

Screening for dementia with the Memory Impairment Screen

H. Buschke, MD; G. Kuslansky, PhD; M. Katz, MPH; W.F. Stewart, PhD;
M.J. Sliwinski, PhD; H.M. Eckholdt, PhD; and R.B. Lipton, MD

Article abstract—*Objectives:* To validate a sensitive and specific screening test for AD and other dementias, assess its reliability and discriminative validity, and present normative data for its use in various applied settings. *Background:* To improve discrimination in screening for AD and dementia, we developed the Memory Impairment Screen (MIS), a 4-minute, four-item, delayed free- and cued-recall test of memory impairment. The MIS uses controlled learning to ensure attention, induce specific semantic processing, and optimize encoding specificity to improve detection of dementia. *Methods:* Equivalent forms of the MIS were given at the beginning and end of the testing session to assess alternate forms reliability. Discriminative validity was assessed in a criterion sample of 483 aged individuals, 50 of whom had dementia according to Diagnostic and Statistical Manual of Mental Disorders (3rd ed., revised) criteria. *Results:* The MIS had good alternate forms reliability, high construct validity for memory impairment, and good discriminative validity in terms of sensitivity, specificity, and positive predictive value. We present normative data for use in settings with different base rates (prevalences) of AD and dementia. *Conclusion:* The MIS provides efficient, reliable, and valid screening for AD and other dementias.

NEUROLOGY 1999;52:231–238

There are urgent scientific and public health reasons to develop effective methods of screening for AD and other dementing disorders, especially in the early clinical stages.¹ Individuals with early disease may not consult physicians or may not mention their cognitive complaints.² They may be unaware of their decline or believe that their memory problems are part of the normal aging process.³ Even if they report symptoms, dementia often goes undetected and untreated in primary care settings.⁴

Dementia is a clinical syndrome with many remediable causes. Two pharmacologic treatments for AD have been approved,^{5–7} and other pharmacologic treatments are being developed.^{8,9} Early detection is important because treatment may be more effective early in disease when useful cognitive function can still be preserved.^{10,11} AD may soon meet criteria for community-based “tertiary prevention” programs.¹ Other treatable causes of dementia include subdural hematoma, vitamin B₁₂ deficiency, hypothyroidism,

other metabolic causes, and infections.^{12,13} Effective recognition of the dementia syndrome is an important step toward establishing and addressing a specific etiology.

Despite several decades of research, the 1996 US Agency for Health Care Policy and Research Clinical Practice Guideline on recognition and initial assessment of Alzheimer’s disease and related dementias¹⁴ indicates that there is no evidence to recommend one screening test over another. Because memory impairment is often the earliest feature of AD, a very brief test for memory impairment with sufficient sensitivity, specificity, and positive predictive value (PPV) should serve as a good screening test for AD and other dementias with memory impairment.

Very brief memory tests that assess (delayed) recall of three or four words are used in clinical practice to screen for memory impairment¹⁵ and are included in mental status assessments.^{16–22} However, such very brief three- or four-word tests for memory

See also page 224

From The Saul R. Korey Department of Neurology (Drs. Buschke and Kuslansky, M. Katz, and Drs. Sliwinski, Eckholdt, and Lipton), Einstein Aging Study (Drs. Buschke and Kuslansky, M. Katz, and Drs. Sliwinski, Eckholdt, and Lipton), Rose F. Kennedy Center for Mental Retardation and Human Development (Drs. Buschke, Kuslansky, Sliwinski, and Eckholdt), and Department of Epidemiology and Social Medicine (Dr. Lipton), Albert Einstein College of Medicine, Bronx, NY; Department of Epidemiology (Dr. Stewart), School of Public Health, The Johns Hopkins University, Baltimore, MD; and Innovative Medical Research (Drs. Stewart and Lipton), Towson, MD.

Supported by NIA grants AG03949 and AG12448, and NICHD grant HD-01799.

Presented in part at the annual meeting of the American Neurological Association; San Diego, CA; September 1997.

Received March 16, 1998. Accepted in final form September 12, 1998.

Address correspondence and reprint requests to Dr. Herman Buschke, Department of Neurology, Kennedy 912, Albert Einstein College of Medicine, Bronx, NY 10461.

impairment have a high rate of false-negative errors (low sensitivity) or false-positive errors (low specificity) and may yield different results depending on the words used.²³⁻²⁵ The 10-word Delayed Word Recall Test²⁶ discriminates between nondemented elderly individuals and mild AD patients with high sensitivity (89%) and specificity (98%) but takes longer to administer because the list is presented twice, and each time the individual must make up a sentence for each word. A very brief, well-standardized, accurate screening test for memory impairment and dementia might find broader use in clinical practice.

We previously demonstrated that the Double Memory Test (DMT) has high discriminative validity for AD and related dementias in comparison with standard memory tests that do not optimize encoding specificity.^{27,28} The advantage of the DMT, relative to conventional memory tests, is greatest for the patients with mild disease; the DMT has a sensitivity of 93% and a specificity of 99% for the diagnosis of mild dementia.²⁸ Although these unusually high sensitivities and specificities suggest that the DMT may be useful in detecting early dementia, it is not suitable as a *screening* test. It has 128 items, must be administered by a trained tester in two parts, and requires approximately 20 minutes to complete. The multi-trial Free and Cued Selective Reminding Test (FCSRT),²⁹ a forerunner of the DMT, also uses controlled learning and cued recall but, because of its length, is also not suitable for screening.

