A 7 Minute Neurocognitive Screening Battery Highly Sensitive to Alzheimer’s Disease

Paul R. Solomon, PhD; Aliina Hirschoff; Bridget Kelly; Mahri Relin; Michael Brush; Richard D. DeVeaux, PhD; William W. Pendlebury, MD

Objective: To determine the validity and reliability of a rapidly administered neurocognitive screening battery consisting of 4 brief tests (Enhanced Cued Recall, Temporal Orientation, Verbal Fluency, and Clock Drawing) to distinguish between patients with probable Alzheimer’s disease (AD) and healthy control subjects.

Subjects: Sixty successive referrals to the Memory Disorders Clinic at Southwestern Vermont Medical Center, Bennington, who were diagnosed as having probable AD and 60 community-dwelling volunteers of comparable age, sex distribution, and education.

Design: Interrater and test-retest reliability, intergroup comparisons between patients with AD and control subjects on the 4 individual tests, and determination of probability of dementia for patients with AD and control subjects using the entire battery of tests.

Setting: Outpatient care.

Main Outcome Measure: Comparison of the probability of dementia on the 7 Minute Screen with the criterion standard of clinical diagnosis established by examination and laboratory studies.

Secondary Outcome Measures: Test-retest and interrater reliability (correlation coefficients), time for administration.

Results: Mean time of administration was 7 minutes 42 seconds. Mean scores for patients with AD and control subjects on all 4 individual tests were significantly different (for each, \( P < .001 \)). When the 4 tests were combined in a logistic regression, the battery had a sensitivity of 100% and a specificity of 100%. A series of 1000 repeated random samples of 30 patients with AD and 30 control subjects taken from the overall sample of 60 patients with AD and 60 control subjects had a mean sensitivity of 92% and a mean specificity of 96%. The battery was equally sensitive to patients with mild AD as demonstrated by correctly classifying all 13 patients with AD using Mini-Mental State Examination scores of 24 or higher. Neither age nor education was a statistically significant factor when added as a covariate. Test-retest reliabilities for individual tests ranged from 0.83 to 0.93. Test-retest reliability for the entire battery was 0.91. Interrater reliability for the entire battery was 0.92.

Conclusions: The 7 Minute Screen appears highly sensitive to AD and may be useful in helping to make initial distinctions between patients experiencing cognitive changes related to the normal aging process and those experiencing cognitive deficits related to dementing disorders such as AD. It has reasonable interrater and test-retest reliability, can be administered in a brief period, and requires no clinical judgment and minimal training.

Arch Neurol. 1998;55:349-355

The increasing prevalence of Alzheimer’s disease (AD) and the emerging treatments for this disease suggest that there is an increasing need for accurate and easily administered screening instruments. Ideally, these instruments could be administered quickly by nurses, physician’s assistants, or other trained personnel with the goal of distinguishing between cognitive changes seen in the normal aging process and cognitive decline consistent with AD and other dementing diseases.

Primary care physicians do not diagnose AD in routine practice. One contributing factor to the underdiagnosis of AD is issues surrounding testing of mental status. Research indicates that primary care physicians either do not use mental status examinations (MSEs) or use MSEs that lack sensitivity. Increasingly abbreviated clinic visits may preclude the use of MSEs, even if cognitive deficits are suspected. Furthermore, currently available MSEs were not designed to screen for or to diagnose AD and should not be expected to possess those capabilities. For example, the Mini-Mental State Examination (MMSE), the most widely used MSE, was originally developed to evaluate elderly psychiatric pa-
SUBJECTS AND METHODS

HEALTHY CONTROL SUBJECTS

Sixty community-dwelling elderly individuals were recruited through newspaper advertisements and a local health maintenance organization. A medical history was obtained from each subject, including current medication use and incidence of head trauma, stroke, mental illness, mental retardation, or life-threatening illness during the last 5 years. None had a history of psychiatric or neurological disorder. None was taking antidepressant or other psychoactive medications. All claimed to be independent in activities of daily living, including shopping, transportation, and managing finances. Cognitive screening was accomplished in all subjects by administering the BIMC. A random sample of 30 subjects underwent more extensive neuropsychological testing, including the Wechsler Memory Scale–Revised (Logical Memory I and II, Visual Reproduction I and II) and the MMSE. Test procedures were completed after informed written consent had been obtained.

