Variability in Annual Mini-Mental State Examination Score in Patients With Probable Alzheimer Disease

A Clinical Perspective of Data From the Consortium to Establish a Registry for Alzheimer’s Disease

Christopher M. Clark, MD; Lianne Sheppard, PhD; Gerda G. Fillenbaum, PhD; Douglas Galasko, MD; John C. Morris, MD; Elizabeth Koss, PhD; Richard Mohs, PhD; Albert Heyman, MD; and the CERAD Investigators

**Objective:** To determine the variability in annual Mini-Mental State Examination scores of patients with Alzheimer disease enrolled in the Consortium to Establish a Registry for Alzheimer’s Disease (CERAD).

**Patients:** A total of 372 patients with probable Alzheimer disease with 1 or more years of follow-up.

**Setting:** Twenty-one CERAD clinical sites throughout the United States.

**Results:** An average annual decline of 3.4 points in CERAD patients returning for longitudinal reassessments was close to the SD of the measurement error of 2.8 points for the Mini-Mental State Examination. There was wide variability in individual rates of decline. Even with 4 years of follow-up, 15.8% of the patients had no clinically meaningful decline in Mini-Mental State Examination score (defined as a change in initial score ≥3, ie, 1 SD of measurement error). Validity of measurements of the rate of change in Mini-Mental State Examination scores improved with longer observation intervals and was reliable for most patients when observations were separated by 3 or more years.

**Conclusions:** Although the Mini-Mental State Examination is a useful screening instrument to assess level of cognitive function, it has limited value in measuring the progression of Alzheimer disease in individual patients for periods less than 3 years because of a large measurement error and substantial variation in change in annual score.

Arch Neurol. 1999;56:857-862

---

The Mini-Mental State Examination (MMSE) of Folstein et al provides a brief evaluation of orientation, registration, attention, recall, language, and constructional praxis. Although it is insensitive in patients with mild cognitive impairment and lacks diagnostic specificity, it is easy to administer and assesses the major cognitive domains affected in Alzheimer disease (AD). It has high test-retest reliability values, ranging from 0.79 to 0.99.

In addition to its value as a screening test for dementia, the MMSE is often used to document cognitive changes over time in individual patients. This is an important clinical measurement, since progressive cognitive loss is a characteristic of neurodegenerative dementing illnesses. Information on the rate of change over time is valuable for assessing the results of therapeutic interventions, predicting the severity of cognitive decline, and planning for long-term health care.

In 9 published studies, the average annual change in MMSE score for a population of patients with dementia varied from 1.8 to 6.7. Reliability of the change measured increases in proportion to the length of observation. Teri et al evaluated the rates of change in patients with a history of alcoholism, agitated behavior, poor general health, or multiple disorders and noted that their MMSE scores declined up to 5 times faster than those who were free of these added burdens. In some instances, however, variations in the expected rate of change could not be explained, but may be due to true biological or clinical heterogeneity (signal) or measurement error (noise).

Although the MMSE is useful as a screening tool and as a marker of cognitive change in patient groups, to our knowledge, its utility as a measure of change in individual patients has not been determined. During follow-up of an individual patient, the standard required for a test to be a useful measure of progression is different than in the analysis of group data. Although score changes for patients can be reliable and meaningful when averaged over the entire group, the same
METHODS

The Consortium to Establish a Registry for Alzheimer's Disease (CERAD) is a multicenter study established in 1986 to develop standardized methods for the assessment of AD. From inception until December 1989, 589 patients 50 years and older with a clinical diagnosis of probable AD, based on modified National Institute of Neurological and Communicative Disorders and Stroke/Alzheimer's Disease and Related Disorders Association criteria, and 422 controls of comparable age were enrolled at 21 participating clinical sites.

Assessment protocol was previously reported.27 Patients and controls were given the CERAD clinical battery, including a brief physical and neurologic examination. For patients, stage of disease was assessed using the Clinical Dementia Rating (CDR) scale.28 Cognitive status was assessed by administration of the CERAD neuropsychology battery, which includes the MMSE, by trained and certified psychometric technicians.29,30 To the extent feasible, the clinical and neuropsychology batteries were repeated annually.

