Diagnosis of Dementia in a Heterogeneous Population

Development of a Neuropsychological Paradigm-Based Diagnosis of Dementia and Quantified Correction for the Effects of Education

Yaakov Stern, PhD; Howard Andrews, PhD; John Pittman, MS; Mary Sano, PhD; Thomas Tatemichi, MD; Rafael Lantigua, MD; Richard Mayeux, MD

• A brief diagnostic battery of neuropsychological tests was developed for a large-scale epidemiological study of dementia. We operationally defined dementia as defective memory and defective performance in at least two other areas, including orientation, abstract reasoning, construction, and language. Criterion scores for defining defective performance on each test were developed. In a pilot study that used 51 different subjects with a working diagnosis based on physicians' assessment (ie, 32 demented and 29 nondemented subjects), the test-based diagnosis agreed with the working diagnosis in all but two cases. The test battery was then applied to 430 healthy elderly subjects. Eighteen percent of those with 8 or less years of education met criteria for dementia compared with 5% of those with more than 8 years of education. We computed education-corrected scores for each test with the use of residuals from the regression of each test score on education. Based on corrected scores, 12 subjects were reclassified as nondemented and 11 as demented. Subjects who were reclassified as demented were significantly more impaired in activities of daily living than nondemented subjects who were not reclassified. Activities of daily living in subjects who were reclassified as nondemented did not differ from those in demented subjects who were not reclassified. These findings suggest that the neuropsychological battery may have utility in the diagnosis of dementia. However, neuropsychological performance may be influenced by education, and some form of adjustment, such as correction for activities of daily living, may be required in epidemiological studies.
(*Arch Neurol.* 1992;49:453-460)

Large-scale epidemiological studies of dementia have relied on screening tests for diagnosis, but a more definitive diagnosis may require extensive neuropsychological evaluation. In planning a community-based study of dementia, we developed a relatively short (generally, <1 hour) but comprehensive neuropsychological evaluation for diagnosis. This evaluation was also translated into Spanish.

See also p 461.

We employed operational criteria for diagnosis that made use of objective test scores to ensure that (1) study criteria for dementia were objective and replicable, and (2) criteria would not "drift" over time as a function of the experience and interaction of the examiners.

This article has two parts. First, we describe the neuropsychological battery and operational criteria, and we report results of a pilot study that was designed to determine if the criteria could produce diagnoses similar to those derived by our more standard and lengthy clinical evaluation. Second, we summarize our experience with this battery in a large group of elderly subjects and explore the possible influence of education on test results.

NEUROPSYCHOLOGICAL BATTERY

The neuropsychological battery was selected from subsets of items from standardized neuropsychological tests to assess intellectual functions that are typically affected in dementia. Average administration time was 1 hour.

For each test, all items and instructions were translated into Spanish and then translated back to ensure accuracy.

Memory

Verbal Memory.—The Selective Reminding Test was used.[1] Subjects were given six trials to learn a list of 12 unrelated words. After each attempt at recalling the list, the subject was reminded only of the words that were not recalled and then asked to attempt again to recall the entire list. To assess short-term verbal memory, two performance measures were used: (1) total recall and (2) retrieval from long-term storage. To assess long-term verbal recall, delayed recall was assessed 15 minutes after completing the Selective Reminding Test; recognition of words that were not recalled was then tested with the use of multiple-choice arrays.

Accepted for publication December 12, 1991.

From the Departments of Neurology (Drs Stern, Sano, Tatemichi, and Mayeux), Psychiatry (Drs Stern, Andrews, and Mayeux and Mr Pittman), and Medicine (Dr Lantigua), Gertrude H. Sergievsky Center (Drs Stern and Mayeux), College of Physicians and Surgeons, Columbia University and the Alzheimer's Disease Research Center (Drs Stern, Sano, and Mayeux), New York, NY.

Reprint requests to Neurological Institute, 710 W 168th St, New York, NY 10032 (Dr Stern).

Nonverbal Memory.—A multiple-choice version of the Benton Visual Retention Test[2] was used to assess nonverbal memory in a format that did not rely on constructional abilities. The subject viewed a design for 10 seconds. It was then removed, and the subject was asked to recognize the design in an array that included three distractors. Ten items were used, corresponding to Form D of the original Benton Visual Retention Test.

Orientation

Ten items from the Mini-Mental State Examination[3] were used to assess orientation to time and place.

Visuospatial Ability

Construction.—In this test, the subject copied five designs that were selected from the Rosen Drawing Test[4] to span a range of difficulty from simple shapes and topological concepts to overlapping, euclidean, and three-dimensional designs.

Benton Visual Retention Test Matching.—For each of 10 items, the subject matched a larger picture to one in an array of four smaller pictures. Items corresponded to Form C of the original Benton Visual Retention Test.[2]

Language

Naming.—Fifteen selected items from the Boston Naming Test[5] were used.