PATIENTS WITH AD

All patients with AD were seen at the Memory Disorders Clinic at Southwestern Vermont Medical Center, Bennington (an affiliated clinic of the Massachusetts Alzheimer’s Disease Research Center, Boston). The 60 patients represented consecutive referrals to the clinic and met the National Institute of Neurological Disorders and Stroke–Alzheimer’s Disease and Related Disorders Association diagnostic criteria for AD based on (1) neurological, medical, psychiatric, and/or social examinations; (2) standard laboratory studies; (3) computed tomographic scans; (4) neuropsychological evaluations; and (5) history from a caregiver indicating at least a 1-year history of progressive cognitive decline. Results from the 7 Minute Screen did not contribute to the diagnosis.

We selected a battery of tests based on recent developments in cognitive neuropsychology and behavioral neurology that have begun to elucidate the differences between cognitive changes in the normal aging process and those that accompany AD. There is general agreement that learning and memory tasks are the most useful diagnostic tests for the detection of AD. However, the 3-item recall on the MMSE has proved not to be sensitive enough. We chose the Enhanced Cued Recall task because it has been shown to distinguish between memory loss that accompanies the normal aging process and memory deficits present in patients with AD. Orientation for date and time is a highly reliable measure of cognitive status, although it lacks sensitivity for AD as it is commonly used in clinical practice. We used the Benton modification of this test to provide greater sensitivity to AD. Also, visuoconstructive deficits are common to patients with AD and constitute a nonverbal area of cognition. A number of investigators have shown that clock drawing is sensitive to AD. In addition, verbal fluency has been shown to discriminate between patients with AD and healthy elderly control subjects. The combination of these measures is brief, and taken together can be administered in less than 10 minutes.

7 MINUTE SCREEN

The 7 Minute Screen consists of 4 tests representing 4 cognitive areas typically compromised in AD: (1) memory, (2) verbal fluency, (3) visuospatial and visuoconstructive, and (4) orientation for time. Each test (or modification of the test) was selected because previous research showed that it had a high degree of sensitivity to AD, could be rapidly administered by personnel with little training, and could be scored objectively.

ENHANCED CUED RECALL

This test, which was initially described in a longer form by Buschke and colleagues, takes advantage of the finding that healthy elderly control subjects benefit from mnemonic strategies that facilitate the storage and retrieval of information (eg, reminder cues), whereas patients with AD show significantly less benefit from these strategies. This test requires the subject to recall 16 items. The items are presented 4 at a time on 4 individual cards. While looking at the card, the subject is asked to identify the picture on the card that best fits with the semantic cue given by the examiner (eg, examiner: “There is a piece of furniture on the card, what is it?” subject: “A desk.”). When all 4 objects have been successfully identified with the semantic cue, the card is removed and immediate recall testing is performed. The examiner provides the cue and the subject names the picture just observed (eg, “I just showed you a piece of furniture, what was it?” “A desk.”). If the subject misses 1 or more items, all items on the card are presented a second time. This repetition is intended to produce the same depth of encoding in all subjects. No card is repeated more than twice. If a subject makes an error on the second presentation, the examiner provides the correct response and continues to the next card. After all 4 cards are presented, the subject is distracted by reciting the months of the year backward. The subject is then asked to free recall as many of the pictures as possible. When the subject cannot recall any additional items and has been criticized both for its level of sensitivity and specificity as well as the influence that education has on performance. The Blessed Information Memory Concentration Test (BIMC) is also widely used but has been criticized for failing to sample a number of cognitive functions (eg, language and visuospatial abilities) commonly deficient in people with AD. Other batteries have been shown to be useful in diagnosing AD or monitoring the progression of the disease, but each of these is too long and/or requires too much clinical judgment to be useful as a screening tool in a primary care setting.

