Alzheimer Patient Evaluation and the Mini-Mental State: Item Characteristic Curve Analysis

J. Wesson Ashford,1 Paul Kolm,1 Jerry A. Collier,1
Cathy Bekian,2 and Lee-Nah Hsu1

1School of Medicine, Southern Illinois University.
2Neuropsychiatric Institute, University of California at Los Angeles.
3The Health Data Institute, Lexington, MA.

To develop a tool for precisely assessing dementia severity, items should be valued according to their relationship to the overall progression of the disease. Using a logistic characteristic curve analysis (ICCA), items were examined from the Folstein Mini-Mental State Exam (MMSE), a useful clinical test for evaluating dementia. MMSE data were available for 50 patients who met DSM-IIIR criteria for primary degenerative dementia—possible or probable Alzheimer's disease (AD). A logistic regression analysis of the probability of correct performance on an item, given the total MMSE score, yielded statistics for difficulty and discrimination. These statistics were comparable respectively as indicators of the point in the progression of the disease at which the mental function tested by that item is lost and the severity of that loss. The data indicated a systematic progression of the development of symptoms in AD related to decline of memory function. Temporal information was lost before spatial and object orientation, and recollection of words was last before ability to repeat them. ICC can help to delineate the loss of mental functions during the course of AD.

MEMORY dysfunction in elderly individuals is an indication of dementia and the two common initial symptoms of Alzheimer's disease (AD). The most recent diagnostic criteria for dementia (DSM-IIIR) recognizes memory as the primary impairment of dementia (American Psychiatric Association, 1987). In very mild cases, neuropsychological evaluation focuses on determining whether memory, especially recall of recently obtained information, is affected. Dementia can be caused by several different diseases. Therefore, psychological status is carefully measured to determine if the pattern of dysfunction is characteristic of AD (Feld, 1983), multi-infarct dementia (Hachinski, Lassen, & Marshall, 1974), left hemisphere stroke (Benson, Cumming, & Tsai, 1982), or another disorder. Although the clinical validity of such determinations is still in doubt (Erkinjuntti, Laaksonen, Syvälahti, & Pollo, 1986), specific differences in the nature of memory impairment between different diseases can be discriminated (Moss, Albert, Butters, & Payne, 1986). Clinical progression is helpful in confirming the diagnostic impression of AD (McKhann et al., 1984).

AD begins with mild symptoms of forgetfulness and progresses over years, highlighted by the degeneration of memory, to cause total cognitive dysfunction. For patients with suspected AD, evaluation is also aimed at assessing the severity of the disease. AD generally progresses in a typical fashion, certain cognitive and social functions usually being lost before others. Therefore, the clinician assesses the patient's psychosocial function to determine severity. Such assessments of severity are useful in management of the patient because of their relation to the level of daily living function (Vitaliano, Breen, Albert, Russo, & Printz, 1984a). However, these assessments currently have little value in predicting future decline. Measurement of disease severity must become more precise and reliable for use in estimating prognosis or for research applications.

The importance of determining dementia severity has led to the development of many assessment tools (Nelson, Fogel, & Faust, 1986). Because dementia is associated with such a broad spectrum of decline, very brief sets of questions can give reliable information about severity (MSQ: Kahn, Goldfarb, Pollack, & Pesk, 1960; SMPSQ: Pfeifer, 1975). Such brief tests also correlate well with the severity of AD pathology in the brain (Blessed, Tomlinson, & Roth, 1968). The Mini-Mental State Exam (MMS) (Folstein, Folstein, & McHugh, 1975) has gained widespread acceptance because of its ease of administration, its assessment of a popular variety of standard mental status items, and the broad range over which it can usefully assess dementia severity (Anthony, LeRiche, Niz, von Korff, & Folstein, 1982). The Blessed Information-Memory-Concentration Test and the MMS show a correlation ranging between .73 and .83 (Thal, Gunning, & Goldson, 1986). Other authors, feeling uncomfortable with such brief assessments, have advocated the use of a battery of tests to assess dementia severity (Pfeifer et al., 1984). Edlinger, Damasio, Benton, & Van Allen, 1985; Klein et al., 1985). However, none of the currently used tests has been examined on an item-by-item basis to validate the utility of each item for judging dementia severity or for providing a precise picture of the progression of AD.

Item analysis techniques are widely used to assess individual test items when groups of tests are being quizzed regarding their knowledge or aptitude for a given subject. After initial usage, each test item can be evaluated to determine the level of performance at which the item dis-

F139
criminates and the sharpness of that discrimination. Items used to test the performance of demented patients can be similarly analyzed. The level of severity at which specific items discriminate could suggest which functions were lost before others during the progression of the disease. To the extent that AD is a uniformly progressive disease, item analysis techniques could be used to assess uniformity and construct a more reliable scale for patient assessment.