The Memory Impairment Screen (MIS) is a brief, four-item delayed free- and cued-recall memory *impairment* test that also uses controlled learning and cued recall to optimize encoding specificity. Controlled learning²⁹ requires the individual to search for and identify a to-be-remembered item in response to its category cue. The category cues are also used to elicit retrieval by cued recall. Using the same cues at acquisition and retrieval coordinates acquisition and retrieval, optimizing encoding specificity because "specific encoding operations performed on what is perceived determine what is stored and what is stored determines what retrieval cues are effective in providing access to what is stored."³⁰ We postulated that improving retrieval by these procedures would provide a screening test with improved discriminative validity. To test the potential of the MIS as a screening test for AD and dementia, we assessed its reliability, construct validity with respect to a standard memory test, and discriminative validity as a screening test for dementia and AD.

Methods. *Participants.* All participants gave informed consent as specified by the Committee on Clinical Investigations at Albert Einstein College of Medicine. Participants were 483 community-residing volunteers, 50 of whom had dementia, who participated in the Einstein Aging Study, a longitudinal study of dementia and normal aging. The sample included 286 community volunteers recruited from senior centers and physician practices. In addition, 197 individuals were systematically sampled from

Health Care Financing Administration (HCFA) lists for the area adjacent to our clinical research center. According to the HCFA, their lists include 97% of individuals over the age of 65. The eligibility criteria for inclusion in this study were age 65 and over, ambulatory, and adequate ability to understand and speak English. In comparison with the entire HCFA sample for the relevant segment of the Bronx, our study sample was slightly older but otherwise similar to the community in terms of sex (36% male), racial distribution (80.7% white, 16.2% African American, and 2.7% other), and education (16.6% had fewer than 9 years of education).

These participants received a neuropsychological test battery as well as medical, epidemiologic, social, and behavioral questions every 18 months. A neurologic evaluation was performed on those participants who had Blessed Information Memory and Concentration (BIMC)¹⁶ test error scores of greater than 8, a total recall score of 44 or less on the FCSRT,^{29,31} or who were suspected of having dementia based on self-report, informant report, or observations of the testers. Additional diagnostic testing, ordered by the neurologist, included neuroimaging, blood tests, and other evaluations as needed. A diagnosis of dementia was made according to Diagnostic and Statistical Manual of Mental Disorders (3rd ed., revised) (DSM-III-R) criteria,³² and a diagnosis of AD was based on National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria.³³ Of the 50 participants with DSM-III-R dementia (yielding a sample base rate of 10.4%), 39 met criteria for AD. The other 433 participants were classified as nondemented.

Procedure. The participants received the MIS as part of a neuropsychological test battery that included the Wechsler Adult Intelligence Scale-Revised,³⁴ the Logical Memory I and Verbal Paired Associates I subtests from the Wechsler Memory Scale-Revised,³⁵ the FCSRT,^{29,31,36} the BIMC test,¹⁶ and the Zung Depression Scale.³⁷ All participants scoring in the depressed range on the Zung Depression Scale (≥ 50) were clinically assessed by a study psychiatrist using the Structured Clinical Interview for the DSM-III-R.³⁸ Alternate forms of the MIS were administered at the beginning and the end of the neuropsychological test session in counterbalanced order. Participants were tested individually and told before presentation of the test items that they should remember the items so that they could recall them later. Each participant was presented with an $8\frac{1}{2} \times 11$ inch sheet with the four MIS items to be recalled printed in 24-point uppercase letters. Each item belonged to a different category. The individual was asked to read the items aloud and then asked to identify and name each item (e.g., "potato") when the tester said its category cue (e.g., "vegetable"). The sheet was then removed. After a nonsemantic interference task (repeated counting from 1 to 20 and back) lasting approximately 2 to 3 minutes, the individual was asked for free recall of the four items in any order. The category cues were then presented to elicit cued recall of only those items that were not retrieved by free recall. The number of items retrieved by free recall and the number retrieved by cued recall were recorded.

Scoring. The MIS score is calculated as $[2 \times (\text{free recall}) + (\text{cued recall})]$. Simple free recall provides a single

score—the number of items retrieved by free recall. Cued recall in the MIS provides a second score—the number of items retrieved by cued recall. If cued recall of all items were tested, a combined free- and cued-recall score would be obtained by adding the total cued-recall score to the total free-recall score. However, in the MIS, cued recall is tested only for those items that were not retrieved by free recall. Because items retrieved by free recall also are retrieved by cued recall, the equivalent of a total cued-recall score can be obtained by adding free recall (i.e., those items that would also have been retrieved by cued recall) to the actual cued recall obtained by testing only those items that were not retrieved by free recall. Adding this equivalent total cued-recall score to the free-recall score results in an MIS score equal to $[2 \times (\text{free recall})]$ plus (obtained) cued recall. Adding cued recall to free recall also increases the range of scores (0 to 8), which may improve discriminative validity.

Statistical methods. The MIS was analyzed for alternate forms reliability using the intraclass correlation. Internal consistency was determined using Cronbach's coefficient alpha.

Construct validity of the MIS for detecting memory impairment was assessed using the chance corrected kappa statistic to measure the concordance between memory impairment on the MIS and impairment as defined by performance on the FCSRT. The MIS is intended to screen for memory impairment and is not intended as an index of memory ability. The FCSRT cut-score was selected based on its ability to distinguish individuals with and without dementia in a criterion sample.³¹

Discriminative validity of the MIS was assessed by calculating the sensitivity and specificity of the MIS for detecting dementia and for detecting AD for various MIS cut-scores. The base rate of dementia in our sample was 10.4% (50 of 483). Of the 50 demented participants, 39 (78%) met criteria for possible or probable AD.