To address these issues, we sought to develop a mental status screening instrument that had the following characteristics: (1) it could be rapidly administered by a technician (since physician time for administering MSEs is greatly limited), (2) it required no clinical judgment and little training to use, (3) it took advantage of the evolving understanding of the cognitive differences between AD and the normal aging process, and (4) it was capable of reliably distinguishing AD from cognitive deficit changes associated with the normal aging process.
pictures, the examiner provides the category cues for the remaining items (cued recall). The score is the total number of items recalled in both free and cued recall (range, 0-16).

**CATEGORY FLUENCY**

The Category Fluency test requires that the subject generate as many words as possible in a fixed time period. This task has been shown to be sensitive to AD. Subjects were asked to generate exemplars from the semantic category “animals” in 1 minute. The total number of animals named produces the score.

**BENTON TEMPORAL ORIENTATION TEST**

The Benton Temporal Orientation Test uses a graduated scoring system to assess the patient’s knowledge of month, date, year, day of the week, and time of day. Unlike the orientation scales on the MMSE or BIMC, in this variation the degree of error is scored. Patients who miss the date by 1 day receive a minimal deduction compared with patients who miss the date by 15 days or the year by 3 years. The scoring system is (1) month—5 error points for each month off to a maximum of 30; (2) year—10 error points for each year off to a maximum of 60; (3) date—1 error point for each date off to a maximum of 15; (4) day of week—1 error point for each day to a maximum of 7; and (5) time—1 error point for each 30-minute deviation to a maximum of 5. The maximum total error score is 113. We should note that although scores on this test were not used to diagnose AD, temporal orientation is included in the MMSE and Alzheimer Disease Assessment Scale-Cognitive that were used in clinical diagnosis.

**CLOCK DRAWING**

Our version of this widely used test requires the patient to draw the face of a clock with all the numbers and set the hands to “twenty to four.” Our standard procedure is to provide the subject with a pen and blank sheet of paper and say, “I want you to draw a clock with all the numbers on it. Make it large.” After this is completed the examiner states, “Now draw the hands set at twenty to four.” There has been some variation in the literature regarding which time to use. Several studies suggest that “ten after eleven” is optimal whereas others favor “two forty-five,” or “twenty to four.” There are also several scoring methods. We have developed a simplified version of that used by Freedman et al. This system requires the examiner to record the presence of 7 attributes (eg, all 12 numbers are present). This can be accomplished as the patient completes the clock drawing.

**NEUROPSYCHOLOGICAL BATTERY**

The neuropsychological battery given to all patients seen at the Memory Disorders Clinic consists of the following: the North American Adult Reading Test, Wechsler Memory Scale–Revised (verbal and visual memory, immediate and delayed, and digit span), Delayed Word Recall, Boston Naming Test (15-item version), Wechsler Adult Intelligence Scale–Revised (Block Design), Alzheimer’s Disease Assessment Scale, MMSE, Geriatric Depression Scale, Cognitive Rating Scale, Global Deterioration Scale, and BIMC.

**STATISTICAL METHODS**

Associations between neuropsychological test scores and demographic variables, interrater reliability, and test-retest reliabilities were calculated using the Pearson correlation coefficient. Differences on individual tests were evaluated using the Student t test. Logistic regression analysis was used to determine the degree to which the battery discriminated between control subjects and patients with AD.

**RESULTS**

**DEMOGRAPHICS**

There were no differences in mean age or education (years of formal schooling) between patients with AD and control subjects (for each, t<1 and P>.05; Table 1). There was also no significant difference in the ratio of male-female subjects (χ²<1; P>.05; Table 1).

**NEUROPSYCHOLOGICAL BATTERY**

Table 2 summarizes the neuropsychological data from patients with AD and control subjects. In all cases, statistical analysis (t tests) indicated that the patients with AD performed significantly worse than control subjects.

**RELIABILITY**

Test-retest reliability was evaluated in 25 randomly selected patients with AD and 25 randomly selected control subjects by readministering the 7 Minute Screen 1 to 2 months after initial administration. All tests were highly reliable as was the overall score on the battery. Individual test-retest reliabilities were Category Fluency (r=0.83), Clock Drawing (r=0.84), Benton Orientation (r=0.93), and Enhanced Cued Recall (r=0.92). The overall test-retest reliability using the predicted probability of dementia from the logistic regression analysis of the battery was r=0.91.

