Inter-rater Reliability Between Expert and Nonexpert Physicians in the Designation of Amnestic MCI in the Community Setting

Rachelle S. Doody,[1] Steven H. Ferris,[2] Stephen Salloway,[3] Thomas M. Meuser,[4] Anita K. Murthy,[5] Mark Chunming Li,[6] Robert Goldman[6]

[1]Alzheimer's Disease and Memory Disorders Center, Baylor College of Medicine, Houston, Texas, USA; [2]Alzheimer's Disease Center, Silberstein Institute, New York University School of Medicine, New York, New York, USA; [3]Memory and Aging Program, Butler Hospital, Warren Alpert Medical School, Brown University, Providence, Rhode Island, USA; [4]Washington University, St Louis, Missouri, USA (now with the University of Missouri – St Louis); [5]Eisai Inc., Woodcliff Lake, New Jersey, USA; [6]Pfizer Inc, New York, New York, USA

STUDY OVERVIEW

INTRODUCTION

- Clinical (Petersen) criteria for amnestic mild cognitive impairment (aMCI) include a memory complaint (corroborated by an informant), evidence of objective memory impairment given the person's age, normal general cognition, preserved activities of daily living, and the absence of dementia.[1]
- Individuals with aMCI progress to clinically detectable Alzheimer's disease (AD) at a rate of about 10% to 15% per year.[2,3]
  - As many as 80% of those with aMCI progress to AD within 6 years of diagnosis.[4]
- The clinical criteria for aMCI have mainly been applied in research settings by specialists with substantial knowledge and expertise in the designation of MCI.
- It is unknown whether community-based, nonexpert primary care physicians (PCPs) can easily differentiate among older individuals with no cognitive impairment (NCI), individuals with aMCI, and those with mild AD.

OBJECTIVE

- Pilot study to determine whether community-based PCPs can reliably designate aMCI using practically operationalized Petersen criteria.

CONCLUSIONS

- **PCPs with minimal diagnostic training can differentiate aMCI from AD and NCI with fair to moderate accuracy.**
- **When nonexperts disagreed with experts, they tended to underrate the level of impairment compared with experts for both MCI and AD cases.**
- **Nonexperts showed lower sensitivity for designation of aMCI and AD than for NCI, suggesting that they had more difficulty differentiating these cases.**
- **More extensive training in the recognition of aMCI and AD in addition to further refinement and standardization of criteria for aMCI would likely improve the reliability of the designation in primary care.**

RESULTS

- 159 subjects were enrolled, including 79 in the aMCI group, 33 in the AD group, and 47 in the NCI group based on the experts' designations (Figure 1).
- Overall, 119 subjects (74.8%) completed the study:
  - 41 subjects (25.8%) discontinued prematurely, including 36 subjects who did not meet consensus committee designation and therefore did not proceed to Visit 2.
  - 123 subjects were enrolled with designations endorsed by the consensus committee, including 52 subjects in the aMCI group, 29 in the AD group, and 42 in the NCI group.
  - 4 subjects did not complete the study for reasons shown in Figure 1.
- The mean concordance between experts and consensus committee was 0.625, and mean concordance between experts and nonexperts was 0.615.
- Demographic characteristics of consensus committee–endorsed subjects are shown in Table 2.

Binary Outcome Analysis (aMCI vs not aMCI)

- Nonexperts correctly designated 31 of the 50 subjects with aMCI and 55 of the 69 subjects without aMCI (Table 3).
- The level of agreement between nonexperts and experts was fair to moderate ($\kappa = 0.4229$; 95% CI, 0.2577-0.5882).
  - The percent agreement was 72.3%, with a sensitivity of 62.0% and a specificity of 79.7%.

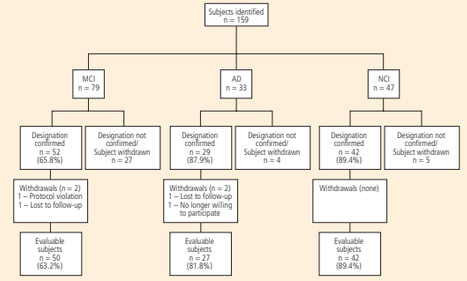

Figure 1. Subject Disposition.