Verbal Fluency.—*Controlled Oral Word Association.*—For this association,[6] the subject was given 1 minute each to name as many words as possible, beginning with the letters C, F, and L. For Spanish-speaking subjects, the letters P, S, and V were used. Percentile scores were derived based on age- and education-adjusted norms.

Category Naming.—The subject was allowed 1 minute each for three categories: animals, food and clothing. Scores were expressed in terms of the mean number of words reported in the three categories.

Comprehension.—The first six items of the Complex Ideational Material subtest of the Boston Diagnostic Aphasia Evaluation[7] were used to assess verbal comprehension. These items required only yes/no answers to relatively simple questions.

Repetition.—The high-frequency items from the Boston Diagnostic Aphasia Evaluation Repetition of Phrases subtest[7] were used.

Abstract Reasoning

Wechsler Adult Intelligence Scales–Revised.—This Wechsler Adult Intelligence Scale–Revised[8] subtest required the subject to identify relevant similarities between pairs of items. Age-scaled scores were used.

Identities and Oddities.—In this subtest of the Mattis Dementia Rating Scale,[9] the subject was asked to identify the two of three items that were the same. After eight trials were completed, the same items were administered again, with the subject identifying the one item that was different.

DIAGNOSIS

We used a series of criterion scores to determine whether a subject's intellectual function was impaired to the extent that was required to meet criteria for dementia. Criterion scores were determined based on a review of the performance of 172 patients and controls who had been evaluated in previous clinical studies or in our Memory Disorders Clinic (New York State Psychiatric Institute, New York City). In this group, 32 were nondemented elderly controls, 77 had probable Alzheimer's disease,[10] 39 had Parkinson's disease (PD), 14 had PD and dementia, and 10 had PD and major depression. For each test, mean scores and variability in each group were inspected, and the score that best separated nondemented and demented groups was selected as the criterion score.

Based on criteria for dementia according to the *Diagnostic and Statistical Manual of Mental Disorders, Revised Third Edition*,[11] memory was considered to be the key-defining feature of dementia; we required that two of three of the defined memory "areas" ([1] short- and [2] long-term verbal memory and [3] short-term nonverbal memory) be defective to meet criteria for dementia. In addition, performance on at least two of the following areas also had to be impaired: orientation, construction, abstract reasoning, and language. The diagnostic paradigm, including criterion scores, is summarized in Table 1. The maximum possible score for each test is also included in Table 1, where applicable.

The requirement that two aspects of memory be impaired, as well as at least two other cognitive functions, was intended to ensure that subjects were not misclassified on the basis of poor performance on a single measure. In addition, it was designed to simulate to some degree the approach that a clinician would take toward investigating a pattern of neuropsychological performance in a clinical dementia evaluation.

STUDY 1: PILOT STUDY OF NEUROPSYCHOLOGICAL BATTERY

Methods

Subjects.—Fifty-one individuals were included in this study. All subjects were well known to us and had established diagnoses based on findings from full clinical evaluation and extensive neuropsychological testing. Five subjects were healthy elderly individuals, 12 had probable Alzheimer's disease,[10] 17 had stroke (11 were demented and six were nondemented), and 17 had PD (eight were demented and nine were nondemented). Demographical information is summarized in Table 2.

Procedures.—The neuropsychological battery, as described above, was administered to each subject by a technician who was unaware of the subject's dementia status. The diagnostic paradigm, as outlined in Table 1, was applied to each subject's test results to determine if that subject's performance met criteria for dementia.

Results

The mean age did not differ significantly across groups, but the group with stroke dementia had significantly less years of education (Table 2). The mean performance of the groups on the neuropsychological battery is summarized in Table 2. The t test comparisons of demented and nondemented subjects were significant for every neuropsychological test ($P<.05$ for all). One-way analyses of variance that compared performance of all groups on each test were significant in each case ($P<.05$ for all). Duncan's multiple-range test post hoc comparisons are summarized in Table 2.

On a case-by-case basis, all but two demented subjects met criteria for impairment in memory and two other cognitive categories. The relationship between the diagnosis derived from the neuropsychological battery and the previously established diagnosis was highly significant ($\chi^2=43$, $P<.01$).

There were no false positives, that is, cases where the paradigm identified nondemented subjects as demented. However, two cases of probable Alzheimer's disease were not correctly classified. Both cases had been diagnosed clinically on the basis of a history of progressive intellectual and functional decline; however, these subjects did not perform at a defective level on most tests in the more extensive neuropsychological battery that was used in our Memory Disorders Clinic, and these cases would probably not have been diagnosed as demented based on results of neuropsychological testing alone.

