Diagnosis of Dementia in a Heterogeneous Population
A Comparison of Paradigm-Based Diagnosis and Physician’s Diagnosis

John Pittman, MS; Howard Andrews, PhD; Thomas Tatemichi, MD; Bruce Link, PhD; Elmer Struening, PhD; Yaakov Stern, PhD; Richard Mayeux, MD

- Each of 430 subjects received a diagnosis using two independent methods: a test-based quantitative paradigm and a semistructured neurological examination by a physician. The paradigm diagnosis was based on a battery of tests that assessed short- and long-term verbal memory and short-term nonverbal memory, orientation, construction, abstract reasoning, and language. The subjects came from a community in Manhattan County, in New York City, and were characterized by diversity with respect to both ethnicity (29.1% black, 33.4% Hispanic) and educational level (23.5% with 6 or fewer years of education, 25.6% college educated). Based on the paradigm, 10.5% of subjects received a diagnosis of dementia, 29.1% of cognitive impairment, and 60.5% of normal. Based on the physician’s diagnosis, 9.8% were demented, 21.6% cognitively impaired, and 68.6% normal. There was agreement between the two diagnostic methods for 71.8% of subjects. Diagnostic disagreement (n = 121) was in most cases between normal and cognitively impaired (71.0%) or between cognitively impaired and demented (21.5%). There were only nine cases (7.5%) in which a subject was judged demented by one method and normal by the other. The reliability of each method with respect to the other was moderate (intraclass correlation coefficient, .62), while the reliability of a composite diagnosis based on both methods was much higher (.77). The paradigm was more likely than the physician to give the diagnosis of dementia to patients with low educational levels. The physician’s diagnosis was strongly influenced by measures of functioning and by the mental status test administered in the semistructured neurological examination. Race and diagnosis were not related when the effect of education was controlled. Strengths and potential weaknesses of each method of diagnosis, and the relationship between education and diagnosis, are discussed. (Arch Neurol. 1992;49:461-467)

An effective quantitative method for the diagnosis of dementia has been described by Stern and his colleagues in the preceding study. Their method applies an explicit, standardized set of decision rules to the results of a complete neuropsychological battery to arrive at a diagnosis. To be sure, there are a number of potential disadvantages to the diagnosis of dementia based on a quantitative paradigm. First, due to its inflexible nature, the paradigm may generate incorrect diagnoses because of its inability to accommodate anomalous conditions. For example, a subject who is intimidated by testing procedures may be labeled demented by a score-driven paradigm. Second, it would be extremely difficult to develop a paradigm that simulated the judgment of an experienced clinician, who can assess the weight a particular test score or reported symptom should be given, based on the unique characteristics of a particular subject.

At the same time, the use of a quantitative, paradigm-based diagnosis offers several important advantages. First, it ensures a high level of consistency in diagnostic criteria. Second, for many disorders that must be diagnosed primarily on the basis of behavioral criteria, a quantitative, “actuarial” approach to diagnosis has proved superior, in terms of reliability, to diagnosis based on clinical judgment alone. However, before a paradigm-based diagnosis can be widely adopted, its performance must be directly measured against the traditional method of diagnosis for dementia, rendered by a trained clinician. We report on research in which each subject was given a diagnosis using the paradigm and, independently, by a physician. The subjects were drawn from a densely populated commu-

See also p 453.
nity in New York City characterized by diversity with respect to race, education, language, and national origin. In analyzing data from this sample, the questions we addressed were as follows: (1) To what degree does a paradigm-based diagnosis of dementia agree with a diagnosis based on clinical judgment? (2) What demographic factors and clinical features account for discrepancies between the two diagnostic methods? The ultimate aim of this research is to ensure a high degree of accuracy and consistency in the diagnosis of dementia in heterogeneous populations.

