

# Sex-specific norms for verbal memory tests may improve diagnostic accuracy of amnesic MCI

Erin E. Sundermann, PhD,\* Pauline Maki, PhD,\* Anat Biegon, PhD, Richard B. Lipton, MD, Michelle M. Mielke, PhD, Mary Machulda, PhD, LP, and Mark W. Bondi, PhD, for the Alzheimer's Disease Neuroimaging Initiative

*Neurology*® 2019;93:e1881-e1889. doi:10.1212/WNL.0000000000008467

## Correspondence

Dr. Sundermann  
esundermann@  
einstein.yu.edu.

## Abstract

### Objective

To examine whether the use of sex-specific norms and cut scores to identify memory impairment improves diagnostic accuracy of amnesic mild cognitive impairment (aMCI) compared to non-sex-specific (typical) norms/cut scores given the female advantage in verbal memory.

### Methods

We calculated sex-specific and typical norms/cut scores (age and education specific) for impairment on the Rey Auditory Verbal Learning Test in the Mayo Clinic Study of Aging. Norms/cut scores were applied to 453 women and 532 men from the Alzheimer's Disease Neuroimaging Initiative. We compared sex differences in rates of aMCI (Jak/Bondi criteria) for sex-specific vs typical norms/cut scores. Using sex-specific cut scores as the true condition and typical cut scores as the predicted condition, we categorized participants as true positives (TPs), false positives (FPs), true negative (TNs), or false negative (FNs). In cross-sectional analyses within sex, we compared positivity rates of CSF hyperphosphorylated tau/ $\beta$ -amyloid (A $\beta$ ) and cortical A $\beta$  deposition ( $[^{18}\text{F}]\text{AV45}$  PET) and *APOE*  $\epsilon$ 4 frequency among diagnostic comparison groups.

### Results

The frequency of aMCI was higher in men when using typical norms/cut scores. Using sex-adjusted norms/cut scores led to the identification of 10% FNs (missed aMCI cases) among women and 10% FPs among men. Biomarker analyses supported the hypothesis that sex-specific diagnostic criteria improves diagnostic accuracy. Biomarker rates were higher in FNs vs TNs and similar in FNs and TPs. Biomarker rates were lower in FPs vs TPs and similar between FPs and TNs.

### Conclusion

Results suggest that non-sex-specific aMCI diagnostic criteria led to a 20% diagnostic error rate. Accounting for sex differences in verbal memory performance may improve aMCI classification.

## MORE ONLINE

### CME Course

[NPub.org/cmelist](http://NPub.org/cmelist)

\*These authors contributed equally to this work.

From the Department of Psychiatry (E.E.S., M.W.B.), University of California, San Diego; Departments of Psychiatry, Psychology, and Obstetrics & Gynecology (P.M.), University of Illinois at Chicago; Department of Radiology and Neurology (A.B.), State University of New York, Stony Brook; Einstein Aging Study and Department of Neurology (R.B.L.), Albert Einstein College of Medicine, Bronx, NY; Departments of Health Sciences Research and Neurology (M.M.M.) and Departments of Psychiatry and Psychology (M.M.), Mayo Clinic, Rochester, MN; and Veterans Affairs San Diego Healthcare System (M.W.B.), CA.

Go to [Neurology.org/N](http://Neurology.org/N) for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article.

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database ([adni.loni.usc.edu](http://adni.loni.usc.edu)). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found in Supplemental Data File and at: [links.lww.com/WNL/A992](http://links.lww.com/WNL/A992).

## Glossary

$A\beta$  =  $\beta$ -amyloid; AD = Alzheimer disease; ADNI = Alzheimer's Disease Neuroimaging Initiative; aMCI = amnesic mild cognitive impairment; ANCOVA = analysis of covariance; CDR-SOB = Clinical Dementia Rating Sum of Boxes; CI = confidence interval; FN = false negative; FP = false positive; NC = normal control; OR = odds ratio; p-tau = hyperphosphorylated tau; RAVLT = Rey Auditory Verbal Learning Test.; SUVR = standardized uptake value ratio; TN = true negative; TP = true positive.

Verbal memory deficits are typically required for diagnosis of Alzheimer disease (AD) dementia and its precursor, amnesic mild cognitive impairment (aMCI). Normative data for verbal memory tests usually adjust for age and/or education but not sex despite the well-documented, lifelong female advantage on these tests.<sup>1–3</sup> We demonstrated a female advantage in verbal memory in aMCI despite no sex difference in disease burden (i.e., hippocampal atrophy, brain hypometabolism, and cortical amyloid deposition).<sup>4–6</sup> These findings suggest that the female advantage in verbal memory may be a proxy for a verbal memory-specific increase in cognitive reserve<sup>7–9</sup> that allows women to sustain what is considered normal verbal memory performance and delay aMCI diagnosis until more advanced disease states compared to men. Consistent with the cognitive reserve theory, the female advantage is eliminated with advanced disease burden,<sup>10,11</sup> suggesting that decline is more accelerated in women than men after aMCI diagnosis.

Because women outperform men on verbal memory tests, we tested the hypotheses that defining memory impairment with the typical non-sex-specific norms and cut scores for verbal memory tests leads to underdiagnosis of aMCI in women and overdiagnosis of aMCI in men and that use of sex-stratified norms/cut scores improves these diagnostic errors. We hypothesized that sex-specific norms/cut scores would reclassify a subset of women from normal control (NC) to aMCI and a subset of men from aMCI to NC and that group comparisons of AD-associated and genetic markers and biomarkers would validate those reclassifications (e.g., women reclassified as aMCI would have more advanced biomarkers than women consistently classified as NCs).