This study evaluated a subset of the CERAD cohort consisting of white community resident patients and controls. The final sample included 892 subjects: 491 with a clinical diagnosis of probable AD with a CDR score of 1 or greater severity and 402 cognitively healthy (CDR score of 0) controls. Patients with AD had mild-to-moderate cognitive impairment at entry, as determined by a score of 10 to 24 on the MMSE. Approximately half the controls were spouses of participating patients.

Baseline characteristics of patients and controls were compared using χ² tests for categorical measures and t tests for continuous measures.

RESULTS

One or more waves of annual follow-up information were available for 343 controls and 372 of the 491 eligible patients. Compared with patients who returned for annual reevaluation, the 119 who failed to return were more likely to be widowed (P<.02), to have a lower MMSE score (P<.005), and to be more severely demented, as assessed by the CDR (P<.05). There were no differences between the 2 groups in age or sex distribution, average level of education, or score on the Blessed Dementia Scale (Table 1).

Compared with controls, patients with follow-up data tended to be older, male, and widowed, with less education and lower MMSE scores. Approximately one fourth of all enrollees reported having hypertension, and less than a fifth reported having heart disease or thyroid disease. The only difference in comorbidity was the presence of informant-reported thyroid disease in 15.1% of controls compared with 11.2% of patients (P<.05).

Change in MMSE was determined by 2 methods. In the traditional clinical approach, change was defined as the difference between 2 scores separated by time. In the statistical model, change was estimated using a random slope and intercept model12 and fit using SAS PROC-MIXED.13 The model estimated an average entry MMSE score and change in MMSE score over time for the patients as a group (fixed effects), and then estimated subject-specific slope and intercept terms to reflect how each individual deviated from the group average (random effects). Additional covariates were included so that fixed effects could vary by demographic or prognostic factors, such as age, sex, comorbidity, and age of onset. We added terms to the fixed-effect model to determine whether comorbid conditions reported by the subjects or caregivers (hypertension, cardiac disease, thyroid disease) and age at which the symptoms of dementia were first noted (onset age) affected MMSE score at entry or its rate of change over time.

Continuous predictors, specifically, age at first evaluation and symptom duration, were included in the model and centered at a value close to the mean or median. This aids interpretation of the model, since every coefficient must be interpreted as conditional on other predictors in the model. We used age, centered at the average age of 71 years, for patients who were followed up annually. Likewise, we centered symptom duration at the median of 4 years. To ensure adequate control for the possibly confounding effects of age, we also included quadratic age terms as appropriate. Centering age also increased the stability of the model when polynomial terms were included because this technique reduced the correlation between the linear and quadratic terms.

Data from some subjects described in this article have been published previously.14 However, all analyses reported herein were performed independent of prior analyses.

Median length of follow-up was 2.4 years for patients and 4.1 years for controls. The proportion of each group returning annually decreased steadily over time (Figure 1).

Patients who did not return to clinic for follow-up evaluations were evaluated by caregivers via a telephone interview. Comparison of the 82 patients in cohort 2 who returned for a fourth-year follow-up with members of the cohort who did not showed no differences with respect to age, sex, marital status, education, or comorbid conditions at entry. However, patients who did not return for their fourth annual assessment had a greater degree of cognitive impairment at the time of enrollment (MMSE mean score, 17.5 vs 20.6; P<.001) and more functional impairment (mean Blessed Dementia Scale score, 4.3 vs 3.8; P<.05). The primary reasons patients did not return were entry into nursing homes (54.3%) or death (37.1%). The reasons for drop out by controls were less clear, but many stopped coming when their spouses became too severely impaired to return to the clinic.