**SUBJECTS AND METHODS**

Subjects were selected from volunteers participating in the Washington Heights-Inwood and Columbia Aging Project. The sample used for this research is described in the preceding article.1

**Diagnostic Methods**

At the time of entry into the Washington Heights-Inwood and Columbia Aging Project, the cognitive status of each subject was evaluated by two independent methods: physician's diagnosis and paradigm-based diagnosis. Each method assigned a subject to one of three ordered diagnostic categories: normal, cognitively disordered, or demented. These categories were assigned the quantitative values of 0, 1, and 2, respectively. This three-point ordinal scale provided the basis for assessing consistency between the two diagnostic methods. The decision to rank subjects on a three-point scale rather than a dichotomy was based on our observation that dementia occurs after a progressive deterioration in cognitive functioning from a normal state. Consequently, many persons who manifest cognitive disorder do not meet all conditions required for a diagnosis of dementia. As noted below, criteria for classifying subjects on the three-point scale were clearly defined for both the paradigm and physician.

The physician's diagnosis of each subject was based on a semi-structured medical and psychiatric history, a physical and neuropsychological examination, and the Short Blessed Information-Memory-Concentration test (BIMC).2 The physician reviewed three measures of activities of daily living (ADL): the Schwab,4 a modified Barthel,5 and the Blessed Functional Activity Scale (BFAS).6 parts 1 and 2. Using the results of the examination and the three ADL scales, the physician, based on clinical judgment, assigned each subject to one of the three diagnostic categories indicated above.

The second diagnostic method was based on a battery of neuropsychological tests. A diagnostic paradigm, using criterion scores on each test, as well as a pattern of performance across tests, determined the diagnostic category to which each subject was assigned. Based on the Diagnostic and Statistical Manual of Mental Disorders, Revised Third Edition definition of dementia,8 in which memory is the key defining feature, a necessary condition for a diagnosis of dementia in the paradigm is poor performance on two of the three memory “areas” assessed: short- and long-term verbal and short-term nonverbal. In addition, performance on at least two of the following areas must fall below criterion scores: orientation, construction, abstract reasoning, and language. Subjects who did not meet the above requirements but who scored below criterion levels on two or more individual tests were classified as cognitively impaired. Details concerning the types of neuropsychological tests used in the paradigm are provided in the preceding article.1

**Statistical Analysis**

The intraclass correlation coefficient (ICC) was chosen as the index of reliability. Following guidelines suggested by Shrout and Fleiss,7 we selected two forms of ICC. The first form, the Shrout and Fleiss ICC(3,1), is an index of consistency and, in effect, measures the reliability of either single diagnostic method with respect to the other method. If the reliability is found to be high using this approach, say, .85 or higher, then the two diagnostic methods can be used interchangeably. We will refer to this measure of reliability as model A. Second, we calculated the composite reliability of both diagnostic methods using the Shrout and Fleiss ICC(3,k). This approach, which will be referred to as model B, assesses diagnostic reliability when information from both diagnostic methods is pooled to yield a single composite diagnosis. If the reliability of model A is low, then model B can be used to determine if reliability can be improved to an acceptable level by employing a composite diagnosis.

Calculation of both ICCs is based on a fixed-effects, subject-by-diagnostic method two-way analysis of variance (ANOVA). See Table 1 for sources of variation and for equations used to calculate reliability using models A and B.

To explore the sources of agreement and disagreement between the physician and paradigm, each subject was assigned to one of nine agree-disagree categories based on the unique combination resulting from the two diagnostic methods (eg, normal by physician's diagnosis and normal by paradigm-based diagnosis; normal by physician's diagnosis and cognitively disordered by the paradigm). In addition, each subject was assigned a composite diagnosis resulting from the sum of the values associated with each diagnostic method. For example, a person diagnosed as cognitively disordered (1) by the physician and demented (2) by the paradigm would have a composite diagnosis of 1 + 2 = 3. The composite score ranged from a value of 0 (diagnosis of normal by both methods) to 4 (demented by both methods). However, for all analyses involving the agree-disagree groups or the composite diagnosis, we eliminated the nine subjects for whom there was maximal disagreement between the two diagnoses: demented according to one, normal according to the other (Table 2). These subjects were excluded because (1) the number of subjects in this category was too small for statistical analysis and (2) the complete lack of agreement between the two diagnostic methods for these few subjects represents a set of anomalous findings that yield the most helpful information when examined adocdatally (see the “Results” section).