## Methods

### Participants and data source

Cross-sectional data were extracted from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu) in October 2016. Detailed information about ADNI can be found at adni-info.org. ADNI began in 2004 as a longitudinal, multisite cohort study that recruited healthy older adults and individuals with early or late MCI and early AD. See loni.ucla.edu/ADNI for recruitment procedures<sup>12</sup> and adni-info.org/Scientists/ADNIGrant/ProtocolSummary.aspx for eligibility criteria. We included participants without dementia (based on conventional ADNI criteria) with neuropsychological and AD pathologic marker data at baseline (n = 1,065). Because we were interested specifically in aMCI as a precursor to AD dementia,

we excluded individuals with a non-aMCI diagnosis based on both the typical (non-sex-specific) and the newly generated sex-specific Jak/Bondi diagnostic criteria (n = 72). Our final sample consisted of 985 participants (289 from ADNI1 and 696 from ADNI2) who were 55 to 90 years of age, not depressed, and diagnosed as either NC (n = 631) or MCI (n = 354) on the basis of Jak/Bondi diagnostic criteria as typically applied (non-sex-specific verbal memory norms/cut scores).<sup>13</sup>

### Standard protocol approvals, registrations, and patient consents

ADNI was approved by the institutional review board at each site and was compliant with the Health Insurance Portability and Accountability Act. Written consent was obtained from all participants.

### Neuropsychological tests

A battery of neuropsychological tests is administered to all ADNI participants at baseline. Our primary focus was the Rey Auditory Verbal Learning Test (RAVLT) verbal memory test. The RAVLT involves learning a list of 15 unrelated words and recalling as many words as possible across 5 immediate recall trials (Learning score, range 0–75), after learning an interference list, and after a 30-minute delay period (Delayed Recall score, range 0–15).<sup>14</sup> The RAVLT has previously shown a strong female advantage.<sup>15</sup> The battery also includes tests of language (Category Fluency Test [animals], Boston Naming Test [30 items]) and speed/executive function (Trail-Making Test Parts A and B). The Geriatric Depression Scale<sup>16</sup> and the Clinical Dementia Rating Sum of Boxes (CDR-SOB) were also administered.

### AD-associated markers

Presence or absence of the APOE  $\epsilon$ 4 allele was determined for each participant. Our biomarker outcomes included the CSF ratio of hyperphosphorylated tau (p-tau<sub>181</sub>) to  $\beta$ -amyloid ( $A\beta_{42}$ ), and cortical  $A\beta$  deposition as measured by [<sup>18</sup>F]AV45 PET. PET methods are described at adni-info.org. Mean AV45 uptake was measured within 4 regions (frontal, anterior/posterior cingulate, lateral parietal, and lateral temporal cortex), and standardized uptake value ratios (SUVRs) were calculated by averaging across cortical regions and dividing by the whole cerebellum. Participants were dichotomized into positive and negative groups on the basis of previously established cutoffs for CSF p-tau<sub>181</sub>/ $A\beta_{42}$  (0.11)<sup>17</sup> and AV45 SUVR (1.11).<sup>18</sup> In exploratory analyses, we examined a clinical marker, the CDR-SOB score, which measures dementia severity (score range 0–18).<sup>19</sup>

## Diagnostic criteria

The Jak/Bondi actuarial neuropsychological diagnostic method<sup>13</sup> was applied to baseline data. Application of the Jak/Bondi diagnosis to aging cohorts (including ADNI) produced more discernible cognitive phenotypes, more stable diagnoses, stronger associations with AD biomarkers, and better prediction of progression to dementia compared to conventional MCI diagnostic criteria.<sup>13,20</sup> The Jak/Bondi diagnostic method relies on performance on 6 neuropsychological tests from ADNI representing 3 cognitive domains: Category Fluency and Boston Naming Test (language), Trail-Making Test Parts A and B (speed/executive function), and RAVLT learning and delayed recall (verbal memory). These tests were chosen because of their sensitivity to early cognitive deficits in AD and their consistent administration across the 3 ADNI phases (ADNI-1, -GO, and -2). The language and speed/executive function tests were converted to an age- and education-adjusted standard score using published normative data from the National Alzheimer's Coordinating Center ( $n = 3,286$ ).<sup>21,22</sup>

We derived 2 sets of normative data for the RAVLT outcomes in a normative sample of participants from the Mayo Clinic Study of Aging (MCSA) who were diagnosed as NC at baseline and at least 2 follow-up visits at 15 months apart. The normative sample consisted of 810 women and 810 men (age range 50–89 years, mean education 14.5 [SD = 2.5] years, 98.8%). Age- and education-specific norms were derived by generating mean RAVLT scores and their SDs separately for less educated (8–15 years of education) and more educated ( $\geq 16$  years of education) participants within each 5-year age bin (e.g., 55–59, 60–64, 65–69 years). The sex-specific norms were derived in the same way except participants were separated by sex as well as age and education categories (e.g., female, highly educated, 55–59 years of age). We calculated RAVLT norms in the MCSA rather than using published sets of norms because published norms vary widely in terms of what was adjusted for, the criteria for inclusion in a normative sample, and the size, age, and sex distribution of the normative sample. By calculating our norms in the MCSA, we were able to standardize the calculation of norms across the compared norm sets (age, education, sex adjusted vs age, education adjusted) to specifically examine the difference in a sex adjustment to norms.

Participants were considered to have aMCI if one of the following criteria was met: (1) they had an impaired score, defined as  $>1$  SD below the normative mean, on both RAVLT Learning and Delayed Recall (regardless of performance in other cognitive domains); or (2) they had at least 1 impaired score, defined as  $>1$  SD below the normative mean, in each of the 3 cognitive domains sampled (memory, language, speed/executive function). We applied both the typical Jak/Bondi diagnostic method (age- and education-specific RAVLT norms/cut scores) and the adapted Jak/Bondi method that incorporated sex-specific RAVLT norms/cut scores to all participants' data, thus generating 2 different diagnostic groupings. When the adapted sex-specific Jak/Bondi diagnostic criteria

were applied, the difference in cut scores between men and women ranged from 3 to 8 points (mean 6 points) for RAVLT Learning and 1 to 4 points (mean 2 points) for RAVLT Delayed across age and education categories.

