.A. (1976). *A standardized road map test of directional sense, manual*. San Rafael: Academic Therapy Publications.
- Morris, J.C., Heyman, A., Mohs, R.C., Hughes, J.P., van Belle, G., Fillenbaum, G., ... Clark, C. (1989). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology, 39*(9), 1159–1165.
- Petersen, R.C., Stevens, J.C., Ganguli, M., Tangalos, E.G., Cummings, J.L., & DeKosky, S.T. (2001). Practice parameters: Early detection of dementia: Mild cognitive impairment (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology, 56*, 1133–1142.
- Ripich, D.N., Pettrill, S.A., Whitehouse, P.J., & Zioli, E.W. (1995). Gender differences in language of AD patients: A longitudinal study. *Neurology, 45*, 299–302.
- SAS Institute Inc. (2002). SAS OnlineDoc 9.1.3, 2007. Retrieved from <http://support.sas.com/onlinedoc/913/docMainpage.jsp>
- Storandt, M., & Hill, R.D. (1989). Very mild senile dementia of the Alzheimer type. II. Psychometric test performance. *Archives of Neurology, 46*(4), 383–386.
- Tang, M.-X., Jacobs, D., Stern, Y., Marder, K., Schofield, P., Gurland, B., ... Mayeux, R. (1996). Effect of oestrogen during menopause on risk and age at onset of Alzheimer's disease. *Lancet, 348*(9025), 429–432.
- Tuokko, H., Hadjistavropoulos, T., Miller, J.A., & Beattie, B.L. (1992). The clock test: A sensitive measure to differentiate normal elderly from those with Alzheimer's disease. *Journal of the American Geriatrics Society, 40*, 579–584.
- Wechsler, D. (1945). A standardized memory scale for clinical use. *The Journal of Psychology, 19*, 87–95.
- Wechsler, D. (1987). *Wechsler memory scale-revised manual*. San Antonio: The Psychological Corporation.
- Wechsler, D. (1997). *Wechsler memory scale III: Administration and scoring manual*. Orlando: Psychological Corporation.
- Woodard, J.L., Seidenberg, M., Nielson, K.A., Antuono, P., Guidotti, L., Durgerian, S., ... Rao, S.M. (2009). Semantic memory activation in amnesic mild cognitive impairment. *Brain, 132*, 2068–2078.
- Yaffe, K., Sawaya, G., Lieberburg, I., & Grady, D. (1998). Estrogen therapy in postmenopausal women: Effects on cognitive function and dementia. *Journal of the American Medical Association, 279*(9), 688–695.
- Yeudall, L.T., Fromm, D., Reddon, J.R., & Stefanyk, W.O. (1986). Normative data stratified by age and sex for 12 neuropsychological tests. *Journal of Clinical Psychology, 42*, 918–946.
- Zandi, P.P., Carlson, M.C., Plassman, B.L., Welsh-Bohmer, K.A., Mayer, L.S., Steffens, D.C., & Breitner, J.C.S. (2002). Hormone replacement therapy and incidence of Alzheimer's disease in older women: The Cache County Study. *Journal of the American Medical Association, 288*(17), 2123–2129.