Table 1.—Operational Definitions of the Impairment in Memory and the Two Other Cognitive Categories Required for the Diagnosis of Dementia*

Function	Measure†	Score: Cutoff	Score: Maximum
Memory impairment	Two of the following		
	Immediate verbal		
	Both of the following		
	SRT total recall[1]	<25	72
	SRT long-term retrieval[1]	<15	72
	Remote verbal		
	Both of the following (if delayed recall = 0, then remote verbal is impaired)		
	SRT delayed recall[1]	<4	12
	Delayed recognition	<8	12
	BVRT multiple-choice recognition[2]	<7	10
Impairment in two of the following			
Orientation	Orientation test[3]	<8	10
Construction	One of the following		
	Rosen Drawing Test[4]	<3	5
	BVRT multiple-choice matching[2]	<7	10
Abstract reasoning	One of the following		
	WAIS-R Similarities (age scaled)[8]	<7	19
	Mattis Identities and Oddities[9]	<12	16
Language	One of the following		
	BDAE repetition of high-probability phrases[7]	<7	8
	BDAE complex ideational material[7]	<5	6
	Boston Naming Test[5]	<11	15
	Verbal fluency (one of the following)		
	CFL (percentile score)[6]	≤16th percentile	. . .
	Category naming (mean of three trials)	<12	. . .

*SRT indicates Selective Reminding Test (12 items, six trials); BVRT, Benton Visual Retention Test (10 items); WAIS-R, Wechsler Adult Intelligence Scale–Revised; BDAE, Boston Diagnostic Aphasia Evaluation; and CFL, Controlled Oral Word Association.
†Superscript numbers refer to test references.

These findings suggested that performance on this battery of tests might effectively determine whether a subject had suffered cognitive decline consistent with a diagnosis of dementia.

We did not attempt to account for the effects that strokes might have had on test performance. Similarly, we did not address the application of this approach to Spanish-speaking subjects or those with low levels of education.

STUDY 2: ADMINISTRATION TO COMMUNITY-DWELLING ELDERS

In this study, the neuropsychological paradigm, as described and piloted above, was administered to a large group of elderly subjects who lived in the northern Manhattan (NY) community of Washington Heights Inwood.

Methods

Subjects.—Subjects were selected from volunteers who participated in the Washington Heights Inwood Project, a community-based prospective investigation of Alzheimer's disease and dementias associated with PD and stroke; this project consisted of three subprojects, with coordinated but independent recruitment procedures. In the North Manhattan Aging Project, potential subjects were referred by community-based service providers, and recruitment was limited to a geographically defined area of northern Manhattan. In the North Manhattan Aging Project, subjects who were free of dementia in the study sample were used as controls. The PD project actively solicited potential subjects with PD through hospital surveillance, community-based service providers, and news media. Recruitment was also restricted to northern Manhattan. The Stroke and Aging Research Project consisted of a hospital-based sample of stroke patients. In the Stroke and Aging Research Project, a "control" subject was recruited for each stroke subject; the control was often a spouse, friend, or neighbor, but controls were also solicited by advertising and mailings. For the PD study, controls who lived in northern Manhattan were solicited in a similar manner.

The present study sample was selected from all individuals who participated in the three subprojects and who met the following criteria: subjects with PD or with a history or clinical signs of stroke were excluded so that the potential influence of these diseases on neuropsychological test performance could be eliminated.

We excluded subjects for whom complete data on all neuropsychological tests were not available, since the paradigm and subsequent education correction approaches could not be fully applied to incomplete data. Among those for whom testing was attempted, 6.4% refused to complete one or more of the tests that constituted the battery, and 14.6% were unable to complete one or more of the tests. Severe cognitive impairment appeared to be the reason for failure to complete the test battery; 73% of those subjects who were unable or unwilling to complete the tests were subsequently diagnosed as demented based on the physician's assessment and other clinical data—more than seven times the rate of dementia among those subjects who were able to complete the battery.

Subjects were also required to have completed a semistructured assessment by a physician. All subjects had to be older than 55 years of age, and they had to speak English or Spanish. A total of 430 individuals met all criteria for inclusion in the study sample. Demographic characteristics of the study sample are summarized in Table 3.

Procedures.—The neuropsychological battery, as described above, was administered to each subject. Before administration, an attempt was made to determine if the subject felt she or he would perform better in English or Spanish, and this determined