We estimated the reliability of the MMSE score by comparing data obtained at entry with that obtained 1 month later in a subsample of 331 patients and 317 controls. Test-retest Pearson correlations were 0.87 and 0.67, respectively. The lower correlation for controls reflects a more limited range of scores due to a ceiling magnitude of change may not be reliable or interpretable when observed in one patient measured on 2 occasions. This study examines the long-term variability of periodic MMSE assessments to determine the utility of the MMSE to follow progression and help guide the clinical management of individual patients with probable AD.
effect. Changes in score among patients ranged from an increase of 7 points to a decrease of 8 points, with an average change of −0.5 (SD, 2.8). During the 1-month interval, 95% of retest MMSE scores for patients were within 6 points of their original scores. Analysis of the test-retest reliability data provided no evidence of a learning effect or measurable deterioration in cognitive function (Figure 2). Changes in scores among controls ranged from +4 points to −6 points (average change, +0.1; SD, 1.3). Ceiling effects reduced the change feasible for them.

Figure 3 demonstrates the distribution of MMSE score changes seen during sequential visits. Each dot represents 1 or more patients with the same score change. The MMSE score change indicated is the difference between 2 contiguous visits, adjusted for the number of years between visits.

Figure 4 shows the overall rate of change for each patient. This was estimated as the difference between the last and first MMSE scores, divided by the number of years between testing. Follow-up intervals ranged from 1 to 6 years. For patients with only 1 year of follow-up, MMSE score changes ranged from a decline of 23 points to an improvement of 7 points. As follow-up interval increased, the range of score changes narrowed, but even after 4 years, 13 (15.8%) of the 82 patients returning had no meaningful decline.

Table 1. Characteristics of the Study Population (Control and Patient Cohorts) at Entry

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Controls (≥1 y Follow-up) (n = 343)</th>
<th>Patient Cohort 1 (No Follow-up) (n = 119)</th>
<th>Patient Cohort 2 (≥1 y Follow-up) (n = 372)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at entry, mean ± SD, y</td>
<td>68.5 ± 7.7</td>
<td>71.8 ± 7.9</td>
<td>70.9 ± 7.8†</td>
</tr>
<tr>
<td>Male, %</td>
<td>33</td>
<td>40</td>
<td>49†</td>
</tr>
<tr>
<td>Education, %</td>
<td>0-11 y</td>
<td>10</td>
<td>25</td>
</tr>
<tr>
<td>12 y</td>
<td>28</td>
<td>35</td>
<td>32†</td>
</tr>
<tr>
<td>&gt;12 y</td>
<td>62</td>
<td>40</td>
<td>48†</td>
</tr>
<tr>
<td>Marital status, %</td>
<td>Married</td>
<td>66</td>
<td>66</td>
</tr>
<tr>
<td>Widowed</td>
<td>8</td>
<td>27</td>
<td>17†</td>
</tr>
<tr>
<td>Other</td>
<td>6</td>
<td>7</td>
<td>5†</td>
</tr>
<tr>
<td>MMSE score, %</td>
<td>&gt;20</td>
<td>100</td>
<td>38‡‡</td>
</tr>
<tr>
<td>16-19</td>
<td>0</td>
<td>21‡‡</td>
<td>25‡‡</td>
</tr>
<tr>
<td>10-15</td>
<td>0</td>
<td>23‡‡</td>
<td>21‡‡</td>
</tr>
<tr>
<td>&lt;10</td>
<td>0</td>
<td>18‡‡</td>
<td>6‡‡</td>
</tr>
<tr>
<td>BDS score, mean ± SD</td>
<td>NA</td>
<td>4.4 ± 2.3</td>
<td>4.1 ± 2.3</td>
</tr>
<tr>
<td>CDR score, %</td>
<td>0</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>51‡‡</td>
<td>58‡‡</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>41‡‡</td>
<td>39‡‡</td>
</tr>
<tr>
<td>&gt;2</td>
<td>0</td>
<td>8‡‡</td>
<td>3‡‡</td>
</tr>
</tbody>
</table>

*MMSE indicates Mini-Mental State Examination; BDS, Blessed Dementia Scale; CDR, Clinical Dementia Rating scale; and NA, not administered.
†Cohort 2 significantly different from the control group at P < .05 based on the χ² analysis for categorized variables and t test for continuous measures.
‡Significant difference between the 2 patient cohorts (P < .05) based on the χ² analysis for categorized variables.
To evaluate how often an MMSE score remains stable during 1 or more years in a patient with AD, we reviewed the data for the 82 patients with 4 years of follow-up to determine the proportion whose scores remained within 3 points of their initial score. At the first annual evaluation, 53.9% of the patients had no significant decline. By the fourth year of follow-up, 15.8% still had no significant decline (Table 2).