Interrater reliability was accomplished by having 2 raters score the same testing session for 25 randomly selected patients with AD and 25 randomly selected control subjects. Interrater reliability for the entire battery was r=0.93. Because the only subscale in the battery that requires any judgment to score is Clock Drawing, we had 2 raters score all the clocks for control subjects and patients with AD. The interrater reliability was r=0.92.

**INDIVIDUAL TESTS**

Figure 1 shows the distribution of scores for each of the individual tests for patients with AD and control subjects. To determine if patients with AD and control subjects differed significantly on individual tests, we performed t tests between the mean test scores. Mean scores
and t values are shown in Table 3. As the table shows, there were significant differences on all measures (for each, P<.001).

LOGISTIC REGRESSION

To determine the degree to which the battery discriminated between control subjects and patients with AD, a logistic regression model

\[ \log \left( \frac{p_i}{1-p_i} \right) = \beta_0 + \beta_1 \cdot \text{ECR} + \ldots \]

was estimated using the 4 tests from the screening battery as the predictor variables. Because of the clear non-normality of the data from the battery, discriminant analysis was rejected as a possible alternative method. Specifically, the following model was estimated where ECR indicates Enhanced Cued Recall; CF, Category Fluency; BTO, Benton Temporal Orientation; and CD, Clock Drawing.

The response

\[ \log \left( \frac{p_i}{1-p_i} \right) \]

can be viewed as the natural logarithm of the odds in favor of having AD. For clarity, we then transformed this response back to the probability of having the disease, p. We classified someone as likely to have AD if \( p_i > 0.7 \) and unlikely to have AD if \( p_i < 0.3 \). We categorized patients with \( 0.3 \leq p_i \leq 0.7 \) (p indicating probability of AD) as requiring further testing (diagnosis deferred).

Estimating the model on the 120 patients (60 control subjects and 60 patients with AD) gave the following model, with SEs of the estimated coefficients given within parentheses:

\[ \log \left( \frac{p_i}{1-p_i} \right) = 35.59 - 1.303 \cdot \text{ECR} - 1.378 \cdot \text{CF} + 3.298 \cdot \text{BTO} - 0.838 \cdot \text{CD} + \varepsilon_i \]

(17.8) (1.11) (0.78) (1.37) (1.19)

This model resulted in a sensitivity of 100% and a specificity of 100% because the patients with AD all had \( p_i > 0.76 \) and all control subjects had \( p_i < 0.3 \). In fact, 38 of the 60 control subjects had \( p_i < 0.1 \), and 58 of the 60 patients with AD had \( p_i > 0.9 \). Using 0.1 and 0.9 as cutoff points resulted in a sensitivity and a specificity of 96%.

Figure 2 shows the distribution of probabilities for patients with AD and control subjects.

To test the robustness of the model, 2 different strategies were used. First, a sample of 30 patients with AD and 30 control subjects was chosen at random, and the logistic regression model fit to them was used to predict the status of the 30 patients with AD and the 30 control subjects not included in the samples. Repeating this random selection 1000 times, using \( r = 0.1 \) and \( r = 0.9 \) as cutoff points, and conservatively considering all “diagnosis-deferred cases” to be misclassifications, the 7 Minute Screen classified 92% of patients with AD correctly and 96% of control subjects correctly (Table 4). These numbers represent the percentages correctly predicted from only the 30 patients and 30 control subjects not included in the model building phase.

To determine how well the model would predict disease in patients with less severe AD, patients with AD who had MMSE scores below 21 were excluded, leaving 35 of 60 patients. The resulting sensitivity and specificity were 98% using 0.1 and 0.9 as cutoff values. A model using only the 13 patients with scores of 24 or higher performed equally well, with a sensitivity of 98% and a specificity of 100%.

We also calculated the positive and negative predictive values for different base rates of AD using the sensitivity and specificity rates from the 1000 random samples. Using hypothetical population base rates of 5%, 10%, 20%, and 50% produced positive predictive values between 54% and 95% and negative predictive values between 92% and 99% (Table 5).