Table 2.—Performance of Study 1 Patients on Neuropsychological Battery*

	Group†					
	Control	pAD	PD	PD + Dementia	Stroke	Stroke + Dementia
Group demographics						
n	5	12	9	8	11	6
Age, y	63.0 (7.6)	69.4 (6.9)	65.4 (7.4)	69.4 (7.9)	66.6 (11.4)	73.8 (8.5)
Education, y	14.6 (2.4)[a]	15.1 (4.6)[a]	13.8 (2.1)[a]	17.1 (3.6)[a]	12.6 (4.5)[ab]	8.5 (4.7)[b]
Abstract reasoning						
WAIS-R Similarities	11.6 (2.6)[a]	8.6 (2.4)[ab]	10.4 (2.2)[a]	6.4 (4.1)[b]	9.4 (3.8)[ab]	6.5 (1.9)[b]
Mattis Identities	15.6 (0.9)[a]	13.0 (2.4)[a]	15.4 (0.9)[a]	8.5 (7.2)[b]	14.8 (1.5)[a]	12.6 (3.4)[a]
Orientation	9.8 (0.4)[a]	4.0 (2.2)[c]	9.8 (0.5)[a]	4.6 (2.8)[c]	9.2 (0.9)[a]	7.0 (1.8)[b]
Memory						
SRT						
Short-term recall						
Total recall	53.4 (11.2)[a]	25.2 (9.8)[c]	43.1 (6.3)[b]	18.8 (7.6)[c]	37.2 (13.2)[b]	24.0 (5.4)[c]
LT retrieval	47.2 (14.9)[a]	7.1 (5.2)[c]	29.2 (11.2)[b]	4.9 (4.0)[c]	26.1 (13.9)[b]	5.8 (4.2)[c]
Delayed recall						
Recall	9.6 (1.3)[a]	0.5 (1.2)[c]	7.0 (3.4)[b]	0.4 (0.4)[c]	5.6 (3.5)[b]	2.0 (1.8)[c]
Recognition	11.8 (0.4)[a]	7.1 (2.8)[b]	11.4 (1.0)[a]	4.3 (3.7)[c]	11.1 (1.6)[a]	9.8 (1.3)[a]
BVRT, recall	7.7 (1.7)[ab]	5.9 (2.2)[bc]	8.0 (1.1)[ab]	2.9 (3.1)[d]	8.2 (1.6)[a]	5.3 (1.4)[c]
Visuospatial						
Rosen Drawing Test	3.6 (0.9)[a]	2.1 (1.2)[b]	2.6 (1.0)[ab]	0.9 (1.0)[c]	2.6 (1.1)[ab]	1.0 (0.6)[c]
BVRT, matching	9.0 (0.7)[a]	6.4 (2.4)[bc]	8.8 (1.3)[a]	4.4 (3.1)[c]	8.1 (2.3)[ab]	5.7 (1.4)[c]
Language						
fluency						
CFL (raw mean score)	10.5 (2.8)[ab]	9.5 (4.6)[a]	12.0 (2.9)[a]	3.6 (1.8)[c]	10.3 (4.9)[ab]	7.2 (2.4)[bc]
Categories	17.0 (2.9)[a]	10.4 (5.4)[b]	15.4 (5.4)[a]	5.4 (2.7)[c]	15.6 (5.0)[a]	6.3 (1.5)[bc]
Repetition	8.0 (0.0)[a]	7.0 (1.1)[a]	7.6 (0.7)[ab]	6.5 (1.5)[b]	7.6 (0.8)[ab]	6.5 (0.8)[b]
Comprehension	6.0 (0.0)[a]	4.5 (2.0)[ab]	5.7 (0.7)[a]	3.3 (2.1)[b]	5.6 (0.7)[a]	4.8 (1.2)[a]
Naming	14.4 (0.9)[a]	11.8 (4.2)[ab]	14.2 (1.4)[a]	11.0 (2.9)[ab]	13.6 (2.3)[a]	8.8 (4.4)[b]

*For each one-way analysis of variance comparing groups on a single variable that is significant at $P < .05$, superscript characters summarize post hoc comparisons; means with the same letter do not differ significantly. Except for n, data are given as mean (SD). pAD indicates probable Alzheimer's disease; PD, Parkinson's disease; WAIS-R, Wechsler Adult Intelligence Scale–Revised; SRT, Selective Reminding Test; LT, long-term; BVRT, Benton Visual Retention Test; and CFL, Controlled Oral Word Association.

†The superscript alphabet letters summarize post hoc comparisons. The means with similar letters do not differ significantly from each other.

the language in which the test was administered. The paradigm, as described above, was used to determine if each subject met criteria for dementia.

Separate from the neuropsychological testing, a physician completed three measures of functional capacity or activities of daily living: the Blessed Dementia Rating Scale (Part I, Sections A and B),[12] the Schwab and England Activities of Daily Living Scale,[13] and the Barthel Index.[14] For the Barthel Index items, scores were reported simply as the number of assessed activities in which disability was reported. The short version of the Blessed Memory Information and Concentration Test[15] was also administered. The physician used this information, along with that obtained during a medical and neurological examination and elicitation of medical/neurological history, to determine separately whether each subject met criteria for dementia.

Data Analysis.—The major focus of the analyses was on the influence of educational attainment on meeting neuropsychological criteria for dementia. Comparisons of dementia rates by educational ranges, as well as by age and ethnicity, were initially done with χ^2 tests for independence.