Although normative data are often presented in the form of percentiles or means and SDs, we present norms in the form of the probability of dementia (or AD) given different MIS cut-scores. To calculate the probability of dementia, one must know the test sensitivity and specificity at each cut-score as well as the base rate of dementia.^{39,40} Because the base rate of dementia varies with setting as well as age, we present normative data that cover a broad range of dementia base rates. Normative data are presented for various dementia base rates as the probability of dementia or AD (i.e., the PPV) using each MIS score as a cutoff.

Probability of dementia or AD was calculated according to Bayes' theorem.³⁹ The formula for PPV yields the proportion of individuals who screen positive at a given cut-score who actually have dementia (for a specific prevalence rate). Note that the PPV also represents the probability of dementia for individuals with scores falling at or below a given score on the MIS in a sample with a given base rate of dementia. PPV is a useful index of the efficiency of a screening test in different applied settings that have different base rates of dementia. The negative predictive value (NPV) is the proportion of individuals who screen negative and do not have disease.

Results. Demographic and neuropsychological characteristics of the study sample are summarized in table 1. The

Table 1 Demographic and neuropsychological characteristics of participant groups

| | Dementia, including AD | AD only | Nondementia |
|-----------------------|---------------------------|--------------|--------------|
| Sample size | 50 | 39 | 433 |
| Age, y | 81.4 (7.0) | 81.1 (7.3) | 79.3 (6.1) |
| Education, y | 11.0 (3.6) | 11.3 (3.5) | 12.2 (3.2) |
| Sex, % male | 34 | 33 | 36 |
| WAIS-R verbal IQ | 86.5 (14.3) | 87.03 (15.1) | 105.8 (13.2) |
| Zung Depression Scale | 52.3 (11.5) | 52.1 (11.8) | 46.2 (10.7) |
| BIMC errors | 14.7 (5.7) | 15.1 (5.5) | 2.8 (2.6) |
| MIS score | 2.5 (2.3) | 2.1 (1.9) | 7.2 (1.2) |

Means (SD) are presented except for sample size and sex.

BIMC = Blessed Information Memory Concentration test; WAIS-R = Wechsler Adult Intelligence Scale-Revised; MIS = Memory Impairment Screen.

dementia group was slightly older and had slightly less education, although the gender composition of the two groups was similar. The dementia group scored approximately one-half SD unit higher ($p < 0.001$) on the Zung Depression Scale than did the nondementia group. Zung Depression Scale scores were not correlated with the MIS scores for the demented (Spearman = 0.09, $p = 0.53$) or for the nondemented groups (Spearman = 0.002, $p = 0.96$). Based on clinical interviews by a psychiatrist of individuals with Zung Depression Scale scores of ≥ 50 , seven participants met DSM-III-R criteria for major depressive disorder, six of whom were in the nondemented group.

Reliability. Alternate forms reliability for the MIS was assessed by administering one of two forms to a subset ($n = 429$) of clinically diagnosed demented and nondemented participants at the beginning and end of a neuropsychological evaluation. The intraclass correlation was high, 0.69, indicating strong concordance between the two forms of the MIS and a high level of alternate forms reliability. Internal consistency also was good for both forms—the coefficient alpha was 0.67 for both forms.

Validity. **Construct validity.** We assessed the MIS as a screening test for memory impairment as determined by the 16-item multi-trial FCSRT.³¹ Memory impairment is identified by the FCSRT by a total free- plus cued-recall score over all three trials of 44 or less. Classification of memory impairment by the MIS (as defined by a score of 4 or less) was strongly associated with memory impairment as defined by the gold standard FCSRT (kappa = 0.62).

Discriminative validity. We assessed the discriminative validity of the MIS first as a screening test for dementia in general and then specifically for clinically diagnosed AD. The trade-off between sensitivity and specificity of the MIS as a screening test for dementia and AD as the cut-score varies is shown by the receiver operating characteristic (ROC) curves^{41,42} in figure 1 (all dementia, including AD, versus nondemented) and figure 2 (only AD versus nondemented). The area under the ROC curve is regarded as a useful index of a diagnostic test's performance. The area under the ROC curve for discrimination of dementia (including AD) is 0.94. The area under the ROC curve for discrimination of AD is even larger, 0.97. These figures

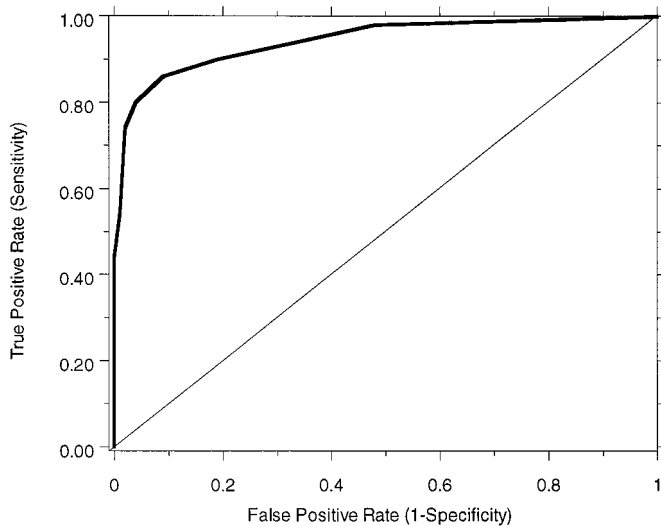


Figure 1. Receiver operating characteristics of the Memory Impairment Screen as a screening test for dementia (including AD).

graphically indicate that the specificity of the MIS is consistently high for a large range of sensitivities. The influence of the sensitivity-specificity trade-off on the usefulness of the MIS as a screening test for AD and dementia is examined in the following section on normative data.