## Statistical analyses

Using the sex-specific diagnostic groupings as the true condition and the typical diagnostic groupings as the predicted condition, we labeled participants as follows: (1) true positive (TP), diagnosed as having aMCI with both sets of cut scores; (2) false positive (FP), diagnosed as having aMCI with typical cut scores but NC with sex-specific cut scores; (3) true negative (TN), diagnosed as NC with both sets of cut scores; and (4) false negative (FN), diagnosed as NC with typical cut scores but aMCI with sex-specific cut scores. Sample characteristics were compared among the resulting diagnostic comparison groups with  $\chi^2$  tests for categorical variables and analyses of variance for continuous variables. Two separate  $\chi^2$  tests were conducted to examine whether the frequency of aMCI diagnoses differed between men and women when the typical diagnostic criteria were applied (first  $\chi^2$ ) and then when the sex-specific diagnostic criteria were applied (second  $\chi^2$ ). In sex-stratified analyses, we used multivariable logistic regressions to compare APOE  $\epsilon 4$  allele frequency and positivity rates of CSF p-tau/A $\beta$  and AV45 SUVR by diagnostic comparison groups, and we used analysis of covariance (ANCOVA) to compare CDR-SOB means by diagnostic comparison groups. Analyses were adjusted for age and education. We used dummy coding to the diagnostic comparison groups of TP, TN, and FN for women and TP, TN, and FP for men. We ran regressions comparing the diagnostic comparison groups to the reference group of TNs. If the overall diagnostic comparison group variable was significant, then regressions were also conducted with the TP group as the reference group to compare the odds of the AD-related markers for the FN (women only) and FP (men only) groups vs both the TN and TP groups.

## Data availability

ADNI data were obtained from [adni.loni.usc.edu](http://adni.loni.usc.edu) and are available to investigators in the scientific community who have been approved by the ADNI Data Sharing and Publications Committee and who agree to the terms of the ADNI Data Use Agreement for purposes of replicating procedures and results.

## Results

The ADNI participants included 453 women and 532 men (94% white, mean age 72.9 years [SD 7.0 years], mean education 16.3 years [SD 2.7 years]). Across diagnostic groups, women had a significantly lower mean age (mean 71.9, SD 6.8 vs mean 73.6, SD 7.0 years) and years of education (mean 15.7, SD 2.6 vs mean 16.7, SD 2.6 years) and a significantly higher mean Mini-Mental State Examination score (mean 28, SD 1.6 vs mean 28.1, SD 1.7) and Geriatric Depression Scale

score (mean 1.5, SD 1.5 vs mean 1.3, SD 1.4) compared to men ( $p < 0.05$ ). As expected, women outperformed men on the RAVLT Learning (mean 42.3, SD 11.5 vs mean 35.6, SD 11.1) and Delayed Recall (mean 6.2, SD 4.5 vs mean 4.5, SD 3.9) outcomes ( $p < 0.001$ ). When typical cut scores were used, the frequency of aMCI diagnosis was significantly higher in men than women ( $p < 0.001$ ); however, this sex difference was eliminated when sex-specific cut scores were used ( $p = 0.62$ , table 1). Among men, 184 (35%) were categorized as TP, 293 (55%) as TN, and 55 (10%) as FP (table 2). No men were categorized as FN. Among women, 120 (26%) were categorized as TP, 288 (64%) as TN, and 45 (10%) as FN. No women were categorized as FP. Table 3 shows sample characteristics and AD-related markers (continuous levels of biomarkers) for each diagnostic group stratified by sex.

## AD-associated markers

### Women

Among women, the likelihood of positivity for cortical amyloid and CSF p-tau/A $\beta$  ratio and carrying an APOE  $\epsilon 4$  allele differed by diagnostic group ( $p \leq 0.001$ ). Specifically, the likelihood of cortical amyloid positivity in FN women was 3.6 times greater compared to TN women (odds ratio [OR] 3.60, 95% confidence interval [CI] 1.58–8.18,  $p = 0.002$ ) but did not differ from TP women (OR 1.43, 95% CI 0.57–3.59,  $p = 0.45$ ; figure 1). Similarly, the likelihood of positivity for the CSF p-tau/A $\beta$  ratio in FN women was >2 times higher than in TN women (OR 2.46, 95% CI 1.13–5.33,  $p = 0.02$ ) but did not differ from TP women (OR 0.65, 95% CI 0.26–1.59,  $p = 0.34$ ). The likelihood of having an APOE  $\epsilon 4$  allele in FN women was almost 5-fold higher than in TN women (OR 4.91, 95% CI 2.46–9.80,  $p < 0.001$ ) but did not differ from TP women (OR 1.36, 95% CI 0.64–2.87,  $p = 0.99$ ). In an ANCOVA, mean CDR-SOB score significantly differed by diagnostic groups ( $F_{2,575} = 96.3$ ,  $p < 0.001$ ). Specifically, mean CDR-SOB score for FN women was significantly higher than that for TN women ( $F_{1,329} = 14.3$ ,  $p < 0.001$ ) and significantly lower than that for TP women ( $F_{1,161} = 9.5$ ,  $p = 0.002$ ).

### Men

Among men, the likelihood of positivity for cortical amyloid, CSF p-tau/A $\beta$  ratio, and carrying an APOE  $\epsilon 4$  allele also differed by diagnostic group ( $p < 0.001$ ). Specifically, the likelihood of cortical amyloid positivity in FP men was less than in TP men (OR 0.45, 95% CI 0.21–0.97,  $p = 0.04$ ) but did not differ from TN men (OR 1.76, 95% CI 0.87–3.54,  $p = 0.11$ ; figure 2). Similarly, the likelihood of positivity for CSF p-tau/A $\beta$  ratio in FP men was significantly less than in TP men (OR 0.47, 95% CI 0.23–0.98,  $p = 0.04$ ) but did not differ from TN men (OR 1.23, 95% CI 0.64–2.28,  $p = 0.53$ ). The likelihood of having the APOE  $\epsilon 4$  allele in FP men was lower than in TP men (OR 0.63, 95% CI 0.38–1.06,  $p = 0.08$ ) and higher than in TN men (OR 1.50, 95% CI 0.91–2.48,  $p = 0.11$ ), but not significantly. In an ANCOVA, mean CDR-SOB score significantly differed by diagnostic groups ( $F_{2,527} = 57.0$ ,  $p < 0.001$ ). Specifically, mean CDR-SOB score for FP men was significantly

**Table 1** Comparison of sex differences in the frequency of aMCI diagnosis when using sex-specific aMCI diagnostic criteria vs conventional aMCI diagnostic criteria

	Women with aMCI, n (%)	Men with aMCI, n (%)	$\chi^2$ Test, p value
aMCI classified via conventional diagnostic criteria	120 (26)	239 (45)	38.5, <0.001
aMCI classified via sex-specific diagnostic criteria	165 (36)	184 (35)	0.25, 0.66

Abbreviation: aMCI = amnesic mild cognitive impairment.

higher than for TN men ( $F_{1,344} = 13.8$ ,  $p < 0.001$ ) and significantly lower than for TP men ( $F_{1,235} = 7.9$ ,  $p = 0.005$ ).