Because of the variable length of follow-up and number of MMSE scores available for each patient with AD, we used a random-effects linear regression model to estimate the average rate of change within patients. Model 1 represents the result of fitting MMSE scores on years in the study, sex, and age (with both linear and quadratic terms). The interaction between years in the study and the effects of sex and age (linear term only) are presented in Table 3.

The MMSE score declined by an average of 3.41 points per year ($P = .001$) at age 71 years. Age had a statistically significant impact, with a 0.06-point-per-year ($P < .001$) additional decline for every year greater than 71 or a corresponding reduction in the rate of decline for every year less than 71. Sex had no significant effect ($P = .43$).

Random-effect variances may be interpreted as those of the subject-specific slopes and intercepts. The random “years in study” effect was of particular interest. It had a variance of 3.65, indicating that 95% of the subjects should have had a rate of change in the range of $-7.15$ to $+0.33$ MMSE points per year ($P = -3.41 \pm 1.96/\sqrt{3.65}$), compared with the 95% confidence interval on the population average rate of change of $-4.23$ to $-2.59$. The correlation of 0.33 between the subject-specific slopes and intercepts indicates that subjects with higher initial MMSE scores tended to have an individual rate of decline that was somewhat flatter (ie, they have less decline) than individuals with lower initial MMSE scores. This relationship between initial MMSE score and rate of change was also observed in an earlier analysis of this same cohort.

Model 2 represents the final model after including comorbidity and age at symptom onset as fixed effects. Each variable was assessed both alone and as an interaction with years in the study. All fixed-effect terms not statistically significant at the $P = .1$ level were dropped. This final model included the estimated duration of dementia at entry, reported heart disease, and reported thyroid disease as intercept terms.

The rate of decline in MMSE score was somewhat greater ($-3.67$ points per year) in model 2 compared with model 1, and the SE was slightly lower, with simultaneous control for other predictors. With this model, 95% of the subjects had a rate of change between $-7.27$ and 0.08 (ie, $-3.67 \pm 1.96/\sqrt{3.37}$) when considering the subgroup of individuals 71 years old with no thyroid symptoms and a 4-year history of AD symptoms before entry into CERAD (for this estimate, age, disease duration, and presence of thyroid disease at entry were set to the median). Other explanatory variables that affected the rate of change were age, with a change of 0.05 points per year of age ($P = .001$); duration of disease, with an increase of 0.12 points per additional year ($P = .008$); and presence of thyroid disease, with a decrease of 0.90 points per year when present ($P = .03$). Variance of the subject-specific slopes was 3.37, essentially unchanged from model 1. The correlation between the subject-specific intercept and slope estimates was 0.35, indicating a tendency for patients who entered with higher MMSE scores to have less decline during the observation period.

### Table 2. Number of Patients With No Meaningful Decline in Follow-up Mini-Mental State Examination (MMSE) Score Compared With Their Initial Evaluation

<table>
<thead>
<tr>
<th>Annual Evaluation</th>
<th>No. of Observations</th>
<th>No. (%) With No Meaningful Decline (ie, ≤3) in MMSE Score</th>
<th>No. (%) With No Decline in MMSE Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>352</td>
<td>190 (54)</td>
<td>87 (25)</td>
</tr>
<tr>
<td>2</td>
<td>227</td>
<td>70 (31)</td>
<td>24 (11)</td>
</tr>
<tr>
<td>3</td>
<td>138</td>
<td>32 (23)</td>
<td>9 (7)</td>
</tr>
<tr>
<td>4</td>
<td>82</td>
<td>13 (16)</td>
<td>5 (6)</td>
</tr>
<tr>
<td>5</td>
<td>54</td>
<td>7 (13)</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>32</td>
<td>1 (3)</td>
<td>0</td>
</tr>
</tbody>
</table>