Normative data. Tables 2 and 3 show the sensitivity, specificity, and PPV (i.e., probability of dementia or AD) for each cut-score on the MIS for different base rates of dementia. Performance at or below the cut-score is taken as evidence of dementia. Sensitivity measures the proportion of those *with* dementia who are correctly identified as demented. Specificity measures the proportion of those *without* dementia who are correctly identified as nondemented. PPV is the proportion of those who meet a given cut-score (i.e., screen positive) and actually have dementia.

Table 2 displays the sensitivity and specificity for detecting any type of dementia using each level of MIS score as a cutoff. Although the selection of an optimal cut-score will depend on the intended clinical or research application

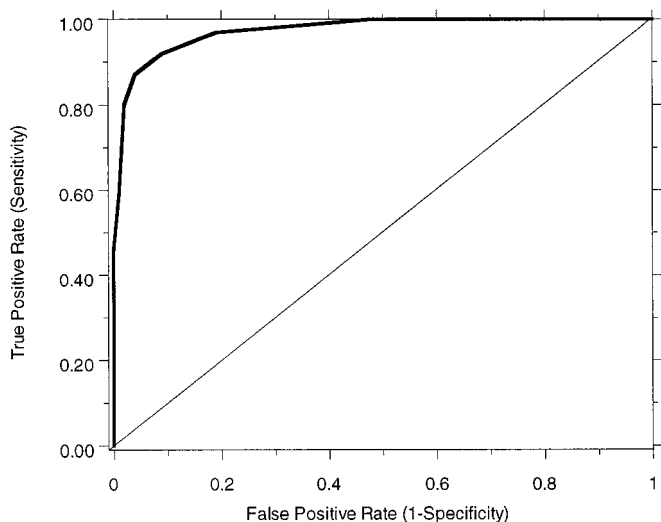


Figure 2. Receiver operating characteristics of the Memory Impairment Screen as a screening test for AD.

Table 2 Dementia sensitivity and specificity of each MIS score with corresponding probabilities of dementia (PPV) at different base rates

| MIS | All dementia | | Probability of dementia (PPV) at different base rates | | | |
|-----|--------------|-------------|---|------|------|------|
| | Sensitivity | Specificity | 5% | 10% | 15% | 20% |
| 0 | 0.24 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 1 | 0.44 | 1.00 | 0.91 | 0.95 | 0.97 | 0.98 |
| 2 | 0.54 | 0.99 | 0.75 | 0.87 | 0.91 | 0.94 |
| 3 | 0.74 | 0.98 | 0.63 | 0.78 | 0.85 | 0.89 |
| 4 | 0.80 | 0.96 | 0.52 | 0.69 | 0.78 | 0.84 |
| 5 | 0.86 | 0.91 | 0.35 | 0.53 | 0.64 | 0.72 |
| 6 | 0.90 | 0.81 | 0.20 | 0.34 | 0.45 | 0.54 |
| 7 | 0.98 | 0.52 | 0.10 | 0.18 | 0.26 | 0.34 |
| 8 | 1.00 | 0.00 | 0.05 | 0.10 | 0.15 | 0.20 |

MIS = Memory Impairment Screen; PPV = positive predictive value.

(see Discussion), a cut-score of 4 provides a high level of sensitivity (0.80), specificity (0.96), and PPV (>0.69) for all base rates except the lowest at 5%. Because the base rate of dementia in the current sample is approximately 10%, approximately 70% of study participants who scored 4 or less had some type of dementia. The NPV reflects the probability that an individual who screens negative does not have disease. A cut-score of 4 provides excellent NPVs at low base rates of dementia. For example, at a base rate of 5% the NPV is 0.99; at a base rate of 10% the NPV is 0.98; at a base rate of 20% the NPV is 0.95. Thus one can be reasonably sure that those who screen negative do not have disease.

The results are similar when the analysis is restricted to cases of AD and controls (see table 3). A MIS cut-score of 4 had a sensitivity of 0.87 for AD, which is slightly higher than the sensitivity for all dementia. This differential sensitivity is expected because (early) memory impairment

Table 3 AD sensitivity and specificity of each MIS score with corresponding probabilities of AD (PPV) at different base rates

| MIS | AD | | Probability of AD (PPV) at different base rates | | | |
|-----|-------------|-------------|---|------|------|------|
| | Sensitivity | Specificity | 5% | 10% | 15% | 20% |
| 0 | 0.26 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 1 | 0.46 | 1.00 | 0.91 | 0.96 | 0.97 | 0.98 |
| 2 | 0.59 | 0.99 | 0.78 | 0.88 | 0.92 | 0.94 |
| 3 | 0.80 | 0.98 | 0.64 | 0.79 | 0.86 | 0.90 |
| 4 | 0.87 | 0.96 | 0.54 | 0.71 | 0.80 | 0.85 |
| 5 | 0.92 | 0.91 | 0.36 | 0.55 | 0.66 | 0.73 |
| 6 | 0.97 | 0.81 | 0.21 | 0.36 | 0.47 | 0.56 |
| 7 | 1.00 | 0.52 | 0.10 | 0.19 | 0.27 | 0.34 |
| 8 | 1.00 | 0.00 | 0.05 | 0.10 | 0.15 | 0.20 |

MIS = Memory Impairment Screen; PPV = positive predictive value.

is the hallmark of AD but not of all other (non-AD) dementias.