In the present study, MMS items were analyzed from an item characteristic curve (ICC) analysis perspective. The relationship between performance on any given item and ability as assessed by overall test performance is described by a plot of the probability of success on the item as a function of overall test performance. A mathematical model fit to these plotted data results in an ICC from which the difficulty and discriminability of each item can be determined.

Difficulty describes the location of an item along the scale of overall test performance (ability). A measure of difficulty is defined as the score at which the expected probability of correct response is 0.5. A small value indicates an easy item and a large value indicates a difficult item. An easy item is one for which the probability of correct response is high for low-ability examinees (or, in this case, severely impaired patients) and approaches 1 for high-ability examinees (in this case, mildly impaired patients). An item of medium difficulty is one for which the probability of correct response is low for low-ability examinees, intermediate for those in the middle of the ability scale, and near 1 at the highest ability levels. A difficult item is one for which the probability of correct response is even lower for high-ability examinees (Baker, 1985, p. 5).

Discrimination describes how well an item can differentiate between examinees having abilities below the item location and those having greater capabilities. Essentially, the slope of the ICC is a measure of discrimination, because the probability of correct response would change more rapidly within a narrower range of ability levels for a curve with a steeper slope. The flatter the curve, the less the item discriminates, as the probability of correct response at low-ability levels approaches the probability of correct response at high-ability levels (Baker, 1985, pp. 5-6). In other words, the probability of correct response changes more slowly over a wider range of ability levels.

Applied to the MMS, difficulty is an indicator of the point at which the mental function assumed to underlie performance on the item is lost in the progression of AD; discriminability is an indication of how quickly that function is lost. An item with high difficulty and high discriminability would indicate an early loss and a loss that occurs quickly in the progression of the disease. High difficulty and low discriminability would indicate an early loss, but a loss that occurs over a longer range of progression. An item with low difficulty and high discriminability would indicate loss of function late in the progression of the disease, but a loss that occurs quickly. An item with low difficulty and low discriminability would indicate late loss and a loss occurring more slowly over the progression of the disease.

The purpose of this study was to examine these item statistics in order to identify: (a) the level or degree of severity of AD, as indicated by MMS total score, at which particular items are lost in the progression of AD, and (b) the rapidity of rate with which they are lost at that level. Level of severity at which particular functions are lost is, in turn, indicated by item difficulty, and the rate of loss by item discriminability. Specific items are considered to represent particular underlying mental functions, and loss of performance on an item presumably indicates impairment of that function.

METHODS

Subjects. — All patients in this study were evaluated in the UCLA Geriatric Psychiatry Outpatient Clinic. Patients presenting with complaints including memory difficulty were given complete dementia evaluations including psychiatric and neurologic examinations and CT scans and EEG (Wells, 1977). Between July 1982 and March 1984, 112 patients had received complete dementia evaluations that included a Mini-Mental State Exam (MMS). Of this group, 86 met DSM-III criteria for a clinical diagnosis of primary degenerative dementia (PDD). Subsequent studies involving this group of 86 patients indicated that 60% met NINCDS-ADRDA criteria for probable AD (McKhann et al., 1984) and 40% for possible AD (Ashford, Rosenblatt, Becak, & Hayes, 1987) [National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer’s Disease and Related Disorders Association]. The present study focused on these 86 patients.

The average age of these patients was 74 years (SD = 8; range = 53-91). There were 23 males and 63 females. Average education was 11 years (SD = 3; range = 0-20). For 81 of these patients a significant other was available, in case to provide information to assess Activities of Daily Living (ADLs), using the format of the GARS, Duke University, 1975).

Measures of dementia severity. — MMS scores (mean = 18.0; SD = 7.1; Range = 1-29) were computed according to the rule that the best score of serial 7’s or “WORLD backward” was used for the total (this is an acknowledged weakness of this test, as the serial 7 task is more difficult). The ADL scores were totaled (scale range = 0-25) and the average score of the group was 21.1 (SD = 5.7; range = 4-29). ADL scores and MMS scores were highly correlated (r = .76; n = 81; p < .001), both for males (r = .57; n = 20; p < .009) and females (r = .77; n = 61; p < .001). Both ADL scores and MMS scores were negatively correlated with age (r = -.35; n = 81; p < .002 and r = -.39; n = 86; p < .001, respectively) (possibly reflecting a greater tolerance for social and cognitive dysfunction in more elderly individuals; older patients thus seem to be brought to clinical attention for the first time at a more advanced stage of the dementing process. No relationship was noted between educational status and either age (r = .10; n = 65; p = .43) or ADL (r = .13); n = 61; p = .32) or MMS (r = .19; n = 65; p = .13) scores.