Subjects in the seven agree-disagree groups were compared on gender, age, years of education, self-reported race, language, BIMC, Schwab, modified Barthel, and BFAS using one-way ANOVA for continuous variables and $x^2$ for categorical variables. Subjects in the five composite diagnosis groups were analyzed in the same manner. Three mutually exclusive categories of race-ethnicity were defined: white, not of Spanish/Hispanic origin; black, not of Spanish/Hispanic origin; and Spanish/Hispanic in origin. A modified form of the Barthel was used to determine whether, in the prior month, the subject needed the help or supervision of another person to complete each of nine activities: eating, dressing, bathing, grooming, getting from bed to chair, walking 45 yd, bladder control, bowel control, and getting in or out of a car, bus, or subway. The yes (1) or no (0) responses were summed to form a score ranging between 0 and 9.

### Table 1. Analysis of Variance of Diagnoses

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between subjects (BMS)</td>
<td>429</td>
<td>0.7304</td>
</tr>
<tr>
<td>Within subjects (WMS)</td>
<td>430</td>
<td>0.1721</td>
</tr>
<tr>
<td>Between methods (IMS)</td>
<td>1</td>
<td>1.6791</td>
</tr>
<tr>
<td>Residual (EMS)</td>
<td>429</td>
<td>0.1686</td>
</tr>
<tr>
<td>Total</td>
<td>859</td>
<td>0.4509</td>
</tr>
</tbody>
</table>

*Calculation of intraclass correlation coefficients: reliability of either single diagnostic method, $(BMS - EMS)/(BMS + (k-1)EMS) = (0.7304 - 0.1686)/(0.7304 + (2 - 1)0.1686) = 0.6249$; reliability of composite diagnosis, $(BMS - EMS)/(BMS + (k-1)EMS) = (0.7304 - 0.1686)/(0.7304 + 0.1686) = 0.7692$. 

---

1. Analysis of Variance of Diagnoses.
Table 2.—Joint Distribution of Physician's Diagnosis and Paradigm Diagnosis*

<table>
<thead>
<tr>
<th>Paradigm Diagnosis</th>
<th>Normal</th>
<th>Cognitive Deficit</th>
<th>Demented</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>53.5 (230)</td>
<td>6.5 (28)</td>
<td>0.5 (2)</td>
<td>60.5 (260)</td>
</tr>
<tr>
<td>Cognitive deficit</td>
<td>13.5 (58)</td>
<td>12.3 (53)</td>
<td>3.3 (14)</td>
<td>30.2 (130)</td>
</tr>
<tr>
<td>Demented</td>
<td>1.6 (7)</td>
<td>2.8 (12)</td>
<td>6.0 (26)</td>
<td>9.3 (40)</td>
</tr>
<tr>
<td>Total</td>
<td>68.6 (295)</td>
<td>21.6 (93)</td>
<td>9.8 (42)</td>
<td>100.0 (430)</td>
</tr>
</tbody>
</table>

*Values are percent (number).