Although the MMSE is valid during a 1-month interval, suggesting that repeated observations over time are not confounded by a practice effect, our analysis of data collected during an observation period of up to 6 years indicated that, because of high measurement error and wide variability in individual rates of change, the MMSE had limited value as a method to mark cognitive changes in individuals with AD who are followed up for less than 3 years. Even after 4 years of follow-up, 15.8% of the remaining 82 patients had no clinically meaningful decline in MMSE score. Until factors affecting clinical or biological heterogeneity are identified, the MMSE will con-

__COMMENT__

The MMSE score declined by an average of 3.41 points per year ($P < .001$) at age 71 years. Age had a statistically significant impact, with a 0.06-point-per-year ($P < .001$) additional decline for every year greater than 71 or a corresponding reduction in the rate of decline for every year less than 71. Sex had no significant effect ($P = .43$).

Random-effect variances may be interpreted as those of the subject-specific slopes and intercepts. The random “years in study” effect was of particular interest. It had a variance of 3.65, indicating that 95% of the subjects should have had a rate of change in the range of $-7.15$ to $+0.33$ MMSE points per year ($P = -3.41 \pm 1.96/\sqrt{3.65}$), compared with the 95% confidence interval on the population average rate of change of $-4.23$ to $-2.59$. The correlation of 0.33 between the subject-specific slopes and intercepts indicates that subjects with higher initial MMSE scores tended to have an individual rate of decline that was somewhat flatter (ie, they have less decline) than individuals with lower initial MMSE scores. This relationship between initial MMSE score and rate of change was also observed in an earlier analysis of this same cohort.

Model 2 represents the final model after including comorbidity and age at symptom onset as fixed effects. Each variable was assessed both alone and as an interaction with years in the study. All fixed-effect terms not statistically significant at the $P = .1$ level were dropped. This final model included the estimated duration of dementia at entry, reported heart disease, and reported thyroid disease as intercept terms.

The rate of decline in MMSE score was somewhat greater ($-3.67$ points per year) in model 2 compared with model 1, and the SE was slightly lower, with simultaneous control for other predictors. With this model, 95% of the subjects had a rate of change between $-7.27$ and 0.08 (ie, $-3.67 \pm 1.96/\sqrt{3.37}$) when considering the subgroup of individuals 71 years old with no thyroid symptoms and a 4-year history of AD symptoms before entry into CERAD (for this estimate, age, disease duration, and presence of thyroid disease at entry were set to the median). Other explanatory variables that affected the rate of change were age, with a change of 0.05 points per year of age ($P = .001$); duration of disease, with an increase of 0.12 points per additional year ($P = .008$); and presence of thyroid disease, with a decrease of 0.90 points per year when present ($P = .03$). Variance of the subject-specific slopes was 3.37, essentially unchanged from model 1. The correlation between the subject-specific intercept and slope estimates was 0.35, indicating a tendency for patients who entered with higher MMSE scores to have less decline during the observation period.
Prevalence of comorbidity was low in this study population (hypertension, 22.0%; heart disease, 16.5%; and thyroid disease, 11.2%). Although the presence of thyroid disease had a statistically significant effect (model 2, P<.03), the comorbid conditions, age at time of testing, and sex did not have a clinically meaningful effect on the MMSE score rate of change or variability in change.

Despite the wide variation in individual annual score changes, as a group, the average MMSE score of patients with AD declined about 4 points per year (−3.8 when calculated by annualizing the difference between sequential visits; −3.7 when estimated using the random-effects linear regression model 2). These values are consistent with an earlier report of this cohort25 and other published studies.