Statistical methods. — MMS items are scored as either correct or incorrect. Therefore a logistic regression model
was applied, in this case, regression of the binary outcome, correct or incorrect, on the total MMS score. Logistic regression is a practical method of calculating a sigmoid curve to fit the data. The logistic regression resulted in an expected probability of success on an item given a particular MMS score. Specifically, the regression estimated the parameters of the model:

$$P(Y=1) = \frac{e^{c+b(MMS)}}{1+e^{c+b(MMS)}}$$

where c is the regression constant and b the regression coefficient. MMS represents the total MMS score. Severity at loss (difficulty) was defined as the score at which the expected probability of correct response is .50. Maximum rate of loss (discrimination) was indicated by the slope of the curve at the 0.5 probability level; the steeper the slope, the greater the rate and the better the item served as discriminative between more severely and less severely affected patients.

A chi-square goodness-of-fit test was used to determine whether the fit of the logistic regression model to the ICC was statistically significant. In addition, a measure of goodness-of-fit was calculated by correlating the predicted probability of success with the observed proportion of success (R). The square of the correlation coefficient ($R^2$) indicated the proportion of variance in the ICC data explained by the logistic regression model. These statistics are indicators of the validity of the model.

RESULTS

Table 1 presents the item analysis statistics for each item. Items are arranged arbitrarily into four blocks based on the rate of loss index (discriminability), the first group having the highest rate and the last group the lowest rate. Within each group, individual items were ordered from earliest loss to latest loss (difficulty). The chi-square goodness-of-fit test indicated that for all items, the fit of the logistic regression model to the ICC was statistically significant (not data not presented). However, from the $R^2$ value, it can be seen that the degree of fit was much better for some items than for others.

The first group of items in Table 1 were those with the highest rate of loss index. Figure 1 presents the ICCs for four of these items with the logistic regression curve overlaid. The high rate of loss index indicates that there was a sharp cut-off of ability level at which the item was answered correctly or incorrectly. For example, the ICC for the easiest item, "pencil," indicated that a patient with an MMS score of about 10 or more would have virtually a 100% chance of getting the item correct. These statistics indicate that the mental function assessed by the item "pencil" is lost at an advanced stage of the disease and that it is lost quite quickly at this stage. At the other extreme, "date" had a very high severity at loss index. An individual with a score of 20 or lower would have a very small chance of getting the item correct. This indicates that the mental function assessed by "date" is lost very early in the progression of AD and is lost quickly.

Figure 2 presents the ICCs for four items from the second group. These items have a fairly high rate of loss, although not as great in degree as the first group. Interpretation of the sequence and rapidity of loss of mental function would be essentially the same as for the first group.

Figure 3 presents four examples of the ICCs from the third group of items. It is readily apparent that rate of loss of these items was not nearly as great as the first two groups of items. In other words, there was a wider range of MMS scores in which subjects answered the item either correctly or incorrectly. This can be interpreted to mean that the mental functions underlying these items are lost more gradually or more variably than those underlying the first two groups of items or that mental functions are tested less succinctly by these items.

The four ICCs presented in Figure 4 are for those items at the extremes of the MMS for this sample, that is, the items lost first and the items lost last. The items lost first, "lag-memory" and "tree-memory," were answered correctly by less than 17% of the sample, whereas the items lost last,
"paper-th" and "paper-th." were answered correctly by over 80% of the sample. The rate of loss indices for these items cannot be interpreted in quite the same way as for the other items. Because responses were mostly correct for the items lost last and incorrect for items lost first, the ICCs are very flat over most of the scale; therefore the rate of loss indices is necessarily low. In other words, because these items are lost almost immediately or not until the end, a concept of rate of loss is not meaningful. These items are of considerable potential utility because they measure function at the extremes of the MMS state.

While Table 2 presents the same information as in Table 1, the items in the former are ordered by severity at loss. The
data are represented in this fashion to more clearly illustrate the progression of loss across all of the items. The items lost first were those where the patient was required to remember information after distraction, recall the date, and count backwards. Items last lost were those requiring the patient to perform an action immediately following some instruction or to repeat information immediately after presentation.