Table 3.—Analysis of Variance of Agree-Disagree Groups*

<table>
<thead>
<tr>
<th>Diagnostic Method</th>
<th>No.</th>
<th>Age (n = 421)</th>
<th>Education (n = 421)</th>
<th>BIMC (n = 389)</th>
<th>BFAS (n = 414)</th>
<th>Barthel (n = 407)</th>
<th>Schwab (n = 342)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physician Paradigm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal Normal</td>
<td>230</td>
<td>70.9</td>
<td>12.0</td>
<td>2.6</td>
<td>0.3</td>
<td>0.2</td>
<td>93.5</td>
</tr>
<tr>
<td>Normal CD</td>
<td>58</td>
<td>75.2</td>
<td>8.4</td>
<td>3.8</td>
<td>0.9</td>
<td>0.7</td>
<td>87.4</td>
</tr>
<tr>
<td>CD Normal</td>
<td>28</td>
<td>74.3</td>
<td>9.6</td>
<td>4.8</td>
<td>1.1</td>
<td>0.9</td>
<td>84.4</td>
</tr>
<tr>
<td>CD CD</td>
<td>53</td>
<td>75.5</td>
<td>7.2</td>
<td>8.2</td>
<td>1.4</td>
<td>1.2</td>
<td>80.4</td>
</tr>
<tr>
<td>CD Demented</td>
<td>12</td>
<td>77.2</td>
<td>4.5</td>
<td>9.8</td>
<td>2.3</td>
<td>1.3</td>
<td>70.0</td>
</tr>
<tr>
<td>Demented CD</td>
<td>14</td>
<td>80.7</td>
<td>9.5</td>
<td>13.4</td>
<td>2.4</td>
<td>3.0</td>
<td>63.8</td>
</tr>
<tr>
<td>Demented Demented</td>
<td>26</td>
<td>82.4</td>
<td>8.0</td>
<td>17.1</td>
<td>3.2</td>
<td>1.3</td>
<td>66.7</td>
</tr>
<tr>
<td>F ratio</td>
<td></td>
<td>15.9</td>
<td>18.8</td>
<td>88.0</td>
<td>34.1</td>
<td>10.5</td>
<td>18.5</td>
</tr>
</tbody>
</table>

*BIMC indicates short Blessed Information-Memory-Concentration test; BFAS, Blessed Functional Activity Scale; and CD, cognitive deficit. All F ratios, P<.0001.

RESULTS

Based on the paradigm, 10.5% of the subjects were given the diagnosis of being demented; 29.1%, cognitively impaired; and 60.5%, normal. Based on the physician's diagnosis, 9.8% were demented; 21.6%, cognitively disordered; and 68.6%, normal. The joint distribution of physician's diagnosis and paradigm-based diagnosis is presented in Table 2. Of the 430 subjects given diagnoses, there was agreement on 309 (71.8%). Of the 121 subjects on whom there was disagreement, 77 (63.6%) were judged more impaired by the paradigm than by the physician; 44 (36.3%) were found more impaired by the physician than by the paradigm. Eighty-six (71.0%) of the 121 subjects on whom there was diagnostic disagreement were found in the normal vs cognitively disordered categories. Of these 86 subjects, 58 (67.4%) were judged normal by the physician but cognitively disordered by the paradigm. Twenty-six (21.5%) of the 121 subjects on whom there was diagnostic disagreement were in the cognitively disordered vs demented categories. In this case, the disagreements were nearly evenly distributed: 12 (46.0%) were judged demented based on the paradigm but cognitively disordered by the physician; 14 (54%) were judged demented by the physician but cognitively disordered based on the paradigm. Maximum disagreement, in which physician and paradigm diagnoses were two positions apart on the three-point scale, occurred with only nine (2%) of the 430 subjects given diagnoses. In seven of these nine subjects, diagnosis was dementia by the paradigm and normal by the physician.

Reliability

The results of the ANOVA and derived ICCs are presented in Table 1. For model A, the measure of reliability of either single diagnosis, moderate consistency was found between physician and paradigm, as indicated by a reliability coefficient of .62. The reliability of the composite diagnosis, calculated in model B, was substantially higher: .77. This is close to the .85 value considered acceptable for clinical tests and is an adequate level of reliability for research purposes.

Agreement and Disagreement Between Diagnostic Methods

A total of seven categories, representing combinations of paradigm and physician diagnoses, were examined (Table 3). Two of these represented disagreement as to whether a subject was normal or cognitively impaired; two reflected disagreement as to whether a subject was demented or cognitively impaired. The remaining three categories represented agreement on dementia, cognitive deficit, and normal status.

With respect to education, subjects for whom there was agreement on a normal diagnosis were significantly better educated than subjects in any of the six other groups. This group had, on average, nearly 2.5 more years of education than subjects in the next most educated group. With respect to disagreement, there was a marked trend in educational level. Within each pair of categories for
which there was disagreement, mean educational level was lower in the group judged more impaired by the paradigm. Of particular interest is the pair in which the disagreement was over whether the subject was demented or had a cognitive deficit. For the group in which the paradigm indicated that the subject was demented, while physician diagnosis was of cognitive deficit, the mean educational level was significantly lower than in any other group, 4.5 years. For those with an identical set of discrepant diagnoses, but rated demented by the physician and as having a cognitive deficit by the paradigm, mean educational level was substantially higher: 9.5 years. These findings suggest that the test-based paradigm may be overestimating the level of cognitive impairment in subjects with little formal education.