## Discussion

Previous work in ADNI showed that sex differences in verbal memory persisted in aMCI and that women with aMCI retained an advantage in verbal memory over men<sup>4–6,23</sup> despite showing similar levels of hippocampal volume, brain glucose hypometabolism, and cortical A $\beta$  deposition.<sup>4–6</sup> These findings suggested a cognitive benefit for women with aMCI because women had better memory than men despite the similar levels of underlying brain pathology. However, a negative consequence of this verbal memory advantage among women is that it might mask early signs of AD and delay aMCI diagnosis. Indeed, it has been suggested that women with aMCI decline faster compared to men with aMCI.<sup>24</sup> Because verbal memory tests have been used clinically to diagnosis aMCI and women perform better on these tests, we tested the hypothesis that the use of typical (non-

**Table 2** Comparison of clinical diagnosis, by sex, when using the typical vs sex-specific aMCI diagnostic criteria

	Typical diagnostic criteria, n (%)	Sex-specific diagnostic criteria, n (%)	FN, n (%)	FP, n (%)
<b>Men (n = 532)</b>				
NC	293 (55)	348 (65)	—	—
aMCI	239 (45)	184 (35)	0 (0)	55 (10)
<b>Women (n = 453)</b>				
NC	333 (73)	288 (64)	—	—
aMCI	120 (26)	165 (36)	(10)	0 (0)

Abbreviations: aMCI = amnesic mild cognitive impairment; FN = false negative; FP = false positive; NC = normal control.

**Table 3** Sample characteristics of male and female diagnostic groups resulting from the comparison of sex-specific aMCI diagnostic criteria vs typical aMCI diagnostic criteria

Parameters	Female subjects (n = 453)				Male subjects (n = 532)			
	TP (n = 120), mean (SD)	FN (n = 45), mean (SD)	TN (n = 288), mean (SD)	p Value	TP (n = 184), mean (SD)	FP (n = 55), mean (SD)	TN (n = 293), mean (SD)	p Value
<b>Demographic/clinical characteristics</b>								
Age, y	72.1 (7.4)	72.1 (6.6)	72.9 (6.7)	0.38	72.9 (7.2) <sup>c,d</sup>	75.7 (7.0) <sup>d</sup>	73.7 (6.8) <sup>c</sup>	0.04
Education, y	15.5 (2.9)	14.7 (2.5) <sup>b</sup>	15.8 (2.6) <sup>b</sup>	0.01	16.5 (2.5) <sup>c</sup>	15.9 (3.0) <sup>e</sup>	17.0 (2.6) <sup>c,e</sup>	0.004
White, %	94.0	97.0	95.3	0.73	99.4	96.2	96.5	0.12
Global cognition (MMSE) score	27.0 (1.8) <sup>a,c</sup>	28.1 (1.7) <sup>a,b</sup>	28.9 (1.3) <sup>b,c</sup>	<0.001	27.3 (1.9) <sup>d</sup>	28.0 (1.8) <sup>d,e</sup>	28.6 (1.4) <sup>e</sup>	<0.001
GDS score	1.7 (1.5)	1.2 (1.4)	1.3 (1.5)	0.06	1.5 (1.5) <sup>c</sup>	1.4 (1.3)	1.1 (1.3) <sup>c</sup>	0.003
RAVLT Learning score	28.3 (6.1) <sup>a,c</sup>	35.8 (4.8) <sup>a,b</sup>	48.4 (8.3) <sup>b,c</sup>	<0.001	25.9 (5.5) <sup>c,d</sup>	30.4 (4.2) <sup>d,e</sup>	42.9 (9.2) <sup>c,e</sup>	<0.001
RAVLT Delayed Recall score	1.4 (1.7) <sup>a,c</sup>	3.9 (2.3) <sup>a,b</sup>	8.5 (3.8) <sup>b,c</sup>	<0.001	1.4 (1.5) <sup>c,d</sup>	2.3 (1.8) <sup>d,e</sup>	6.9 (3.7) <sup>c,e</sup>	<0.001
<b>AD-related markers</b>								
CDR-SOB score	1.6 (0.9) <sup>a,c</sup>	1.1 (1.0) <sup>a,b</sup>	0.5 (0.8) <sup>b,c</sup>	<0.001	1.6 (1.1) <sup>c,d</sup>	1.1 (0.9) <sup>d,e</sup>	0.6 (0.8) <sup>c,e</sup>	<0.001
APOE ε4 carrier, %	60.7 <sup>c</sup>	60.7 <sup>b</sup>	30.2 <sup>b,c</sup>	<0.001	52.9 <sup>c</sup>	42.8	33.9 <sup>c</sup>	<0.001
CSF p-tau/Aβ <sub>42</sub>	0.35 (0.25) <sup>c</sup>	0.31 (0.27) <sup>b</sup>	0.19 (0.16) <sup>b,c</sup>	<0.001	0.29 (0.18) <sup>c</sup>	0.26 (0.18) <sup>e</sup>	0.19 (0.15) <sup>c,e</sup>	<0.001
Cortical Aβ <sub>42</sub> burden (AV45 SUVR)	1.28 (0.23) <sup>c</sup>	1.25 (0.21) <sup>b</sup>	1.16 (0.19) <sup>b,c</sup>	<0.001	1.26 (0.23) <sup>c</sup>	1.18 (0.20)	1.12 (0.18) <sup>c</sup>	<0.001

Abbreviations: Aβ<sub>42</sub> = amyloid-β peptide; AD = Alzheimer disease; aMCI = amnesic mild cognitive impairment; CDR-SOB = Clinical Dementia Rating Sum of Boxes; FN = false negative (diagnosed cognitively normal by conventional cut scores but diagnosed as aMCI by sex-specific cut scores); FP = false positive (diagnosed aMCI by conventional cut scores but diagnosed as cognitively normal by sex-specific cut scores); GDS = Geriatric Depression Scale; MMSE = Mini-Mental State Examination; p-tau = hyperphosphorylated tau; RAVLT = Rey Auditory Verbal Learning Test; SUVR = standardized uptake value ratio; TN = true negative (diagnosed as cognitively normal by both conventional and sex-specific cut scores); TP = true positive (diagnosed aMCI by both conventional and sex-specific cut scores). Significance was defined as  $p < 0.05$ .