To examine how the MIS performed for different levels of dementia severity, we assessed discriminative validity separately for demented individuals who fell at or below 13 (mild) and at or above 14 (moderate) on the BIMC test. The sensitivity for mildly demented individuals with any type of dementia at an MIS cut-score of 4 is 0.69; for moderately demented individuals the sensitivity is 0.92. When analyses are restricted to only AD dementia, the sensitivity for mildly demented individuals is 0.79, and for moderately demented individuals the sensitivity is 0.95.

The demographic variables age, education, and gender as well as their interactions (with each other and with the MIS) were added to a logistic regression model as covariates to determine their effect on predictive accuracy. None of these demographic variables or interaction terms was significant at the 0.05 level. When the Zung Depression Scale score and its interaction were added to the logistic regression model, they were not significant at the 0.05 level. The ethnicity of the 17 false positives was examined to determine the possible influence of ethnic background on classification by the MIS. Of the 17 nondemented individuals who failed the MIS (cut-score = 4), 2 (11.7%) were African Americans, which is less than expected given the sample composition of 16.2% African Americans. Thus, individuals of minority backgrounds were no more likely to be incorrectly identified as memory impaired by the MIS than were white individuals.

Discussion. The MIS is a new screening test for memory impairment, dementia, and AD that is brief, easy to use, reliable, and valid. It is based on our previous research demonstrating that memory tests with high encoding specificity, such as the DMT and the FCSRT, have higher discriminative validity for AD and for dementia in general.^{27-29,31,36} The MIS uses category cues during acquisition (learning) to ensure attention, induce semantic processing, and optimize encoding specificity, and during retrieval to benefit from encoding specificity, increase retrieval, and improve discrimination. Because the MIS is a screening test, individuals who fail it are referred for definitive diagnostic assessment. Individuals who pass the test may be rescreened in the future depending upon the context of screening. In this section we summarize the psychometric properties of the MIS, illustrate how the normative data might be used in particular clinical and research applications, and discuss limitations.

We demonstrated that the MIS has high alternate forms reliability, indicating that the two forms are similar and that individuals perform in a highly repeatable fashion. The high reliability scores show that the alternate forms of the MIS can be used for repeated administration. The availability of equivalent alternate forms is an advantage in applications that require serial screening. There was a high degree of internal consistency for the two forms of the MIS (Cronbach's alpha = 0.67 for both).

Construct validity. Construct validity of the MIS as a screen for memory impairment was evaluated

using the FCSRT as the gold standard; the FCSRT has an empirically based cut-score for dementia-related memory impairment.³¹ Classifications by the MIS and by the longer FCSRT were quite similar, supporting the view that the MIS measures the same domain as standard tests of memory.

Discriminative validity. Discriminative validity of the MIS for dementia and AD was assessed in a large sample of individuals with and without clinically diagnosed dementia. In clinical practice, when older patients present with memory or cognitive complaints, the first issue is to diagnose or exclude dementia of any sort (e.g., AD, vascular dementia, dementia with Lewy bodies). If dementia is present, a more specific etiologic workup is essential. Thus, it is clinically important for a screening instrument to identify dementia syndromes with diverse etiologies. Although we present normative data separately for detecting all dementia (with memory impairment) or only AD, we believe that the discriminative validity data for detecting all dementia is most relevant for the majority of clinical and research applications. Discriminative validity of the MIS is influenced by dementia severity. In moderate dementia, sensitivity is 0.92 for all dementia and 0.95 for AD. Although sensitivity for mild dementia is lower, it is reasonably high—0.69 for all dementia and 0.79 for AD. The number of participants with non-Alzheimer dementia is relatively small; further experience with this group is needed.

The sensitivity and specificity of the MIS compares favorably with results in other samples for longer screening tests for dementia such as the BIMC test¹⁶ and the Mini-Mental State Examination (MMSE).¹⁷ Because the BIMC test was used by clinicians in this study to assign diagnoses, we cannot compare the MIS and the BIMC test obtained in this study and must rely on analysis of the MMSE and the BIMC test in other studies. Recent analyses of neuropsychological tests for diagnosis of dementia by Wilder et al.⁴² and Meiran et al.⁴⁰ reported the discriminative validity of the MMSE and the BIMC test for detecting AD. When Wilder et al.⁴² selected cut-scores for these tests so that sensitivity is high (0.90), their specificities are unimpressive—0.44 for the MMSE and 0.60 for the BIMC test. Data from Meiran et al.⁴⁰ indicate that when MMSE sensitivity is 0.91, MMSE specificity is only 0.25. The specificity of MIS for detecting AD is 0.96 when sensitivity is set to 0.87 (see table 3). These data suggest that the MIS may be a more sensitive screen for AD than these commonly used mental status examinations. Because the sensitivity and specificity of the MIS in this study are compared with the sensitivity and specificity of the MMSE and the BIMC test in other studies, these comparisons must be interpreted with caution; the populations used in different studies may have been different. In particular, the influence of education and cultural background on the relative performance of the MIS, the MMSE, and the BIMC test requires further study.^{43,44} Although the MMSE

and the BIMC test are influenced by education, at least in this sample education did not affect the discriminative validity of the MIS.

Influence of depression. Depression can influence performance on cognitive tests and enters into the differential diagnosis of dementia.³⁸ In the current study, Zung Depression Scale scores were not correlated with MIS scores for the demented or for the nondemented groups. Only seven participants met DSM-III-R criteria for major depression, and of these six were in the nondemented group. Of the 17 false-positive instances in the study (MIS of 4 or less, but not clinically demented), only one had depression. We do not have adequate data to draw firm conclusions about the influence of moderate or severe depression on MIS performance, because of small numbers of depressed individuals. However, there is little evidence to suggest that depression influenced MIS performance in our study results.