DISCUSSION
The results of the analysis of MMS items show a differential pattern of loss of items which implies that there is a pattern of deterioration of mental functions in the progression of AD. The items with the highest severity at loss index (MMS score above 20), indicating earliest loss, are recent memory items: the three detail memory items (ball, flag, and tree), recall of the date, and the serial 7 calculations beyond the first subtraction. The latter are difficult because they require a functional recent memory — the patient must recall what he or she is supposed to do next after being distracted by performing the first subtraction. Indeed, memory difficulty is the most common initial problem affecting patients (Liston, 1979a; 1979b), and memory is rapidly lost early in the course of the disease (Folstein & Whitehouse, 1983; Vitaliano, Breen, Russo, Albert, Visiello, & Prinz, 1984b; Storandt, Botwinick, & Danziger, 1986). Items that became impaired in the middle category of severity level (MMS greater than 10 and less than 20) were time and place orientation items that utilize longer term memory functions and involve many more cues for developing acquisition. The item requiring the perception and reproduction of intersecting pentagons was affected quite variably, perhaps because it involves many different cerebral systems. The items lost later in the progression of AD (severity at loss index of 10 or less on the MMS) are those requiring use of the most solidly stored memories: early-learned, verbal mimicking (the repetition of simple words), over-learned associations (the naming of simple objects), and frontal lobe procedural functions (the following of simple commands). Naming objects, writing, and reading are lost even later in AD progression than the repetition and command items. Indeed, expressive language deficits indicate later disease stage (Kazinski et al., 1976; Folstein & Whitehouse, 1983; Heyman, 1984). Thus, the pattern of loss of performance on MMS items is consistent with the observed clinical course of AD. A particularly striking and well-known comparison clearly documented by these results is the difference between the ability to repeat the names of three objects (ball, flag, tree) and to recall them later. Most of the patients in our group, particularly those with mild and moderate impairment, clearly perceived these items and could repeat them easily (87%; 84%, and 86% correct, respectively). Yet only a few of the patients were able to recall these items after distraction (24%; 9%, and 16% correct, respectively). This clear example and the overall pattern of deficit development support the notion that the single underlying factor in Alzheimer dementia is a disorder of memory. The most recent, volatile memory storage is the first to be disrupted, followed by longer term memories, with the pathological process affecting over-learned associations and highly practiced motor skills appearing late in the disease.

The importance of memory dysfunction in Alzheimer's disease has been well documented. The disease affects both recent memory (Corkin, 1982) and remote memory (Wilson, Bacon, Fox, & Kasnak, 1983). Both recall and recognition memory is impaired (Moss et al., 1986). Numerous other functions have been documented to be impaired in this disease, including language function (Hoff, Corkin, & Growdon, 1986; Cummings, Benson, Hill, & Read, 1985), object perception (Flekoay, 1976), and motor skills (Foster, Chase, Paterson, Gillespie, & Fedio, 1986). However, the analysis of the present study supports an alternative clinical impression that aphasia, agnosia, and apraxia, which are frequently diagnosed in AD, are really difficulty in remembering words, what things are for, and how to do things. This leads to the hypothesis that Alzheimer's disease is quintessentially a disorder of language.