<sup>a</sup> Significant difference between TP and FN.

<sup>b</sup> Significant difference between FN and TN.

<sup>c</sup> Significant difference between TP and TN.

<sup>d</sup> Significant difference between TP and FP.

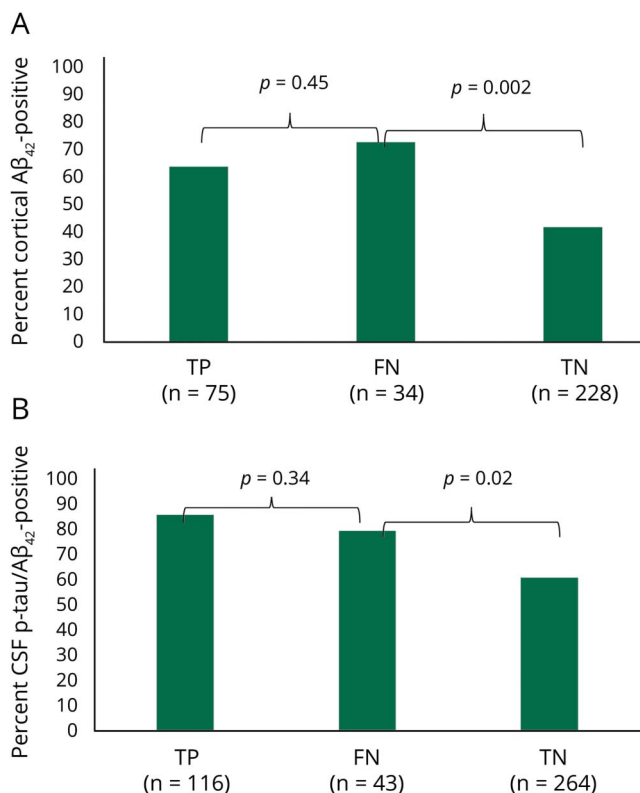
<sup>e</sup> Significant difference between FP and TN.

sex-specific) verbal memory norms/cut scores to define memory impairment in aMCI diagnostic criteria leads to aMCI being underdiagnosed in women and overdiagnosed in men. Furthermore, we sought to determine whether the use of sex-specific norms/cut scores would provide better diagnostic accuracy compared to the typical non-sex-specific norms/cut scores based on comparisons of AD-related markers.

As hypothesized, we found that when the new, sex-specific norms/cut scores were incorporated into aMCI diagnostic criteria, a subset of 10% of women previously classified as NC with typical diagnostic criteria were reclassified as having aMCI (FNs). The new, sex-specific diagnostic criteria also reclassified a subset of 10% of men previously classified as having aMCI with typical diagnostic criteria to NC (FPs). Our hypotheses were further supported because the addition of sex-specific verbal memory norms/cut scores to aMCI diagnostic criteria provided arguably more accurate diagnoses based on genetic and pathologic AD-related markers.

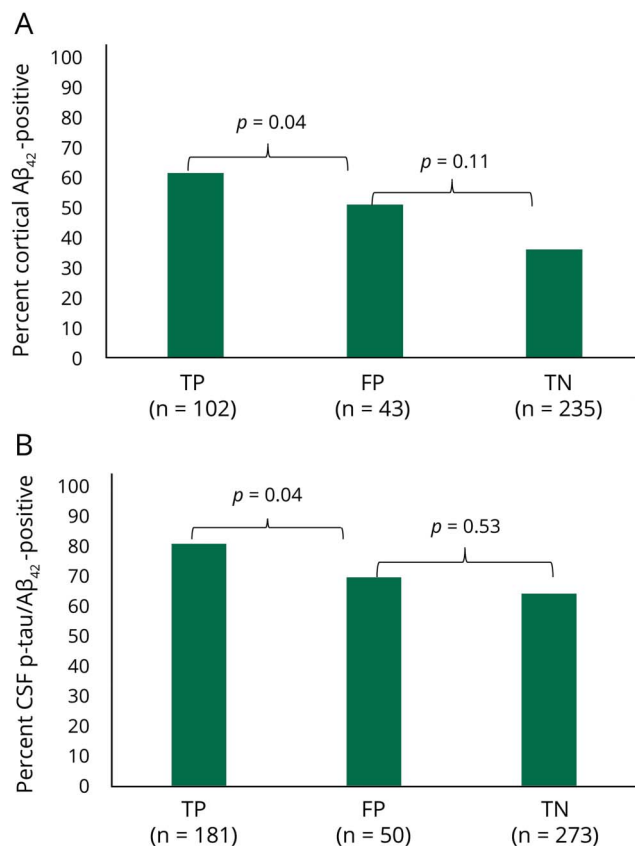
Specifically, the FN women had positivity rates for AD-specific pathologic markers and an APOE ε4 allelic frequency that were higher than true NC cases and nearly as high as true aMCI cases. Similarly, the FP men had positivity rates for AD-specific pathologic markers that were lower than TP men and comparably low to TN men. The APOE ε4 allelic frequency in FP men was intermediate between TN and TP men without significantly differing from either group. When considering the sex-specific diagnostic groupings as the true condition and the typical diagnostic groupings as the predicted condition, these results suggest that typical, non-sex-specific aMCI diagnostic criteria lead to a 20% diagnostic error rate in opposite directions across sexes. A particularly noteworthy finding was the almost 5-fold higher likelihood of having an APOE ε4 allele in FN vs TN women. This suggests that the cognitive reserve offered by the female advantage in verbal memory may be most salient for APOE ε4 carriers and that female APOE ε4 carriers are most at risk of a missed aMCI diagnosis. This raises the possibility that the vulnerability to developing aMCI

**Figure 1** Differences in Alzheimer disease-related biomarkers among TP, TN, and FN women



Percent of women who are positive for (A) cortical  $\beta$ -amyloid peptide ( $A\beta_{42}$ ) burden and (B) CSF hyperphosphorylated tau (p-tau)/ $A\beta_{42}$  and by diagnostic comparison category. The *p* values are from the logistic regression models that examined the effect of diagnostic comparison group on odds of biomarker positivity while covarying for age and education. FN = false negative; TN = true negative; TP = true positive.