To determine the potential impact of education on the neuropsychological test-based diagnoses, the following approach was developed to derive an education-adjusted diagnosis. First, the linear regression equation for the prediction of each test score by education was calculated. For these calculations, subjects whom the neurologist considered to be demented were excluded ($n = 42$), since the relationship between education and performance would be attenuated among demented individuals. We considered adjusting education for the presumed variation in educational attainment in different age groups, but in the present group there was only a minimal relationship between age and education ($r = .21$, $P < .01$); adjusting education for age had no influence on the education-adjusted diagnostic outcome, and we wished to limit the complexity of the adjustment process.

Based on the linear regression equations, each subject's residual score for each test was calculated. This residual score was the difference between the actual score and the score predicted by education in the regression equation. The use of residual scores had two advantages. First, they were, in effect, "education free" in that the correlation between the residual scores and education was 0. Second, the residual scores corrected for education at both ends of the range: subjects with higher levels of education must have performed at a higher level on a test to receive a residual score that was equivalent to that based on a lower raw score in an individual with fewer years of education.

The residual scores for each test were then transformed so that the new distribution had the same mean and SD as the distribution of the original raw test scores. A set of education-adjusted test scores was calculated for each of the 430 subjects, and then the criterion scores that were used in the neuropsychological paradigm were applied. The relationship between a test score, before and after education adjustment, and education is demonstrated in the Figure.

Results

Overall, 45 subjects fulfilled the neuropsychological paradigm criteria for dementia.

As might be expected, there was an increased frequency of in-

Table 3.—Demographic Characteristics of Subjects for Study 2*

Characteristic	Mean/ Frequency	SD/%
Age, y	73.6	7.9
Education, y	10.2	4.7
Score		
Short Blessed (BMIC)	4.9	5.0
BDRS	1.3	1.9
Schwab and England's ADL scale[13]	86.3	18.5
Barthel Sum Index[14]	0.6	1.6
Sex		
M	117	27.2
F	313	72.8
Predominant language spoken		
English	276	64.2
Spanish	130	30.2
Other	22	5.1
Unknown	2	0.5
Race ethnicity		
Non-Hispanic W	158	36.7
Non-Hispanic B	123	28.6
Hispanic	141	32.8
Other/unknown	8	1.9

*Data for age, education, and scores are given as means and SDs; data for sex, language, and race ethnicity, as frequencies and percentages. BMIC indicates Blessed Memory Information and Concentration Test; BDRS, Blessed Dementia Rating Scale; and ADL, Activities of Daily Living.

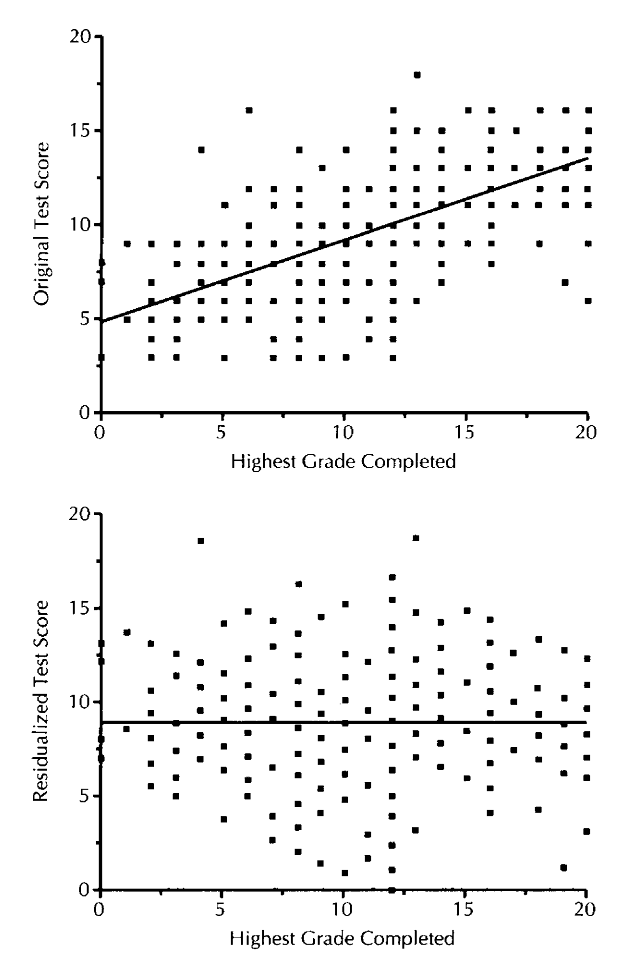
tellectual impairment in older subjects when age was subdivided into five ranges ($\chi^2=36.2$, $P<.01$) (Table 4). The relation between ethnic self-identification and the diagnosis of dementia (Table 5) approached significance ($\chi^2=7.4$, $P<.06$). However, as an analysis described below will indicate, this was a function of the differing educational level for these groups.