Two of the predictors (Enhanced Cued Recall and Clock Drawing) in the model in the first equation were not statistically significant at the \( \alpha = 0.05 \) level using the \( t \) statistic. A stepwise procedure resulted in a model dropping the clock score with all 3 remaining predictors statistically significant. Moreover, in the 1000 randomly resampled runs, Enhanced Cued Recall was positive (the wrong sign) only 39 times, indicating an approximate \( P \) value of .04. Clock Drawing was positive 240 times. Given the a priori evidence of the diagnostic value of the Clock Drawing score and the behavior of the cross-validation exercise, we left in all 4 predictors for the final model.

Other variables may have predictive ability. A model using age, years of education, and sex of patient was also considered. Since sex showed no predictive ability, a second model with only age and years of education was estimated. Using a probability of disease of only 0.5 as a cutoff probability for diagnosis, the model had a sensitivity of only 57% (34/60) and a specificity of 65% (39/60). Probabilities for the patients with AD ranged from 0.29 to 0.83 (mean, 0.52), while the control subjects
ranged from 0.28 to 0.70 (mean, 0.47). Thus, while age and years of education have some ability to predict AD, it is weak. Additionally, when these terms were added to the model above as covariates, neither age, education, nor sex was statistically significant.

**ADMINISTRATION TIME**

The mean time to complete the test for all subjects was 7 minutes and 42 seconds (range, 6 minutes and 40 seconds to 11 minutes and 32 seconds). Patients with AD took somewhat longer than control subjects, with a mean of 7 minutes and 57 seconds (range, 6 minutes and 51 seconds to 11 minutes and 32 seconds). The control subjects took a mean of 7 minutes and 27 seconds (range, 6 minutes and 40 seconds to 9 minutes and 43 seconds).

**COMMENT**

This study provides initial reliability and validity data for an empirically derived brief screening battery to identify patients with AD. The data from the present study suggest that (1) the battery is highly sensitive, 100% in the initial sample and 92% in a series of random samples taken from the overall group, in detecting patients with AD; (2) the battery shows a high degree of specificity, 100% in the initial sample and 96% in a series of random samples taken from the overall group, in detecting healthy subjects; (3) there is a high degree of test-retest reliability and interrater reliability; (4) the ability of the battery to discriminate patients with AD from control subjects is not affected by age, sex, or education; and (5) the battery shows high sensitivity and specificity in patients with very mild, mild, and moderate disease.

The ability of the battery to discriminate between patients with AD and control subjects appears to be high. This, however, may not be surprising given the components of the test. Recent developments in the neuropsychological features of AD have identified a number of sensitive cognitive tests. The present battery attempted to take advantage of this research by truncating and combining these tests. For example, a longer version of the Enhanced Cued Recall test originally described by Grober and Buschke compared a group of patients with AD with a group of community-dwelling control subjects and found 94% sensitivity and 99% specificity. In this sample, however, the patients with dementia appeared more impaired than in the present sample and the control subjects were very high functioning (eg, mean intelligence quotient, 120.5). This test was developed to take advantage of the well-documented finding in cognitive psychology that either self-generated or experimenter-generated encoding retrieval cues facilitate subsequent recall in healthy aged subjects. The initial work by Grober and Buschke indicated that 3 trials of 16 items were necessary to discriminate between patients with AD.
and control subjects, but the data from the present study suggest that 1 trial is sufficient. The time savings of using only 1 trial makes the test appropriate for a short screening battery. Similarly, recent data suggest that verbal fluency, clock drawing, and orientation (using the Benton scoring method) all have a high degree of sensitivity and specificity for AD. Given the high degrees of individual sensitivity and specificity of these tests, it is not surprising that combining them identifies patients with AD.

Because the Enhanced Cued Recall test alone appears to yield relatively good sensitivity and specificity, it may be tempting to use this single test as a screening instrument. There are, however, 2 arguments against this. First, the diagnosis of AD requires cognitive deficits in at least 2 areas. Second, the Enhanced Cued Recall test is not as sensitive or specific as the entire battery. Using a cutoff point of 13/16 items correct, this test correctly classifies 59 of 60 patients with AD and 54 of 60 control subjects.