There have been numerous reports of diversity in the impairments associated with Alzheimer's disease. Many of these studies have stressed the presence of asymmetries in the temporal lobes of patients, with more severe impairment on the right side in patients with right-handedness (Foster et al., 1983; Benson et al., 1983; Friedland, Budinger, Koss, & Ober, 1985) or on neuropsychological testing (Martin, Cox, Bowers, & Fedio, 1985; Filley, Kelly, & Heath, 1986;
Grady, Haxby, Sehlgren, Berg, & Rapport, 1986). In our analysis of MMS data, the rate of loss index is a measure of the
severity of the progressive loss of memory function. This variability accommodates observations of disproport-
tonate cerebral involvement at any one phase of the disease, even if that means that one hemisphere is much more
impaired than the other at that time. However, evidence that language or visuospatial functions may be selectively im-
paired at some point in the illness does not dispute that memory impairment is at the root of these deficits. On the
contrary, across the full course of the disease a progressive pattern emerges in which functions dependent on more
recent memory are lost before functions that are dependent on more solidly stored memories.
Identifying the progressive nature of AD as related to a sequential disruption of memory function rather than a
deterioration of certain anatomical structures (Whitehouse, 1982) is relevant to understanding the diverse cellular neuro-
pathology of AD. Several authors have proposed an anatom-
ical spread of Alzheimer pathology (Saper & German, 1987), that the disease process affects cells because of their
degree of connectivity (Gajdusek, 1985; Hyman, Van
Hoesen, Kromer, & Ganshoj, 1986; Lewis, Campbell,
Terry, & Morrison, 1987) we relate to effector connec-
tions (Pearson, Eusti, Horn, Wilcock, & Powell, 1985;
Roberts, 1986). However, these hypotheses fail to account
for the preservation of basic functions, such as perception,
in regions that have totally lost the capacity to store new
information [this point is based on the concept that memories are stored in the same area that information is originally
perceived (Ashford & Fuster, 1985; Mahlkin & Appenzel-
ier, 1987)]. Furthermore, these hypotheses do not explain
why the tempo-parietal regions, where detailed memories are
stored, are disproportionately affected compared to the fron-
tal lobes (Braun, 1983). The solution to this dilemma of
"what system is affected by AD" is that neural plasticity
itself, the neural foundation of memory storage, predisposes
to the development of Alzheimer pathology in a particular
neuron (Ashford & Jarvik, 1985). Just as the cells with the
higher cell division rates are the most vulnerable to cancer,
those cells most vulnerable to neurofibillary disruption are
those cells which produce the most new neurofibils. Neural
plasticity has recently been conceived of as an active process
involving continuous production and loss of new synapses
(Lee, Schottler, Oliver, & Lynch, 1980; Chang &
Greenough, 1984; Cottman & Neto-Sampredo, 1984). This
process is a likely target for the neurologist's agent which
causes AD, because the productivity of neural plasticity
reaches its greatest extent in the human brain, and it is the
human that is most vulnerable to AD.
The MMSE has two weaknesses, one at each extreme of test performance, which were particularly problematic for
the logistic regression analysis (Table 1, group A). First, the
MMSE is not an adequate test to distinguish patients with
very mild AD from normal patients. Problems in making this
distinction include variation in educational and socioeco-
nomic status (Cavanaugh, 1983). There may be too much variability early in the course of dementia to rely on any
simple scale to reliably identify demented patients, and this
task should be left to clinical judgment. However, usual
clinical procedures have done poorly in detecting dementia
(Klein et al., 1985), and PET scan data suggest that those
patients who were not diagnosed at the time were probably
already have substantial metabolic impairment (Kull et al.,
1985, Haxby et al., 1986). Thus, there is a need to expand
early patient evaluation with more difficult but still-sharply
discriminating test items that lack external bias, to assist the
clinician in improving the early detection of dementia. The
second difficulty with the MMSE is that the score reaches
zero at a stage in the disease after which a patient may continue to deteriorate for several years. Thus, there is
little left to evaluate in the way of cognitive functions, and
the more relevant measures are ADL scales (Katz, Ford,
Moskowitz, Jackson, & Jaffe, 1963; Long & Lynn, 1983;
Lawton, 1983), which follow a parallel sequence of func-
tional loss but are more sensitive to severe impairment
(Ashford, Hsu, Becker, Kumar, & Bekian, 1986).
The ICC analysis technique offers an approach for im-
proving the measurement of dementia severity. A large
number of items could be administered to AD patients, and
those items displaying the best discrimination across the
spectrum of deterioration could be used for developing an
ever-expanding test to more accurately assess patients with
early and late AD. More accurate assessment would lead to
better measurement of rate of decline and improve precision of future deterioration.
As is true with all AD studies, autopsy confirmation of AD in each case would strengthen the findings. However,
quantification of pathology at late stage death and correlation with performance near that event (Blessed et al., 1968)
does not imply validation of a test across a broad spectrum of function and the long course of AD (though the Blessed scale
does correlate well with the MMSE; Thal et al., 1986). Studies must focus on early detection, long-term follow-up,
and autopsy confirmation to determine the robustness of AD.
ACKNOWLEDGMENTS
Address correspondence to D. W. Wesson Ashford, Department of
Psychiatry, School of Medicine, Northern Illinois University, P.O. Box
19230, Springfield, IL 62794-9230.
REFERENCES
American Psychiatric Association. (1987). Diagnostic and statistical man-
Author.
Anthony, J. C., Lebowitz, L., Nau, U., Von Korff, M. R., & Rotenstein, M.
dementia and delirium among hospital patients. Psychological Medi-
cine, 12, 371-408.
Ashford, J. L., & Fuster, J. M. (1986). Occipital and inferotemporal
responses to visual stimuli in the elderly. Experimental Neurology, 90,
442-449.
Mental status and activities of daily living: Cross-validation of scales
and item analysis techniques. The Gerontologist, 26, 143A.
plasticity predominate to avoid neurofibillary degeneration? New Eng-
land Journal of Medicine, 313, 386-389.
complete dementia evaluation: Complications and complexities. Ameri-
can Journal of Alzheimer's Care and Research, 2, 9-15.
Heitenschel Publishing.


Received February 19, 1980
Accepted November 15, 1980.