Normative data. Tables 2 and 3 provide the normative data needed to select cut-scores for the MIS. The PPVs, which reflect the probability of dementia for individuals falling at or below a MIS cut-score for a given base rate, provide an estimate of how likely those who screen positive will prove to have disease upon definitive assessment. We provide the following examples to illustrate how to use the MIS or any screening test to classify individuals. A clinician could be confident about classifying individuals scoring 4 or less on the MIS as demented because the probability of being demented is 0.84 (at a base rate of 20%; see table 2). However, if the base rate were only 5%, a clinician could not make such a confident classification because the probability of dementia is only 0.52. In general, clinicians will need to adopt lower (i.e., more stringent) cut-scores to maintain a given level of confidence for samples with lower base rates of dementia.

To select an optimal cut-score, a clinician or researcher needs to consider the base rate of the dementia in the screened population as well as the clinical or research goals of the screening program. For example, a clinician might choose a cut-score to maximize sensitivity so that impaired individuals are not missed by the screen, whereas a researcher conducting clinical trials might wish to maximize the PPV so that only those with a very high probability of dementia screen positive and are enrolled in the study.

Applications of screening. Interpretation and use of the normative data presented in tables 2 and 3 will depend on the clinical or research goals of the screening program. For a *clinical* application in which early diagnosis is important and follow-up testing for definitive diagnosis is safe and inexpensive, optimal *sensitivity* is important. In this case, a missed diagnosis would be undesirable given the lost opportunity for treatment of AD or other types of dementia. Accordingly, a clinician might use the data in table 2 to select a MIS cut-score of 5, which has a reasonably high sensitivity (0.86) and specific-

ity (0.91). How accurately this cut-score classifies patients will depend on the dementia base rate in the clinical setting. In a clinic-based sample of individuals with memory complaints we might reasonably expect a dementia base rate of 20%. At that base rate, a MIS cut-score of 5 corresponds to a 0.72 probability of dementia. This means that 72% of those meeting a MIS cut-score of 5 will have dementia (i.e., PPV = 0.72), and the sensitivity of 0.86 means that only 14% of all the demented patients will be missed by the screen. In this setting the NPV of 0.96 means that of those who screen negative, only 4% would prove to have dementia on a definitive assessment.

In contrast to a purely clinical application, *research* applications may require higher *specificity* to efficiently identify potentially diseased individuals. When the base rate of dementia is low, screening can be used to identify individuals likely to have dementia on definitive evaluation. For example, if the base rate of dementia is 10% in the target population, a researcher might select from table 2 a MIS cut-score of 3, which has a very high specificity (0.98). At that specificity the PPV is 0.78, which means that 78% of those who screened positive will meet criteria for dementia. Unlike the clinician, whose choice of cut-score will often be driven by the need to maximize sensitivity, the researcher may require high specificity to yield a high PPV. But even with this more restrictive criterion, the NPV of 0.97 means that only 3% of those who screen negative would prove to have dementia with a definitive assessment. Depending on the context, individuals who initially screen negative may be rescreened to identify individuals missed on the first round of screening.

Limitations and future directions. The current results apply only to use of the MIS as a screening test for detection of dementia-related memory impairment, dementia, and AD in community-residing adults aged 66 to 97 years. Cut-scores selected from tables 2 and 3 may not have the same level of discrimination in specific clinical applications because the distribution of MIS scores for demented and nondemented older adults in primary care settings may not be equivalent to that observed in a sample of community volunteers. Although the current results are encouraging, additional field studies are required to verify the validity of the MIS in screening for memory impairment and dementia in primary care settings. An important direction for future research would be to develop MIS norms for use in specific clinical and research settings.

An additional concern is that the clinical diagnosis of dementia is a fallible gold standard, although we assigned diagnoses with great care using contemporary research criteria. In this sample, individuals with preclinical dementia may not have crossed the threshold required for diagnosis. If such individuals have memory impairment on the MIS, they appear as false positives in our assessment of discriminative validity. Of the 17 individuals considered false posi-

tives (MIS score of 4 or less but not clinically demented), 12 had a Clinical Dementia Rating⁴⁵ score of 0.5 (i.e., questionable dementia) and four developed DSM-III-R dementia by the next wave of follow-up 18 months later. Although we do not have adequate data to comment on the discriminative validity of the MIS for preclinical AD, it appears that the preclinical group may account for a substantial proportion of the false positives, arguably leading to an underestimation of specificity.

One final and important limitation of the MIS is that it is a *screening* test and should not be used in isolation to make clinical diagnoses of dementia. By identifying individuals with a very high (or very low) likelihood of dementia, a good screening test can play an important part in the diagnostic process. No screening test should replace a comprehensive clinical evaluation. The MIS is intended to identify for clinical evaluation individuals who are likely to have dementia.

Acknowledgment

The authors thank Anna Sapozhnik Miller, Betty Forro, Charlotte Magnotta, Maria Luisa Giraldo, and Mary Joan Sebastian for their assistance with data collection, and Drs. Randall Berliner and Howard Crystal for assistance with clinical evaluations and diagnoses.

Disclosure

The Albert Einstein College of Medicine owns the copyright and makes this test available as a service to the research community but licenses the test for commercial use.