There was a strong relationship between years of education, subdivided into four ranges, and dementia, with higher frequencies in those subjects with lower levels of education ($\chi^2=21.0$, $P<.01$) (Table 6).

To evaluate the extent to which this finding was a function of the influence of educational attainment on test performance, we derived an "education-corrected" diagnosis by applying the criterion scores to test scores that were corrected for education by the statistical methods described above. Performance on every test in the neuropsychological battery correlated significantly with education. Correlation coefficients ranged from a low of .22 for repetition to a high of .67 for the Similarities subscale of the Wechsler Adult Intelligence Scale–Revised. As summarized in Table 7, the education correction resulted in a change in diagnosis in 23 cases (5.3%): 11 from not demented to demented and 12 from demented to not demented. Subjects who were reclassified as demented had more years of education: two had some high school, four were high school graduates, and five had at least some college. Conversely, subjects who were reclassified as nondemented all had 8 or less years of education.

In the subjects who were reclassified as demented, six were white, one was black, three were Hispanic, and one classified himself as neither white, black, or Hispanic. In those subjects who were reclassified as nondemented, one was white, six were black, and five were Hispanic. All subjects were at least 70 years old.

To explore how reclassified subjects differed from comparable subjects who were not reclassified, two sets of comparisons were made. Originally nondemented subjects who were reclassified as demented were compared with nondemented subjects who were not reclassified. Similarly, demented subjects who were reclassified as nondemented were compared with demented subjects who remained in that category.



Relationship between the Wechsler Adult Intelligence Scale–Revised Similarities subscale score (age scaled) and educational level of subject before (top) and after (bottom) adjustment for educational level. The line through each figure part represents the least squares regression line for predicting test scores from education. Each point on the graph represents one or more of the 430 subjects. Top, R=*.43. Bottom,* R=*.00.*

Table 8 summarizes comparisons of the 11 subjects who were reclassified from nondemented to demented with 237 selected subjects who remained classified as nondemented. The comparison group was restricted to those in the same education range and to subjects aged 60 years and older. Distributions of ethnicity and language spoken were comparable in the two groups. The two groups differed significantly on the Schwab and England Activities of Daily Living Scale. There was also a large, significant disparity between the two groups on the short version of the Blessed Memory Information and Concentration Test.

In a similar set of comparisons, the 12 subjects who were reclassified from demented to nondemented were contrasted with 19 who remained classified as demented (Table 8). Since no reclassified subjects had more than 8 years of education, subjects with more than 8 years of education were excluded from the comparison group. In addition, the comparison group was restricted to subjects who were at least 70 years old. Neither ethnicity nor language spoken differentiated the two groups, nor did the measures of functional capacity and mental status.

The relation between self-designated ethnicity and the

Table 4.—Distribution of Dementia Diagnosis by Age Ranges*

	Age Range, y					
	≤59	60-69	70-79	80-89	90+	Row Total
Not demented	7	140	163	67	8	**385**
	1.8	36.4	42.3	17.4	2.1	89.5
	87.5	98.6	90.6	76.1	66.7	. . .
Demented	1	2	17	21	4	**45**
	2.2	4.4	37.8	46.7	8.9	10.5
	12.5	1.4	9.4	23.9	33.3	. . .
Column Total	**8**	**142**	**180**	**88**	**12**	**430**
	1.9	33.0	41.9	20.5	2.8	100.0

*Data are given as values for frequencies and row and column percentages. In each cell, row, and then column, percentages are listed beneath the frequency.

Table 5.—Distribution of Dementia Diagnosis by Self-designated Ethnicity*

	Race Ethnicity				
	Non-Hispanic W	Non-Hispanic B	His-panic	Other/ Unknown	Row Total
Not demented	148	103	127	7	**385**
	38.3	26.8	33.0	1.8	89.5
	93.7	83.7	90.1	87.5	. . .
Demented	10	20	14	1	**45**
	22.2	44.4	31.1	2.2	10.5
	6.3	16.3	9.9	12.5	. . .
Column Total	**158**	**123**	**141**	**8**	**430**
	36.7	28.6	32.7	1.9	100.0

*Data are given as values for frequencies and row and column percentages. In each cell, row, and then column, percentages are listed beneath the frequency.

education-corrected diagnosis of dementia was explored with a χ^2 analysis, and it was no longer significant ($\chi^2=3.5$, $P<.33$).

COMMENT

Our results suggest that this neuropsychological paradigm can be used in a large population with a wide range of age, ethnicity, and education. The validity of the paradigm-based diagnoses will require further study, however. In the first part of this study, we compared diagnoses based on the paradigm with those derived by more standard clinical methods, and we found good consistency between the two diagnostic methods. To the extent that the working clinical diagnoses were accurate, this suggests that the neuropsychological paradigm produces valid diagnoses. An accompanying report[16] evaluates the reliability of the neuropsychological paradigm in reference to a clinical diagnosis derived by a physician's standard dementia evaluation, including a mental status screen.