This study had several limitations. Only white patients with AD and an initial MMSE score of 10 to 24 were studied. Information was restricted by the willingness and ability of patients to provide in-person follow-up data. This limits the ability to apply our conclusions to the larger universe of patients with AD, particularly those at the very high and very low ends of the MMSE score range. Nevertheless, the large number of carefully evaluated patients enrolled in this CERAD study and the multiple participating sites provide an excellent opportunity to explore questions that cannot be addressed easily using smaller, single-site cohorts.

At entry into CERAD, patients were at different stages of disease. We presumed that the rate of MMSE change was constant over time and minimally affected by the point of entry into CERAD (relative to age at onset of dementia symptoms and hypertension as a comorbidity. Ellipses indicate data not applicable.

This threshold makes it more likely that the difference reflects an actual change in cognitive abilities rather than testing imprecision. Second, the MMSE, when used as the sole measure of cognitive change for an individual patient, may not be a reliable measure for intervals of less than 3 years. Procedures that might enhance the utility of the MMSE as a measure of change are not obvious. Other structured instruments for measuring cognitive change in patients with dementia have similar limitations. From a practical standpoint, a clinician may want to continue to use the MMSE, particularly if the assessment can be done often enough to allow averaging of the changes that may be due to random variation. Although not assessed in this study, clinicians should always consider other factors that may affect individual MMSE scores, such as changes in the patient's living environment, medications, and the presence of problem behaviors. In addition, changes in a patient's functional abilities and the observations of a knowledgeable caregiver are impor-

<p>| Table 3. Linear Mixed Models to Assess Mini-Mental State Examination Change Over Time for Patients With Alzheimer Disease* |</p>
<table>
<thead>
<tr>
<th>Parameter (β)</th>
<th>SE (β)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years in study</td>
<td>−3.41</td>
<td>0.42</td>
</tr>
<tr>
<td>Sex by years in study</td>
<td>−0.22</td>
<td>0.26</td>
</tr>
<tr>
<td>Age of 71 y by years in study</td>
<td>0.06</td>
<td>0.02</td>
</tr>
<tr>
<td>Disease duration of 4 y by years in study</td>
<td>0.12</td>
<td>0.05</td>
</tr>
<tr>
<td>Thyroid disease by years in study</td>
<td>−0.90</td>
<td>0.40</td>
</tr>
</tbody>
</table>

* Variables found to be nonsignificant and therefore excluded from the final model included age at onset of dementia symptoms and hypertension as a comorbidity. Ellipses indicate data not applicable.
tant components in reaching an overall conclusion about changes in dementia severity. Third, the rate of change in MMSE score is not influenced to a clinically meaningful extent by the age at onset of dementia symptoms or the presence of relatively mild medical comorbidity. Furthermore, in this cohort, neither age at the time of assessment nor sex contributed to (or predicted) either the annual rate of MMSE score change or the variability associated with that annual change. Fourth, although the hallmark of AD is progressive cognitive impairment, the rate of change for individual patients varies, and it is not uncommon for patients to have a stable or even an improved score during a 1-year interval. Although we have found that the degree of variation narrows with increasing length of follow-up, this may simply reflect a self-censoring process associated with drop out as patients are admitted to a nursing home, become too severely impaired to return for follow-up evaluations, or die.

Thus, although the MMSE remains a convenient instrument for the rapid assessment of cognitive status, its ability to document changes over time in individual patients with AD is limited. The most important limitations are the high measurement error, which almost equals the average annual score change, and the highly variable individual annual change. As a consequence, improvement in the MMSE score during 1 year and/or stability in the score for periods up to 4 years may still be consistent with a clinical diagnosis of AD.

Accepted for publication November 30, 1998.

This study was partially supported by grants AG06790 and AG10124 from the National Institute on Aging, Bethesda, Md.

We thank the participating Consortium to Establish a Registry for Alzheimer’s Disease sites that contributed data to this study.

Corresponding author: Chris Clark, MD, Department of Neurology, University of Pennsylvania, 3615 Chestnut St, Philadelphia, PA 19104 (e-mail: ClarkC@mail.med.upenn.edu).

REFERENCES