With respect to mental status, as measured by the BIMC, those judged demented by the physician but cognitively impaired by the paradigm scored significantly higher, i.e., more impaired, on the BIMC than subjects judged demented by the paradigm but cognitively impaired by the physician. Although not statistically significant, the same tendency was found in the two groups where the disagreement was normal vs cognitively impaired: those whom the physician judged more impaired tended to score higher on the BIMC. These findings suggest that the physician was more influenced by the BIMC than was the paradigm.

**Individual Cases of Maximum Disagreement**

As indicated above, there were nine cases in which a subject was judged normal by one method and demented by the other method. In two cases, the subject was demented according to the physician, normal by the paradigm. In both instances the subject performed well on the neuropsychological tests, which are the basis of paradigm diagnosis. In one of the cases, the BFAS was performed poorly, and the physician thought the patient was borderline but eventually decided on a diagnosis of dementia. In the second case, the subject performed poorly on the BIMC, which, as shown above, was highly influential for the physician, and was judged demented by the physician.

In seven cases, the subject was judged demented by the paradigm and normal according to the physician. In one of these cases, the physician concluded that the patient tested poorly due to a psychiatric condition. In three other cases, the patient performed quite well on the BFAS and reported few functional problems, resulting in a diagnosis of normal by the physician. In the three remaining cases, the subject performed somewhat more poorly on the BFAS (a score of 10 or greater), and there were some functional complaints, but the physician believed there was no cognitive change. In all seven cases, these subjects performed very poorly on the neuropsychological battery and were diagnosed demented by the paradigm. Three of these seven had very low levels of education.

These anecdotal findings are consistent with the results of statistical analyses focused on disagreement between diagnoses rendered by the two methods: those with little education tended to be judged more severely by the paradigm, while those who did not perform well on the BIMC and/or who showed evidence of functional deficit tended to be judged more harshly by the physician—even when performance on the neuropsychological battery was acceptable.

**Characteristics Associated With Severity of Composite Diagnosis**

The composite diagnosis is given as a function of age, education, the BIMC, and three ADL measures in Table 4. The ordered scale of the composite diagnosis, from 0 (agreement that the subject is normal) to 4 (agreement that the subject is demented), reflects increasing severity of cognitive impairment. The agreement categories (composite score, 0, 2, and 4) are identical in Tables 3 and 4, while the four disagreement categories in Table 3 are collapsed into two categories in Table 4.

As shown in Table 4, there was a progressive increase in age with severity of cognitive disorder. Those judged as normal by both methods were significantly younger than subjects in any other group, while those judged as demented by both methods were significantly older. Subjects in each composite diagnostic category also performed as expected in terms of mental status: there was a significant increase in BIMC score (i.e., decrease in performance ability) with increasing severity of diagnosis. Similarly, functional capacity, as measured by the BFAS and Schwab, declined with increasing severity of composite diagnosis, from agreement on normal (0) to agreement on dementia (4). The Barthel Index also indicated a decline in functioning with increased severity of composite diagnosis, with the exception that those with a composite diagnosis of severe cognitive disorder were rated more functionally impaired than those subjects judged to be demented. The fact that the composite diagnosis yields results that are consistent with what is known about the relationship among dementia, age, mental status, and functioning indicates that this diagnosis has good construct validity in addition to reliability.

---

*Abbreviations are as in Table 3. All F ratios, P<.0001.
However, while educational level differed significantly between the diagnostic groups, there was no clear ordering of educational level with respect to composite diagnosis. For example, those with the most severe composite diagnosis—agreement with respect to dementia—were intermediate with respect to the other groups in terms of education, with a mean of 8 years.