Table 2. Subject Demographics.

	aMCI n = 50	AD n = 27	NCI n = 42
Mean age, years ± SD (range)	73.5 ± 8.6 (52-89)	78.0 ± 9.4 (55-92)	70.5 ± 9.3 (52-92)
Female subjects, n (%)	30 (60.0)	18 (66.7)	26 (61.9)
Race, n (%)			
Caucasian/White	40 (80.0)	22 (81.5)	39 (92.9)
African American/Black	4 (8.0)	2 (7.4)	0
Hispanic or Latino	3 (6.0)	1 (3.7)	3 (7.1)
Other	3 (6.0)	2 (7.4)	0
Education, mean years ± SD	14.4 ± 2.7	13.0 ± 2.5	15.1 ± 2.9

Three-Category Analysis (aMCI vs NCI vs AD)

- Table 4 summarizes the inter-rater agreement for all evaluable subjects with respect to all 3 categories of classification.
- A moderate agreement level was maintained ($\kappa = 0.5698$) with an overall percent agreement of 68.9%.
- Sensitivity was highest for the NCI category (88.1%), whereas specificity was highest for AD (95.6%).
  - Specificity was greater than sensitivity for the aMCI (specific % agreement 65.3%) and AD (specific % agreement 62.2%) designation categories.
- In cases where the expert designation was:
  - aMCI, nonexperts incorrectly designated 15 subjects with NCI and 4 subjects with AD.
  - AD, nonexperts incorrectly designated 9 subjects with aMCI and 4 subjects with NCI.
  - NCI, nonexperts incorrectly designated 5 subjects with aMCI; none were incorrectly designated with AD.

Log-Linear Model

- When the MMSE and mCDR were included as covariates in the log-linear model, the expert's mCDR was the only statistically significant factor ($\chi^2 = 4.76$, $df = 1$, $P = .029$).
  - This demonstrates a strong relationship between the expert's mCDR and the pattern of observed designations given by experts and nonexperts.

Analysis of Assessment Scores

- The diagnostic instruments used by experts and nonexperts were used consistently as evidenced by Pearson correlations.
- Correlations between experts and nonexperts were relatively high for both the MMSE (0.6875) and the mCDR (0.6885).

Table 3. Analysis of Inter-rater Agreement – Evaluable Subjects (Binary Outcome).

	Expert Designation				Overall
	aMCI	Not aMCI[1]	Total	Statistic	95% CI
Nonexpert designation					
aMCI	31 (26.1%)	14 (11.8%)	45 (37.8%)		
Not aMCI*	19 (16.0%)	55 (46.2%)	74 (62.2%)		
Total	50 (42.0%)	69 (58.0%)	119 (100.0%)		
Kappa				0.4229	(0.2577-0.5882)
% Agreement				72.3%	(71.9%-72.6%)
Sensitivity				62.0%	(48.5%-75.5%)
Specificity				79.7%	(70.2%-89.2%)

Note: Percentages for frequency counts were based on the number of subjects in the evaluable subjects analysis set.
*AD and not cognitively impaired subjects were pooled to comprise the not aMCI column.

Table 4. Analysis of Inter-rater Agreement – Evaluable Subjects (3-Category Outcome).

	Expert Designation					Overall
	aMCI	AD	NCI	Total	Statistic	95% CI
Nonexpert designation						
aMCI	31 (26.1%)	9 (7.6%)	5 (4.2%)	45 (37.8%)		
AD	4 (3.4%)	14 (11.8%)	0	18 (15.1%)		
NCI	15 (12.6%)	4 (3.4%)	37 (31.1%)	56 (47.1%)		
Total	50 (42.0%)	27 (22.7%)	42 (35.3%)	119 (100.0%)		
Specific % agreement	65.3%	62.2%	75.5%			
Sensitivity	62.0%	51.9%	88.1%			
Specificity	79.7%	95.6%	75.3%			
Weighted Kappa*					0.5698	(0.4496-0.6900)
Overall % agreement					68.9%	(68.6%-69.3%)

Note: Percentages for frequency counts were based on the number of subjects in the evaluable subjects analysis set.
The aMCI, AD, and NCI classifications were based on the expert designations.
*Calculated using Cicchetti-Allison (Cicchetti Allison, 1971) weights.

METHODS

Design

- A diagnostic reliability study.

Study Entry Criteria

- Ambulatory or ambulatory-aided subjects aged ≥ 50 years, with ≥ 9 years of education:
  - who resided in the community and had an informant who had daily contact with the subject of ≥ 10 waking hours per week
  - hearing and vision sufficient to comply with testing procedures
  - who were on allowable psychotropic medications at a stable dose for at least 1 month prior to the first visit, with no dose changes expected during the study.
- Exclusion criteria
  - Other medical, psychiatric, or neurologic disorders that could impair cognition
  - Drug or alcohol abuse or dependence within the last 5 years
- Criteria for aMCI, NCI, and AD are summarized in Table 1.

Protocol

- Cases were identified at Visit 1 by 1 of 14 physician experts (ie, physicians with extensive experience in the designation of aMCI) using laboratory results, a modified Clinical Dementia Rating[5] (mCDR; modified by RSD), the Mini-Mental State Examination (MMSE), and the Alzheimer's Disease Assessment Scale-cognitive subscale (ADAS-cog) Delayed Word Recall to designate aMCI and AD. The mCDR emphasized domains not already assessed by the MMSE and ADAS-cog Delayed Word Recall. The mCDR boxes were scored after all assessments had taken place.
- A consensus committee of 4 physician experts who were blind to the expert designation reviewed each case within 72 hours of Visit 1, and endorsement was established if the majority of the consensus committee agreed with the expert's designation.