The major focus of the second study was to evaluate factors that could potentially influence test performance and consequently bias a paradigm-based diagnosis. Dementia was more common with advancing age, an expected observation. However, the higher prevalence of dementia in the subjects with fewer years of education is

Table 6.—Distribution of Dementia Diagnosis by Education Range*

	Level of Education				
	≤8 y	Some High School	High School Graduate	Some College	Row Total
Not demented	139	58	80	108	**385**
	36.1	15.1	20.8	28.1	89.5
	81.8	93.5	90.9	98.2	. . .
Demented	31	4	8	2	**45**
	68.9	8.9	17.8	4.4	10.4
	18.2	6.5	9.1	1.8	. . .
Column Total	**170**	**62**	**88**	**110**	**430**
	39.4	14.4	20.4	25.6	100.0

*Data are given as values for frequencies and row and column percentages. In each cell, row, and then column, percentages are listed below each frequency.

Table 7.—Comparison of Original and 'Education-Corrected' Diagnoses

Original Diagnosis	Corrected Diagnosis: Not Demented	Demented	Row Total
Not demented	374	11	**385**
Demented	12	33	**45**

more problematical. Several studies have obtained similar findings.[17,18] It is possible that in individuals with more education, the dementing process must be further advanced before it is clinically detectable, since there is a higher baseline from which intellectual function must decline.[19] Education could also be in the causal pathway for dementia, but this has not been determined. While these concepts are intriguing, it must first be demonstrated that the diagnostic process is not biased against subjects with lower educational levels. For example, if subjects simply cannot comprehend the tasks that they are confronted with because of lack of prior exposure to test materials, then the diagnostic process is at fault.

While the present study cannot definitively address these issues, it does attempt to quantify the relation between educational attainment and the paradigm-based diagnosis. Findings indicate that only a small percentage of individuals, ie, 5%, were potentially misclassified due to education's effects on test performance. However, the relatively large number of low-education subjects who were reclassified as nondemented suggests that the unadjusted neuropsychological paradigm might be overdiagnosing dementia in the lower range of education. In addition, underdiagnosis of dementia may occur in the higher education ranges.

It is simplistic to assume that a statistical manipulation could adequately adjust for the effects of education. Education may serve as a proxy for many other sociocultural variables, including general intelligence, social opportunity, and societal expectations. Our ability to measure educational attainment is poor. The present analyses assumed the equivalence of education no matter where it was obtained. Also, the calculations treated years of education as an interval-level variable in which each year of

Table 8.—Comparison of ADL and Mental Status Scores in Subjects Who Were Reclassified and Education-Matched Subjects Who Were Not Reclassified*

	High Education: Reclassified From Nondemented to Demented?		Low Education: Reclassified From Demented to Nondemented?	
	Yes	No (Still Nondemented)	Yes	No (Still Demented)
Schwab and England's ADL scale[13]	78.2 (22.3)	91.5 (15.1)†	73.3 (21.2)	68.9 (23.5)
BMIC	9.0 (6.6)	2.8 (3.1)†	11.6 (5.4)	14.4 (6.9)
BDRS	2.0 (2.1)	0.6 (1.1)	2.5 (2.2)	3.5 (2.8)
Barthel Sum Index[14]	2.0 (3.6)	0.2 (1.0)	0.4 (1.2)	1.2 (1.7)

*ADL indicates Activities of Daily Living; BMIC, Blessed Memory Information and Concentration Test; and Blessed Dementia Rating Scale. Score values are given as means (SDs).
†$P<.05$ (for t tests).

education contributed an equal amount to total educational attainment. Given these drawbacks, the present analysis might best be viewed as an estimation of the degree to which educational attainment might possibly influence a neuropsychological paradigm-based diagnosis. We therefore do not consider the education correction paradigm, as described here, as definitive, and we await follow-up data on our subject cohort to confirm diagnoses and subsequently to refine the diagnostic paradigm.

One measure in the neuropsychological battery is a percentile score that is education adjusted (controlled oral word association). When the battery was constructed, we found that this score discriminated between demented and nondemented subjects better than the raw score. At that time, we did not anticipate the application of the education correction approach, as reported here. Still, we think that it is appropriate to include this percentile score in the education correction process because it still correlated strongly and significantly with education ($r=.41$, $P<.01$).

We excluded subjects who did not complete the entire neuropsychological battery from this study because it would have complicated the attempt at education correction. In the context of our epidemiological studies, the presumed reason for the subject not completing a test was recorded. If it was clear that the subject was incapable of completing a particular test because of a visual, auditory, or motor impairment, then that test was excluded from diagnostic consideration. Memory and other cognitive impairment in intact sensory modalities were required for the diagnosis of dementia. However, if a subject could not grasp or comply with task demands when there was no mitigating sensory, motor, or language disability, we considered that subject to have scored below the criterion for that test.