Relationships between composite diagnosis and categorical demographic variables are given in Table 5. There was a statistically significant relationship between composite diagnosis and race-ethnicity: most striking is the fact that 71% of white subjects but only 47.5% of black subjects and 44.3% of Hispanic subjects were judged to be normal. There was also a significant relationship between diagnosis and the language spoken in the subject’s home: as would be expected from the findings with respect to ethnicity, 60.1% of English-speaking subjects but only 42.6% of Spanish-speaking subjects were judged to be normal. There was no relationship between gender and diagnosis.

The strong relationship between ethnicity and diagnosis, a finding that has been reported in previous research as well, was examined further to determine whether a confounding variable, such as age or education, might explain the observed relationship. A one-way ANOVA indicated that the mean educational level for whites (13.0 years) was significantly greater than for either blacks (9.9 years) or Hispanics (7.2 years). An analysis was therefore performed to examine the relationship between race and diagnosis while controlling for the effect of education. The subject’s age was also controlled in the analysis because of the strong relationship that exists between age and diagnosis (Table 4).

Two logistic regression models were developed using a dichotomous diagnostic outcome: agreement on normal vs all other diagnoses. In the first model, two indicator variables representing minority status (black, 1; other, 0; and Hispanic, 1, non-Hispanic, 0) were entered. Both race and Spanish origin were significantly related to the outcome ($P<.001$), confirming the results described above, based on the $\chi^2$ statistic. In the second model, age and education, the potential confounding variables, were entered in the first step. As expected, both age and education were significantly related to the outcome ($P<.001$). A forward stepwise selection procedure was then used in an attempt to introduce the minority status variables into this second model. Neither variable was “allowed” to enter, using a .05 probability of a score statistic for variable entry. This indicates that when age and education were statistically controlled, minority status was unrelated to diagnosis. Thus, the tendency to diagnose white subjects as normal at a higher rate than minority subjects appears to reflect the lower educational level of members of minority groups in our study sample, rather than a genuine relationship between race and diagnosis.

**Strength of Relationships**

Further to assess the strength of relationship between each characteristic and the individual diagnoses, the ANOVAs were repeated separately for the physician and paradigm. In each of these ANOVAs, there were three classification categories: normal, cognitively impaired, and demented. The results are given in Table 6. Because of the moderate agreement between the two diagnostic methods, demographic and other characteristics that were associated with the outcome of one method of diagnosis tended to be associated in the same direction with the other diagnosis. As with the composite diagnosis, the relationship between diagnosis rendered by each method and each of the characteristics examined was highly significant.

However, there were distinct differences between diagnoses with respect to $R^2$, the proportion of variability in diagnosis that can be accounted for by a key characteristic. Age was related to physician diagnosis about as strongly as to paradigm diagnosis. However, education was more strongly associated with paradigm diagnosis ($R^2 = .18$) than with physician diagnosis ($R^2 = .11$). Performance on the memory status test (BIMC) was more strongly associated with each type of diagnosis than was any other variable examined. However, the relationship between the BIMC and physician diagnosis ($R^2 = .53$) was much stronger than the relationship between BIMC and paradigm diagnosis ($R^2 = .37$). Finally, all three measures of functional impairment were more strongly associated with physician diagnosis than with paradigm diagnosis.

In Table 6, the strength of relationship between each key characteristic and the composite diagnosis categories (from Table 3) is also given (row 3). In each case, the relationship between the characteristic and these categories is stronger than for the three categories of each diagnosis taken separately (rows 1 and 2). This is due at least in part to the increased reliability of the composite diagnosis: because there is less error variance associated with the composite than with the individual diagnoses (Table 1), there is more reliable variance available for association with key subject characteristics.

**COMMENT**

We have used ICC and other procedures to assess the relationship between independent diagnoses rendered by physicians’ assessment and neuropsychological testing.