References

1. Wilcock G, Ashworth D, Langfield J. Detecting patients with Alzheimer's disease suitable for drug treatment. *Br J Gen Pract* 1994;44:30-33.
2. Ross G, Abbott R, Petrovitch H, et al. Frequency and characteristics of silent dementia among elderly Japanese-American men. *JAMA* 1997;277:800-805.
3. Kotler-Cope S, Camp C. Anosognosia in Alzheimer's disease. *Alzheimer Dis Assoc Disord* 1995;9:52-56.
4. Callahan C, Hendrie HC, Tierney WM. Documentation and evaluation of cognitive impairment in elderly primary care patients. *Ann Intern Med* 1995;122:422-429.
5. Farlow M, Gracon S, Hershey L, Lewis K, Sadowsky C, Dolan-Ureno J. A controlled trial of tacrine in patients with probable Alzheimer's disease. *JAMA* 1992;268:2523-2529.
6. Raskind M, Sadowsky C, Sigmund W, Beitler P, Auster S. Effects of tacrine on language, praxis, and noncognitive behavioral problems in Alzheimer's disease. *Arch Neurol* 1997;54:836-840.
7. Gelmacher D. Donepezil (Aricept) therapy for Alzheimer's disease. *Compr Ther* 1997;23:492-493.
8. Schordert M. Alzheimer's disease: fundamental and therapeutic aspects. *Experientia* 1995;51:99-105.
9. Van Reekum R, Black S, Conn D, Clarke D. Cognition-enhancing drugs in dementia: a guide to the near future. *Can J Psychiatry* 1997;42(suppl 1):35s-50s.
10. Khachatorian Z, Phelps C, Buckholtz N. The prospect of developing treatments for Alzheimer's disease. In: Terry R, Katzman R, Bick K, eds. *Alzheimer's disease*. New York: Raven Press, 1994:445-454.
11. Gauthier S, Thal L, Rosser M. The future diagnosis and treatment of Alzheimer's disease. In: Gauthier S, ed. *Clinical diagnosis and management of Alzheimer's disease*. London: Martin Dunitz Ltd, 1996:354-368.
12. Siu A. Screening for dementia and investigating its causes. *Ann Intern Med* 1991;115:122-132.
13. Tatemichi T, Sacktor N, Mayeux R. Dementia associated with cerebrovascular disease, other degenerative diseases, and metabolic disorders. In: Terry R, Katzman R, Bick KE, eds. *Alzheimer's disease*. New York: Raven Press, 1994:123-166.
14. U.S. Agency for Health Care Policy and Research. Recognition and initial assessment of Alzheimer's disease and related dementias. Clinical Practice Guideline No. 19. Silver Springs, MD: U.S. Agency for Health Care Policy and Research, 1996.
15. Petersen RC. Memory assessment at the bedside. In: Yanagihara T, Petersen R, eds. *Memory disorders*. New York: Marcel Dekker, 1991:137-152.
16. Blessed G, Tomlinson R, Roth M. The association between quantitative measures of dementia and of senile changes in the cerebral gray matter of elderly subjects. *Br J Psychiatry* 1968;114:797-811.
17. Folstein M, Folstein S, McHugh P. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975;12:189-198.
18. Strub R, Black F. *The mental status examination in neurology*. Philadelphia: FA Davis, 1985.
19. Teng E, Chui H. The modified Mini-Mental State (3MS) examination. *J Clin Psychiatry* 1987;48:314-317.
20. Kokmen E, Naessens J, Offord K. A short test of mental status: description and preliminary results. *Mayo Clin Proc* 1987;2:281-288.
21. Weintraub S, Mesulam M. Mental state assessment of young and elderly adults in behavioral neurology. In: Mesulam M, ed. *Principles of behavioral neurology*. Philadelphia: FA Davis, 1985:71-85.
22. Kokmen E, Smith G, Petersen R, Tangalos E, Ivnik R. The short test of mental status: correlations with standardized psychometric testing. *Arch Neurol* 1991;48:725-728.
23. Guilmette T, Tshoh J, Malcolm C. Orientation and three-word recall in predicting memory: age effects and false-negative errors. *Neuropsychiatry Neuropsychol Behav Neurol* 1995;8:20-25.
24. Cullum C, Thompson L, Smerhoff E. Three-word recall as a measure of memory. *J Clin Exp Neuropsychol* 1993;15:321-329.
25. White T, Bauer R, Bowers D, Crosson B, Kessler H. Recall of three words after five minutes: its relationship to performance on neuropsychological memory tests. *Appl Neuropsychol* 1995;2:130-138.
26. Knopman DS, Ryberg S. A verbal memory test with high predictive accuracy for dementia of the Alzheimer type. *Arch Neurol* 1989;46:141-145.
27. Buschke H, Sliwinski M, Kuslansky G, Lipton RB. Aging, encoding specificity and memory changes in the Double Memory Test. *J Int Neuropsychol Soc* 1995;1:483-493.
28. Buschke H, Sliwinski M, Kuslansky G, Lipton RB. Diagnosis of early dementia by the Double Memory Test: encoding specificity improves diagnostic sensitivity and specificity. *Neurology* 1997;48:989-997.
29. Buschke H. Cued recall in amnesia. *J Clin Neuropsychol* 1984;6:433-440.
30. Tulving E, Thomson D. Encoding specificity and retrieval processes in episodic memory. *Psychol Rev* 1973;80:352-372.
31. Grober E, Buschke H, Crystal H, Bang S, Dresner R. Screening for dementia by memory testing. *Neurology* 1988;38:900-903.
32. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*. 3rd ed, revised. Washington, DC: American Psychiatric Association, 1987.
33. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan E. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 1984;34:939-944.
34. Wechsler D. *Wechsler Adult Intelligence Scale-Revised*. New York: The Psychological Corporation, 1981.
35. Wechsler D. *Wechsler Memory Scale-Revised*. San Antonio, TX: The Psychological Corporation, 1987.
36. Buschke H. Criteria for identification of memory deficits: implications for the design of memory tests. In: Gorfein D, Hoffman RR, eds. *Memory and learning: the Ebbinghaus Centennial Conference*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1987:331-344.