**Figure 2** Differences in Alzheimer disease-related biomarkers among TP, TN, and FP men



Percent of men who are positive for (A) cortical amyloid- $\beta$  peptide ( $A\beta_{42}$ ) burden and (B) CSF hyperphosphorylated tau (p-tau)/ $A\beta_{42}$  and by diagnostic comparison category. The *p* values are from the logistic regression models that examined the effect of diagnostic comparison group on odds of biomarker positivity while covarying for age and education. FN = false negative; TN = true negative; TP = true positive.

in female *APOE*  $\epsilon 4$  carriers may have been underestimated in studies reporting no sex difference in the effect of the *APOE*  $\epsilon 4$  allele on AD risk.<sup>25–27</sup>

Notably, means and positivity rates of AD-associated markers in the FN women and FP men were intermediate between the TP and TN groups. This was expected considering that the FN and FP groups are closer to the borderline of AD trajectory vs normal cognitive aging when cognitive deficits are examined on a continuum. However, in support of our hypotheses, rates of FPs were typically closer to the TN vs TP group, and rates of FNs were closer to the TP vs TN group. The one exception was the exploratory analysis of mean CDR-SOB scores in both FN women and FP men, which was midway between the TN and TP groups.

Consistent with others who have reported a higher incidence rate of aMCI in men vs women,<sup>28,29</sup> we found a higher frequency of aMCI in men when applying the typical diagnostic criteria but not the sex-specific diagnostic criteria. This finding suggests that the higher rates of aMCI in men vs women may be an artifact of the lack of adjustment for the female

advantage in memory. With diagnostic criteria that use the typical, non-sex-specific cut scores, impairment in verbal memory may appear to be delayed in women, resulting in greater burden of disease during the aMCI stage and consequently a shorter duration in aMCI and a quicker transition to AD dementia. In addition, with typical diagnostic criteria, men falsely diagnosed with aMCI would remain in the aMCI group for longer. Combined, these effects may lead to an overrepresentation of men vs women in aMCI groups<sup>28,29</sup> despite a higher overall frequency of AD in women. Consistent with this view, our and others' previous work in ADNI suggests that, after the onset of clinically evident verbal memory impairment, women decline more rapidly.<sup>4–6,24</sup> In addition, the Einstein Aging Study found that women were less likely to transition from NC to MCI but more prone to transition from normal to dementia than men.<sup>30</sup>

Organizational and activational effects of hormones on brain structure and function contribute to the female advantage in verbal memory, which appears early in life and is sustained until late life.<sup>31</sup> The female advantage in verbal memory in

aMCI may stem from some combination of brain and cognitive reserve mechanisms. Brain reserve suggests that certain quantitative brain factors such as brain size or number of synapses confer a capacity to sustain brain damage and still perform normally. In addition, there is a critical threshold associated with brain reserve capacity such that clinical deficits begin to emerge once brain reserve is depleted past this threshold. Although men have larger whole-brain volumes,<sup>32–34</sup> women have a higher ratio of gray to white matter<sup>35</sup> and a larger hippocampal volume relative to brain size.<sup>36,37</sup> Therefore, greater depletion of brain resources may be necessary in women vs men for the emergence of hippocampal-mediated cognitive deficits. Cognitive reserve suggests that the brain actively compensates for brain damage by using neural networks more efficiently or activating alternative cognitive strategies or networks. Applied here, women may be more adept at invoking alternate brain networks or cognitive strategies during verbal memory tasks. Neuroimaging and brain lesion studies demonstrate sex differences in the neural correlates of language processing, whereby processing is more bilateral in women and left-lateralized in men.<sup>38,39</sup> Furthermore, women engage more than men in an advantageous encoding strategy of clustering stimuli in a verbal memory task by semantic and phonologic characteristics.<sup>40,41</sup>

These initial finding of a 20% difference in the diagnosis of aMCI vs NC with the use of sex-specific vs non-sex-specific diagnostic criteria warrants further investigation because of its clinical and research implications. If women with a considerable AD-related pathologic burden are inaccurately identified as cognitively normal, then pharmaceutical and cognitive interventions, as well as care, financial, and legal planning, are delayed. Similarly, false diagnosis of aMCI in men without substantial AD-related pathology can lead to undue stress and decreased quality of life, as well as unneeded medications. For research, diagnostic errors lower the reliability and replicability of AD findings from cohort studies and clinical trials. For example, the use of sex-specific diagnostic criteria as entry criteria in clinical trials could better ensure that men and women are at similar stages of disease, an assumption that may be violated with the use of typical diagnostic criteria. With typical diagnostic criteria, women might respond less to the therapeutic intervention than men because they are at a more advanced stage of disease. Conversely, men might not respond because a proportion have been falsely diagnosed with aMCI. These combined factors would result in a reduced estimate of the efficacy of the treatment for both men and women.

Our study has limitations. First, this study used cross-sectional comparisons of AD-associated markers. A longitudinal comparison of progression and reversion rates from aMCI among diagnostic comparison categories will more definitively determine the diagnostic accuracy of the sex-specific aMCI diagnostic criteria. This analysis is currently underway with ADNI data. Second, we evaluated the incorporation of sex-specific, verbal memory norms/cut scores specifically in the

Jak/Bondi diagnostic method for aMCI. Because other aMCI diagnostic methods, including the Petersen/Winblad method,<sup>42,43</sup> involve a criterion of memory impairment as typically determined by non-sex-specific verbal memory cut scores, it is likely that our findings generalize to other methods. However, future efforts are needed to confirm and to characterize the degree of diagnostic changes resulting from the use of sex-specific verbal memory norms/cut scores in other diagnostic methods. It is possible that the differences are stronger with the Jak/Bondi diagnostic approach vs others because it relies solely on neuropsychological performance; however, this neuropsychologically based approach has produced stronger associations with AD biomarkers and better prediction of progression to dementia compared to conventional MCI diagnostic criteria that rely on rating scales, global cognitive screens, subjective cognitive complaints, and impairment on a single cognitive test.<sup>13,20</sup> Lastly, ADNI is a convenience sample of mostly white and well-educated volunteers compared with the general US population, which limits generalizability of results. Study strengths include a large sample that is well characterized for AD-related neurocognitive deficits and biomarkers, the use of an empirically based statistical approach to defining aMCI, the use of an outside normative reference group, and associating the sex-specific diagnostic method to both CSF and neuroimaging AD-specific pathologic markers.