---

*Table 5.—Association of Composite Diagnosis With Demographics*

| Composite Diagnosis | Race-Ethnicity | | | Language | | | Gender |
|---------------------|---------------|--------------|--------------|--------------|---------------|---------------|
|                     | W  | B  | Hispanic | English | Spanish | M  | F  |
| Normal              | 71.0 (110) | 47.5 (56) | 44.3 (62) | 60.1 (161) | 42.6 (55) | 53.9 (62) | 54.9 (168) |
| Mild CD             | 14.8 (23)  | 28.0 (33) | 20.7 (29) | 19.8 (53)  | 21.9 (28) | 24.3 (28) | 19.0 (58)  |
| CD                  | 5.8  (9)   | 7.6  (9)  | 23.6 (33) | 6.7  (18)  | 24.8 (32) | 11.3 (13) | 13.1 (40)  |
| Severe CD           | 3.9  (6)   | 7.6  (9)  | 5.7  (8)  | 6.3  (17)  | 6.2  (8)  | 3.5  (4)  | 7.2  (22)  |
| Demented            | 4.5  (7)   | 9.3  (11) | 5.7  (8)  | 7.1  (19)  | 4.7  (6)  | 7.0  (8)  | 5.9  (18)  |

*CD indicates cognitive deficit. Race-ethnicity, $x^2 = 42.95, P < .0001$; language, $x^2 = 28.48, P < .0001$; and gender, $x^2 = 3.43, P < .49$. Values are expressed as percent (number).
One purpose of this investigation is to establish a method to assess and refine a neuropsychological diagnostic paradigm for cognitive deficit, including dementia. A second goal is to identify strategies for maximizing diagnostic reliability, using appropriate information from both physicians' assessment and neuropsychological testing.

**Physician vs Paradigm Diagnosis**

In 71.8% of the cases examined, there was diagnostic agreement between the two methods. This level of consistency is reflected in a moderate ICC, using model A (.62). This value may be interpreted as the proportion of reliable variance in either of the diagnostic methods, with respect to the other. From one standpoint, these findings indicate that the current version of paradigm-based diagnosis is not acceptable because it yields results that are not sufficiently consistent with the traditional method used to diagnose cognitive deficit. On the other hand, the reliability obtained between the traditional diagnosis and the current version of the paradigm is high enough to anticipate that further refinement of the paradigm could raise reliability to an acceptable level.

In fact, the level of consistency obtained in the results is probably an underestimate, since subjects who did not represent a challenge for the diagnostician were not included in the study sample. The sample did not include subjects who were too demented to complete one or more of the neuropsychological tests. Had these unambiguously demented subjects been included in the research sample and assigned failing scores on the tests they were unable to complete, the percentage of agreement between physician and paradigm diagnoses, and the value of the ICC for model A, would have been even higher. The fact that level of agreement was so high even in a sample in which the difficulty of rendering an accurate diagnosis was elevated is particularly encouraging.

**Reliability of the Composite Diagnosis**

When there is at least some consistency (positive correlation) between diagnostic methods, a composite diagnosis will always yield greater reliability than a diagnosis rendered by one method alone. In research, one consideration in deciding between using a single method and a composite diagnosis is whether the "benefit" of the composite diagnosis in terms of increased reliability is worth the additional "cost" of examining each patient using two different and independent diagnostic methods. For our research, we believe that the single diagnosis reliability (model A) is too low and that the increase in reliability of the composite diagnosis (model B) is worth the cost of obtaining two independent diagnoses.

**Discrepancies Between Diagnoses**

While the composite diagnosis has good reliability and construct validity, one goal of this research effort is to improve the reliability and accuracy of a paradigm-based diagnosis. The analysis of cases in which there were discrepancies between physician’s and paradigm-based diagnoses yielded two related findings that suggest possible shortcomings in the current version of the paradigm. First, when there were discrepancies in diagnosis, those with low educational levels tended to be judged more impaired by the paradigm than by the physician. The results also indicate that the educational level of the subject was more strongly associated with paradigm than with physician diagnosis.

In the absence of evidence from postmortem pathological examination, it is not possible to determine on a case-by-case basis which of the discrepant diagnoses is the correct one. It might be argued that the paradigm tends to overdiagnose dementia among those with little formal education, and that this shortcoming is particularly problematic in samples like ours, which contains a substantial number of subjects with little or no formal education. It would be expected, according to this argument, that subjects who have no organically based cognitive deficit, but who lack rudimentary test-taking skills, tend to perform poorly on the neuropsychological tests, and that some would obtain scores low enough to meet paradigm-based criteria for cognitive deficit or even dementia. When the diagnosis is made by a physician, on the other hand, the subject's educational level can be weighed in making a final determination of cognitive status. A quantitative method for adjusting the paradigm score for educational level is presented in an accompanying study.