Table 1. Criteria for aMCI, AD, and NCI.

aMCI	• Subject expressed a memory complaint that was recognized by the informant and represented a change from previous functioning • mCDR score of 0.5 and a "memory" box score of 0.5 or 1.0, with no more than 2 box scores other than "memory" rated as high as 1.0 and no box score rated > 1.0 • Scored at least 1.5 standard deviations below the norm on the ADAS-cog Delayed Word Recall • MMSE score ≥ 24 • General cognition and functional performance sufficiently preserved to preclude a diagnosis of possible or probable AD • Subjects and their informants had not been previously informed of diagnosis of aMCI before the first visit
AD	• MMSE score between 17 and 26 and a CDR score of 1.0 • Diagnostic evidence of probable AD per *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV)* and National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria either before or at the initial visit • Computer tomography or magnetic resonance imaging scan (taken within the past 12 months before the first visit) consistent with a diagnosis of AD without any clinically significant comorbid pathologies • Subjects or their informants had not been previously informed of a diagnosis of AD prior to the first visit
NCI	• Subjective memory complaint • MMSE score ≥ 24 and CDR score of 0 with no individual domain of greater than 0.5 • No previous history or diagnosis of MCI, AD, or other dementia

- Each endorsed case was randomly assigned to 1 of 27 nonexperts and was evaluated by the nonexpert within 2 weeks of the confirmed designation (Visit 2) using the same information and assessment instruments.
  - Nonexpert PCPs were provided the laboratory study results obtained by the expert, but performed their own history and physical examination, as well as their own cognitive testing.
  - Nonexpert PCPs regularly treated patients in the age range of the study but were not board certified in psychiatry or neurology, did not have a certificate of added qualification in geriatric medicine, had not previously participated as an investigator in a clinical research study, had no prior medical education in the area of MCI, and had no previous experience using specialized AD diagnostic tools.
  - Prior to the first case, the nonexperts received 3 hours of live or Web-based, standardized training in the differentiation of aMCI from AD and normal aging, accomplished through training in the assessment instruments used in the study as well as by using case studies.
- Following the visit with the nonexpert PCP, each subject returned to the original expert physician for communication of the designation and the customary clinical follow-up (Visit 3).
- Subjects, informants, and nonexperts were blind to expert designation.

Outcome Measures

- *Primary end point:* Inter-rater agreement (chance-corrected inter-rater reliability measure κ and its 95% confidence interval [CI]) between experts and nonexperts with respect to the designation of aMCI based on the binary outcome (aMCI/non-aMCI).
  - κ values interpreted as follows: ≤ 0.0, poor agreement; 0.0 to 0.2, slight; 0.2 to 0.4, fair; 0.4 to 0.6, moderate; 0.6 to 0.8, substantial; 0.8 to 1.0, almost perfect agreement.
  - Percent agreement (without chance correction) and its corresponding 95% CI.
    - The percent agreement calculated as the number of cases for which experts and nonexperts agreed as a percentage of all cases.
  - Sensitivity and specificity of designation.
    - Sensitivity defined as the number of correctly designated aMCI subjects divided by the sum of correctly identified aMCI subjects and aMCI subjects incorrectly designated by nonexperts.
    - Specificity defined as the number of subjects correctly designated as having not-aMCI divided by the sum of not-aMCI subjects and not-aMCI subjects incorrectly designated as having aMCI.
- *Secondary end point:* Inter-rater agreement with respect to the 3 categories (AD, aMCI, and NCI).
  - Overall κ (based on a weighted average) for the 3 designation categories with corresponding agreement rate, sensitivity, and specificity calculations.
  - Percent agreement without chance correction, sensitivity, and specificity.
- *Other end points:*
  - A statistical log-linear model, assessing the independence between raters, was employed to extend the overall agreement provided by the κ statistic.
  - Summary statistics by visit and expert site for the mCDR, the MMSE, and ADAS-cog Delayed Word Recall.
  - Summary of concordance rates of the consensus committee with respect to the expert designation.

REFERENCES

1. Petersen RC. *J Intern Med.* 2004;256:183-94.
2. Petersen RC, et al. *Arch Neurol.* 1999;56:303-8.
3. Petersen RC, et al. *N Engl J Med.* 2005;352:2379-88.
4. Petersen RC, et al. *Arch Neurol.* 2001;58:1985-92.
5. Morris JC. *Neurology* 1993 ;43 :2412-14.

Research supported by Eisai Inc. and Pfizer Inc. Editorial support was provided by W. Watkins, MS, of PAREXEL and was funded by Eisai Inc. and Pfizer Inc.