Our results suggest that the application of sex-specific cut scores for defining verbal memory impairment improves diagnostic accuracy in both sexes and may result in earlier detection of memory impairment in women and avoid false diagnoses in men. Replication of these results and additional longitudinal analyses are warranted in other aging cohort studies. If replicated, this work could have important implications for both clinical and research practices.

### Author contributions

E.E.S., P.M., A.B., R.B.L.: study concept. E.E.S., M.W.B., A.B., P.M., R.B.L., LR: study design. E.E.S., M.M.M., M.M.: data acquisition. E.E.S: statistical analysis. E.E.S., P.M., M.W.B., A.B., R.B.L, M.M.M, M.M.: data interpretation. E.E.S: initial manuscript preparation. All authors provided a critical review of manuscript for important intellectual content and contributed to and approved the final manuscript.

### Study funding

This work was supported by the NIH (grants AG049810, AG05131, K24 AG026431, U01 AG006786, RF1 AG55151, U54 AG44170). Data collection and sharing for this project were funded by the ADNI (NIH grant U01 AG024904) and Department of Defense ADNI (Department of Defense award W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, by the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc; Biogen; Bristol-Myers Squibb Co;

CereSpir, Inc; Cogstate; Eisai Inc; Elan Pharmaceuticals, Inc; Eli Lilly and Co; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc; Fujirebio; GE Healthcare; IXICO Ltd; Janssen Alzheimer Immunotherapy Research & Development, LLC; Johnson & Johnson Pharmaceutical Research & Development LLC; Lumosity; Lundbeck; Merck & Co, Inc; Meso Scale Diagnostics, LLC; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corp; Pfizer Inc; Piramal Imaging; Servier; Takeda Pharmaceutical Co; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. Dr. Lipton's time was supported by funding for the Einstein Aging Study; National Institute on Aging grants AG003949, AG026728, TL1RR000087, and T32-GM007288.

## Disclosure

E. Sundermann reports no disclosures relevant to the manuscript. P. Maki has received speaking honorarium from Mylan. A. Biegon reports no disclosures relevant to the manuscript. R. Lipton reports research support from the NIH: 2PO1 AG003949 (program director), 5U10 NS077308 (principal investigator [PI]), RO1 NS082432 (investigator), 1RF1 AG057531 (site PI), RF1 AG054548 (investigator), 1RO1 AG048642 (investigator), R56 AG057548 (investigator), K23 NS09610 (mentor), K23AG049466 (mentor), and 1K01AG054700 (mentor). He also receives support from the Migraine Research Foundation and the National Headache Foundation. He serves on the editorial board of *Neurology*, as senior advisor to *Headache*, and as associate editor to *Cephalalgia*. He has reviewed for the National Institute on Aging and National Institute of Neurological Disorders and Stroke; holds stock options in eNeura Therapeutics and Biohaven Holdings; and serves as consultant or advisory board member or has received honoraria from the American Academy of Neurology, Alder, Allergan, American Headache Society, Amgen, Autonomic Technologies, Avanir, Biohaven, Biovision, Boston Scientific, Dr. Reddy's, Electrocore, Eli Lilly, eNeura Therapeutics, GlaxoSmithKline, Merck, Pernix, Pfizer, Supernus, Teva, Trigemina, Vector, and Vedanta. He receives royalties from *Wolff's Headache*, 7th and 8th editions, Oxford Press University, Wiley, and Informa. M. Mielke and M. Machulda report no disclosures relevant to the manuscript. M. Bondi is paid royalties from Oxford University Press and serves as a consultant for Eisai and Novartis. Go to [Neurology.org/N](http://Neurology.org/N) for full disclosures.

## Publication history

Received by *Neurology* November 21, 2018. Accepted in final form June 5, 2019.