The sensitivity and specificity levels of the present battery seem to compare favorably with the most commonly used tests of mental status in elderly patients. A recent review of the MMSE found that the majority (about 75%) of studies using the 23/24 cutoff points reported sensitivity in the 80% to 90% range. However, this range decreases substantially (44%-68%) when the group with dementia is less impaired (ie, mean MMSE score >20). The present battery was well within the range of the MMSE for sensitivity and specificity for the entire sample, but appears to be more accurate for less impaired patients. For example, we were able to detect 13 of 13 patients with MMSE scores above 24, a group of patients who would generally be considered to be mildly demented. These patients would have been classified as within normal limits by the MMSE using the 23/24 cutoff criterion.

One ongoing criticism of the MMSE is its sensitivity to education. For example, Anthony et al found that for patients with less than an eighth-grade education, specificity decreased from 82% to 63%. Similarly, O’Connor et al reported that false negatives are much more likely to occur in patients with high educational levels. The present battery does not appear to be as sensitive to education. Adding years of education to the logistic regression analysis as a covariate did not significantly affect the predictions.

The MMSE has also been shown to be sensitive to age. Most of the age-related changes begin at 55 to 60 years, and dramatically accelerate over the age of 75 to 80 years. The present battery does not appear to show as much sensitivity to age. Adding age to the logistic regression analysis as a covariate did not significantly affect the predictions.

Reliability in the present battery appears reasonably high after a 1- to 2-month interval. Test-retest reliabilities for individual tests ranged between r=0.83 and r=0.93. Test-retest reliability for the entire battery was r=0.91 and interrater reliability for the entire battery was r=0.93. These reliabilities would seem to compare favorably with the MMSE where test-retest reliabilities have typically ranged from 0.80 to 0.95 for short intervals (minutes, hours, and days), but significantly lower for intervals of 1 to 2 months (r=0.50).

It is likely that the degree of sensitivity and specificity reported for this battery is higher than what would be seen in day-to-day practice. The patients with dementia in the present study were all referrals to a memory disorders clinic and were all suspected to have cognitive impairment. Ultimately, the utility of this screening test will be determined by how well it discriminates in primary care settings. Nevertheless, the data from our study...
suggest that the screening battery may have sufficient predictive power to be useful in this type of setting. Studies are currently under way to determine if this is the case.

A number of questions remain regarding the present battery. Although the battery appears sensitive to AD, we do not know if it will be sensitive to other dementing disorders. Similarly, we do not know if it will be useful in distinguishing AD from other dementing disorders. One distinction that may be important is whether the battery is to be useful as a screening instrument is its ability to distinguish between AD and clinical levels of depression. Based on what is known about some of the individual tests, there is reason to believe that this may be the case. For example, the type of memory test included in the battery has been useful in making this distinction.\(^2^3\)

In summary, the 7 Minute Screen appears to be highly sensitive and specific in its ability to discriminate between patients with AD and community-dwelling control subjects. It can be rapidly administered and scored and has adequate reliability. Although a number of questions remain regarding the battery's ability to distinguish between other forms of dementia, screening batteries such as this ultimately may be useful in helping make initial distinctions between cognitive changes that accompany the normal aging process and those that are suggestive of dementing disorders, such as AD. This may lead to earlier diagnosis and better management.

Accepted for publication June 11, 1997.

This work was supported by grant AG 05134-0852 from the National Institute on Aging, Bethesda, MD, a grant from the Howard Hughes Medical Foundation, Chevy Chase, MD, and a grant from the Essel Foundation, Mamaroneck, NY.

All the materials necessary for administering and scoring the 7 Minute Screen are available on request and at no charge from Paul R. Solomon, PhD. Distribution of these materials is supported by Janssen Pharmaceutica Research, Titusville, NJ.

We are grateful to Felicity Adams, Vivian Calvo, and MaryEllen Grocias for their help in collecting the data, to Judy Kessler for preparing the figures, and to Anthony Giuliano, PhD, David Knopman, MD, and Martin Samuels, MD, for their helpful comments on an earlier version of the manuscript.

Reprints: Paul R. Solomon, PhD, Bronfman Science Center, Williams College, Williamstown, MA 01267 (e-mail: psolomon@williams.edu).

REFERENCES