37. Zung W. A self-rating depression scale. *Arch Gen Psychiatry* 1965;12:63–70.
38. Spitzer R, Williams J, Gibbon M, First M. Structured Clinical Interview for DSM-III-R. Nonpatient Edition (SCID-NP, version 1). Washington, DC: American Psychiatric Press, 1990.
39. Elwood R. Clinical discriminations and neuropsychological tests: an appeal to Bayes' theorem. *Clin Neuropsychologist* 1993;7:224–233.
40. Meiran N, Stuss D, Guzman A, Lafleche G, Willmer J. Diagnosis of dementia: methods for interpretation of scores of 5 neuropsychological tests. *Arch Neurol* 1996;53:1043–1054.
41. Hanley J, McNeil B. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; 143:29–36.
42. Wilder D, Cross P, Chen J, et al. Operating characteristics of brief screens for dementia in a multicultural population. *Am J Geriatr Psychiatry* 1995;3:96–107.
43. Katzman R. Education and the prevalence of dementia and Alzheimer's disease. *Neurology* 1993;43:12–20.
44. Kittner SJ, White LR, Farmer ME, et al. Methodological issues in screening for dementia: the problem of education adjustment. *J Chron Dis* 1986;39:163–170.
45. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology* 1993;43:2412–2414.

Tacrine use in nursing homes

Implications for prescribing new cholinesterase inhibitors

D.R. Gifford, MD, MPH; K.L. Lapane, PhD; G. Gambassi, MD; F. Landi, MD;
V. Mor, PhD; and R. Bernabei, MD, for the SAGE Study Group*

Article abstract—*Objective:* To describe the use of tacrine in nursing home residents using data from a clinically based resident assessment instrument used by all US nursing homes. *Methods:* Data were from the Systematic Assessment of Geriatric Drug Use via Epidemiology (SAGE) database, a population-based data set with information on 329,520 patients admitted to all Medicare/Medicaid certified nursing homes in four US states (Maine, Mississippi, New York, and South Dakota) from 1992 through 1995. The SAGE database combines information from the Minimum Data Set (MDS) and the On-Line Survey and Certification Automated Record. We identified all residents receiving tacrine and up to five control residents per case matched on state, date of tacrine use, cognitive function, and dementia diagnosis. *Results:* A total of 1,640 (0.5%) nursing home residents received tacrine at least once. Only 38% of these residents had a diagnosis of AD documented on the MDS; regardless of dementia diagnosis, 25% had severe cognitive impairment, 35% were severely dependent in activities of daily living (ADL), and 17% had both severe cognitive and ADL impairment. Only 8% achieved a therapeutic dose of at least 120 mg/d. After adjusting for confounding variables, wandering and being physically abusive were the strongest predictors of tacrine use. *Conclusions:* A minority of nursing home residents received tacrine. Of those who did, a significant proportion were unlikely to benefit from its use because of their level of cognitive and ADL impairment, or because low doses were used. As new medications become available for dementia, MDS data can be used by nursing homes to monitor the use of these therapies.

NEUROLOGY 1999;52:238–244

The first cholinesterase inhibitor, tacrine, was approved in 1993 for the treatment of AD in the United States.¹ A second cholinesterase inhibitor, donepezil, was approved in 1997.² According to current guidelines,^{3,4} cholinesterase inhibitors are only indicated for patients with a diagnosis of AD who have mild to moderate dementia (e.g., Mini-Mental State Examination [MMSE] score of 10 to 26).^{1,2,5} In the ambula-

tory setting, patients with a confirmed diagnosis of AD who receive tacrine exhibit, on average, statistically significant improvement on cognitive function tests.^{6–9} Data from these trials and others suggest that tacrine also may affect other outcomes,^{7,10–12} but these findings are less conclusive and more controversial.⁹

Little is known about the pharmacologic manage-

See also page 227

*See the Appendix on page 243 for a list of members of the SAGE Study Group.

From the Departments of Medicine (Dr. Gifford) and Community Health (Drs. Gifford, Lapane, Gambassi, and Mor) and the Center for Gerontology and Health Care Research (Drs. Gifford, Lapane, Gambassi, and Mor), Brown University, Providence, RI; and Istituto di Medicina Interna e Geriatria (Drs. Gambassi, Landi, and Bernabei), Università Cattolica del Sacro Cuore, Rome, Italy.

This study was conducted while D.R.G. was a Pfizer/American Geriatric Society Postdoctoral Fellow and SURDNA fellow at the Brown Center for Gerontology and Health Care Research. The study was also supported in part by the "Do Good Nursing Homes Achieve Good Outcomes?" project (National Institute on Aging #AG11624) and by the "Changing Roles in Nursing Homes" project (#500-89-0046; US Department of Health and Human Services, Health Care Financing Administration, to University of Michigan with subcontract to Brown University).

Received May 18, 1998. Accepted in final form October 24, 1998.

Address correspondence and reprint requests to Dr. David R. Gifford, Center for Gerontology and Health Care Research, Brown University, Box G-B 222, Providence, RI 02912; e-mail: David_Gifford@brown.edu