Subjects who were reclassified as demented were rated as less able to perform independently and to take care of daily activities than their peers who remained classified as nondemented. This observation adds support to the reclassification and points to the utility of a multidisciplinary approach to diagnosis, with convergent evidence from several domains required for the diagnosis of dementia. Relatively acceptable performance on the tests might be misleading in a poorly functioning individual who is well educated.

This observation was not repeated in the subjects who were reclassified as nondemented. The functional and activities of daily living measures did not differ significantly between the reclassified group and those who remained classified as demented. In theory, measures of function should be less biased by education than measures of cognition, because they typically assess the performance of activities that have little apparent relation to educational status. There is some evidence, however, that these measures are also subject to bias.[20] In any case, the present analyses suggest that ancillary functional testing would not have influenced or modified the neuropsychological paradigm-based diagnosis before application of education correction. One important consideration is that functional capacity might be given different weighting in the diagnostic process in individuals of different educational or cultural background. This issue is addressed in part in an accompanying article.[16]

The correction for education, as reported here, resulted in diagnostic reclassification in only a small percentage of the subjects. This bodes well for the utility of the present neuropsychological battery as part of a multidisciplinary diagnostic approach that uses neuropsychological performance, along with functional, medical, and psychiatric information. However, the preponderance of reclassification of presumably demented individuals suggests that the diagnostic paradigm requires further attention and development.

This work was supported by federal grants AG07370, AG07232, and AG08702 from the National Institute on Aging, Bethesda, Md, and by the Charles S. Robertson Memorial Gift for Alzheimer's Disease from the Banbury Fund.

References

1. Buschke H, Fuld PA. Evaluating storage, retention, and retrieval in disordered memory and learning. *Neurology*. 1974;24:1019-1025.
2. Benton AL. *The Visual Retention Test*. New York, NY: The Psychological Corp; 1955.
3. Folstein MF, Folstein SE, McHugh PR. 'Mini-mental State': a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12:189-198.
4. Rosen W. *The Rosen Drawing Test*. Bronx, NY: Veterans Administration Medical Center; 1981.
5. Kaplan E, Goodglass H. Weintraub S. *Boston Naming Test*. Philadelphia, Pa: Lea & Febiger; 1983.
6. Benton AL, Hamsher K deS. *Multilingual Aphasia Examination*. Iowa City, Iowa; University of Iowa; 1976; revised manual, 1978.
7. Goodglass H, Kaplan D. *The Assessment of Aphasia and Related Disorders*. 2nd ed. Philadelphia, Pa: Lea & Febiger; 1983.
8. Wechsler D. *Wechsler Adult Intelligence Scale–Revised*. New York, NY: The Psychological Corp; 1981.
9. Mattis S. Mental status examination for organic mental syndrome in the elderly patient. In: Bellak L, Karasu TB, eds. *Geriatric Psychiatry*. New York, NY: Grune & Stratton; 1976:77-121.
10. McKhann G, Drachman D, Folstein M, et al. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of the Department of Health and Human Services Task force on Alzheimer's disease. *Neurology*. 1984;34:939-944.

11. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Revised Third Edition.* Washington, DC: American Psychiatric Press Inc; 1987.
12. Blessed G, Tomlinson BE, Roth M. The association between quantitative measures of dementia and of senile changes in the cerebral grey matter of elderly subjects. *Br J Psychol.* 1968;225:797-811.
13. Schwab JF, England AC. Projection technique for evaluating surgery in Parkinson's disease. In: Gillinghan FS, Donaldson MN, eds. *Third Symposium on Parkinson's Disease.* Edinburgh, Scotland: E & S Livingstone; 1969:152-157.
14. Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. *Md Med J.* 1965;14:61-65.
15. Katzman R, Brown T, Fuld P. Validation of a short orientation-memory-concentration test of cognitive impairment. *Am J Psychiatry.* 1983;140:734-738.
16. Mayeux R, Denaro J, Hemenegildo N, et al. A population-based investigation of parkinson's disease with and without dementia: relationship to age and gender. *Arch Neurol.* 1992;49:492-497.
17. Zhang MY, Katzman R, Salmon D, et al. The prevalence of dementia and Alzheimer's disease in Shanghai, China: impact of age, gender, and education. *Ann Neurol.* 1990;27:428-437.
18. Berkman LF. The association between educational attainment and mental status examinations: of etiologic significance for senile dementias or not? *J Chronic Dis.* 1986;39:171-175.
19. Stern Y, Alexander G, Prohovnik I, Mayeux R. Inverse relationship between education and parietotemporal perfusion deficit in Alzheimer's disease. *Ann Neurol.* In press.
20. Snowdon DA, Ostwald SK, Kane RL. Education, survival, and independence in elderly Catholic sisters, 1936-1988. *Am J Epidemiol.* 1989;130:999-1012.