## References

- Blecker ML, Bolla-Wilson K, Agnew J, Meyers DA. Age-related sex differences in verbal memory. *J Clin Psychol* 1988;44:403–411.
- Kramer JH, Delis DC, Daniel MH. Sex differences in verbal learning. *J Clin Psychol* 1988;44:907–915.
- Barrett-Connor E, Kritz-Silverstein D. Gender differences in cognitive function with age: the Rancho Bernardo study. *J Am Geriatr Soc* 1999;47:159–164.
- Sundermann EE, Biegon A, Rubin LH, Lipton RB, Landau S, Maki PM. Better verbal memory in women than men in MCI despite similar levels of hippocampal atrophy. *Neurology* 2016a;86:1368–1376.
- Sundermann EE, Maki PM, Rubin LH, Lipton RB, Landau S, Biegon A. The female advantage in verbal memory: evidence for a sex-specific cognitive reserve. *Neurology* 2016b;87:1916–1924.
- Sundermann EE, Biegon A, Rubin LH, Lipton RB, Landau S, Maki PM. Does the female advantage in verbal memory contribute to underestimating Alzheimer's disease pathology? *J Alzheimers Dis* 2017;56:947–957.
- Stern Y, Gurland B, Tatemichi TK, Tang MX, Wilder D, Mayeux R. Influence of education and occupation on the incidence of Alzheimer's disease. *JAMA* 1994;271:1004–1010.
- Stern Y. What is cognitive reserve? Theory and research application of the reserve concept. *J Int Neuropsychol Soc* 2002;8:448–460.
- Stern Y, Zarahn E, Hilton HJ, Flynn J, DeLaPaz R, Rakitin B. Exploring the neural basis of cognitive reserve. *J Clin Exp Neuropsychol* 2003;25:691–701.
- Beinhoff U, Tumani H, Brettschneider J, Bittner D, Riepe MW. Gender-specificities in Alzheimer's disease and mild cognitive impairment. *J Neurol* 2008;255:117–122.
- Irvine K, Laws KR, Gale TM, Kondel TK. Greater cognitive deterioration in women than men with Alzheimer's disease: a meta-analysis. *J Clin Exp Neuropsychol* 2012;34:989–998.
- Petersen RC, Aisen PS, Beckett LA, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. *Neurology* 2010;74:201–209.
- Bondi MW, Edmonds EC, Jak AJ, et al. Neuropsychological criteria for mild cognitive impairment improves diagnostic precision, biomarker associations, and progression rates. *J Alzheimers Dis* 2014;42:275–289.
- Schmidt M. *Rey Auditory Verbal Learning Test: A Handbook*. Los Angeles: Western Psychological Services; 1996.
- Gale SD, Baxter L, Connor DJ, Herring A, Comer J. Sex differences on the Rey Auditory Verbal Learning Test and the Brief Visuospatial Memory Test-Revised in the elderly: normative data in 172 participants. *J Clin Exp Neuropsychol* 2007;29:561–567.
- Yesavage JA, Brink TL, Rose TL, et al. Development and validation of a geriatric depression screening scale: a preliminary report. *J Psychiatr Res* 1983;17:37–49.
- Shaw LM, Vanderstichele H, Knapik-Czajka M, et al. Cerebrospinal fluid biomarker signature in Alzheimer's Disease Neuroimaging Initiative subjects. *Ann Neurol* 2009;65:403–413.
- Joshi AD, Pontecorvo MJ, Clark CM, et al. Performance characteristics of amyloid PET with florbetapir F 18 in patients with Alzheimer's disease and cognitively normal subjects. *J Nucl Med* 2012;53:378–384.
- Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology* 1993;43:2412–2414.
- Jak AJ, Preis SR, Beiser AS, et al. Neuropsychological criteria for mild cognitive impairment and dementia risk in the Framingham Heart Study. *J Int Neuropsychol Soc* 2016;22:937–943.
- Shirk SD, Mitchell MB, Shaughnessy LW, et al. A webbased normative calculator for the Uniform Data Set (UDS) neuropsychological test battery. *Alzheimers Res Ther* 2011;3:32.
- Weintraub S, Salmon D, Mercaldo N, et al. The Alzheimer's Disease Centers' Uniform Data Set (UDS): the neuropsychologic test battery. *Alzheimer Dis Assoc Disord* 2009;23:91–101.
- Caldwell JZK, Berg JL, Cummings JL, Banks SJ. Alzheimer's Disease Neuroimaging Initiative. Moderating effects of sex on the impact of diagnosis and amyloid positivity on verbal memory and hippocampal volume. *Alzheimers Res Ther* 2017;9:72.
- Lin KA, Choudhury KR, Rathakrishnan BG, et al. Marked gender differences in progression of mild cognitive impairment over 8 years. *Alzheimer's Dement* 2015;1:103–110.
- Neu SC, Pa J, Kukull W, et al. Apolipoprotein E genotype and sex risk factors for Alzheimer disease: a meta-analysis. *JAMA Neurol* 2017;74:1178–1189.
- Combarros O, Leno C, Oterino A, et al. Gender effect on apolipoprotein E epsilon4 allele-associated risk for sporadic Alzheimer's disease. *Acta Neurol Scand* 1998;97:68–71.
- Beydoun MA, Boueiz A, Abougergi MS, et al. Sex differences in the association of the apolipoprotein E epsilon 4 allele with incidence of dementia, cognitive impairment, and decline. *Neurobiol Aging* 2012;33:720–731.
- Roberts RO, Geda YE, Knopman DS, et al. The incidence of MCI differs by subtype and is higher in men: the Mayo Clinic Study of Aging. *Neurology* 2012;78:342–351.
- Brodsky H, Heffernan M, Kochan NA, et al. Mild cognitive impairment in a community sample: the Sydney Memory and Ageing Study. *Alzheimers Dement* 2013;9:310–317.
- Song C, Kuo L, Derby CA, Lipton RB, Hall CB. Multi-stage transitional models with random effects and their application to the Einstein Aging Study. *Biom J* 2011;53:938–955.
- Andreano JM, Cahill L. Sex influences on the neurobiology of learning and memory. *Learn Mem* 2009;16:248–266.



32. Sherwin BB, Tulandi T. "Add-back" estrogen reverses cognitive deficits induced by a gonadotropin-releasing hormone agonist in women with leiomyomata uteri. *J Clin Endocrinol Metab* 1996;81:2545–2549.
33. Cosgrove KP, Mazure CM, Staley JK. Evolving knowledge of sex differences in brain structure, function, and chemistry. *Biol Psychiatry* 2007;62:847–855.
34. Peters M. Sex differences in human brain size and the general meaning of differences in brain size. *Can J Psychol* 1991;45:507–522.
35. Gur RC, Turetsky BI, Matsui M, et al. Sex differences in brain gray and white matter in healthy young adults: correlations with cognitive performance. *J Neurosci* 1999;19:4065–4072.
36. Filipek PA, Richelme C, Kennedy DN, Caviness VS Jr. The young adult human brain: an MRI-based morphometric analysis. *Cereb Cortex* 1994;4:344–360.
37. Szabo CA, Lancaster JL, Xiong J, Cook C, Fox P. MR imaging volumetry of subcortical structures and cerebellar hemispheres in normal persons. *Am J Neuroradiol* 2003;24:644–647.
38. Kimura D. Sex differences in cerebral organization for speech and praxic functions. *Can J Psychol* 1983;37:19–35.
39. Wallentin M. Putative sex differences in verbal abilities and language cortex: a critical review. *Brain Lang* 2009;108:175–183.
40. Berenbaum SA, Baxter L, Seidenberg M, Hermann B. Role of the hippocampus in sex differences in verbal memory: memory outcome following left anterior temporal lobectomy. *Neuropsychology* 1997;11:585–591.
41. Weiss EM, Ragland JD, Bressinger CM, Bilker WB, Deisenhammer EA, Delazer M. Sex differences in clustering and switching in verbal fluency tasks. *J Int Neuropsychol Soc* 2006;12:502–509.
42. Petersen RC, Morris JC. Mild cognitive impairment as a clinical entity and treatment target. *Arch Neurol* 2005;62:1160–1163.
43. Winblad B, Palmer K, Kivipelto M, et al. Mild cognitive impairment: beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *J Intern Med* 2004;256:240–246.