A relationship between low educational level and the diagnosis of dementia or cognitive deficit has been observed in previous research. However, several investigators have pointed out that no findings to date have ruled out the possibility that rates of dementia are truly higher among those with little or no education than among the better educated. Also, it is important to note that the physician as well as the paradigm tended to rate as demented or as having a cognitive deficit a disproportionately large number of subjects with little or no education. Thus, the only firm conclusions that can be drawn from the observed disparities between physician and paradigm diagnosis are (1) that those with low educational levels are more likely to be given diagnoses of dementia; (2) that there is a stronger association between paradigm diagnosis and education than between physician diagnosis and education; and (3) that there is clearly a need for further research focused on the assessment of dementia among subjects with little or no formal education. A second source of disparity between physician and paradigm diagnoses is the functional capacity of the subject. The version of the paradigm used in this research included no measures of the subject's day-to-day functional capabilities. Physicians, on the other hand, had access to three ADL measures. The findings we obtained indicate...
that physicians were incorporating information about functioning in making a final diagnosis: subjects rated more cognitively impaired by the physician than by the paradigm evidenced poor functioning on all three measures, and all measures of functional assessment were more closely associated with physician diagnosis than with paradigm diagnosis. This interpretation is consistent with findings reported in the preceding article.1

The analysis of the cases in which there were discrepant diagnoses also suggests that physicians were strongly influenced by the BIMC, the mental status test administered in the semistructured interview on which the physician's diagnosis was based. Among those subjects judged demented by the physician but as having a cognitive disorder by the paradigm, the BIMC score was worse (mean, 13.4) than among those subjects judged demented by the paradigm but as having a cognitive disorder by the physician (mean, 9.8). Moreover, the association between BIMC score and diagnosis, as measured by $R^2$, is much stronger for the physician than for the paradigm (Table 6).

Finally, the finding of a strong influence of the BIMC on physician diagnosis suggests a possible deficiency with respect to this method of diagnosis. Since the BIMC is the only mental status test administered in the semistructured examination, it is understandable that the subject's performance on this test would be a key factor in the physician's diagnosis. On the other hand, performance on the BIMC is not always indicative of performance on the full battery of neuropsychological tests required for a paradigm diagnosis, as indicated in the analysis of the two cases in which the physician indicated the subject was demented and the paradigm determined the subject was normal. The neuropsychological tests, which are not part of the semistructured examination, measure a broad range of cognitive abilities. A physician's diagnosis may be too dependent on the results of a single mental status test.

**Conclusions**

Taking the results presented herein together with the findings described in the preceding article, there is good reason to believe that a modified version of a paradigm-based diagnosis for cognitive deficit and dementia would be a viable tool for diagnosing dementia in a heterogeneous population. Reliability, and probably the validity of the paradigm, could be improved by adding one or more measures of function to the battery of neuropsychological tests. The results also suggest that the physician's diagnosis might be improved by using additional tests of cognitive performance in the semistructured examination.

However, using the two methods as they are currently defined, reliability will be maximized when a paradigm-based diagnosis is used in conjunction with, rather than instead of, a physician's diagnosis. The strength of the current paradigm lies in the comprehensive quantitative measure of cognitive status in the neuropsychological battery. A physician's diagnosis is valuable because functional measures are considered, and because there appears to be at least some adjustment made for the educational level of the subject. In creating a composite diagnosis, the strength of each method is utilized and the impact of each method's weaknesses is attenuated.

This work was supported by grants AG07232 and AG08702 from the National Institute on Aging, Bethesda, MD, by grant NS26179 from the National Institute of Neurological Disorders and Stroke, Bethesda, and by the Charles S. Robertson Memorial Gift for Alzheimer's Disease, New York, NY.

We thank Anne Palestoni for developing and managing the Washington Heights–Inwood and Columbia Aging Project database and Bianca Williams for data entry and assistance in preparation of the manuscript.

**References**


