Clinical Criteria for the Diagnosis of Vascular Dementia

A Multicenter Study of Comparability and Interrater Reliability

Helena C. Chui, MD; Wendy Mack, PhD; J. Edward Jackson, MD†; Daniel Mungas, PhD; Bruce R. Reed, PhD; Jared Tinklenberg, MD; Fen-Lei Chang, MD; Kate Skinner, MD, PhD; Cora Tasaki, MD; William J. Jagust, MD

Background: Several clinical criteria have been developed to standardize the diagnosis of vascular dementia (VaD). Significant differences in patient classification have been reported, depending on the criteria used. Few studies have examined interrater reliability.

Objective: To assess the concordance in classification and interrater reliability for the following 4 clinical definitions of VaD: the Hachinski Ischemic Score (HIS), the Alzheimer Disease Diagnostic and Treatment Centers (ADDTC), National Institute of Neurological Disorders and Stroke—Association Internationale pour la Recherche et l’Enseignement en Neurosciences (NINDS-AIREN), and Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV).

Methods: Structured diagnostic checklists were developed for 4 criteria for VaD, 2 criteria for Alzheimer disease (AD), and 4 criteria for dementia. Twenty-five case vignettes, representing a spectrum of cognitive impairment and subtypes of dementia, were prepared in a standardized clinical format. Concordance in case classification using different criteria and interrater reliability among 7 ADDTCs given a specific set of criteria was assessed using the $\kappa$ statistic.

Results: The frequency of a diagnosis of VaD was highest using the modified HIS or DSM-IV criteria, intermediate using the original HIS and ADDTC criteria, and lowest using the NINDS-AIREN criteria. Scores for interrater reliability ranged from $\kappa = 0.30$ (ADDTC) to $\kappa = 0.61$ (original HIS).

Conclusions: Clinical criteria for VaD are not interchangeable. Depending on the criteria selected, the reported prevalence of VaD will vary significantly. The traditional HIS has higher interrater reliability than the newer criteria for VaD. Prospective longitudinal studies with clinical-pathological correlation are needed to compare validity.

Arch Neurol. 2000;57:191-196
MATERIALS AND METHODS

STUDY DESIGN

Several criteria for VaD, AD, and dementia were selected for study. Twenty-five cases were chosen to represent a spectrum of severity of cognitive impairment and subtypes of dementia. The cases were prepared in a uniform format, then rated at 7 independent ADDTCs using a diagnostic checklist. A computerized algorithm was used to generate the final clinical diagnoses. Comparisons were made between classifications resulting from the application of various diagnostic criteria and between raters (centers) given specific diagnostic criteria.

CLINICAL CRITERIA

We studied 4 clinical criteria for VaD, 2 clinical criteria for AD, and 4 definitions of dementia. The clinical criteria were formatted as itemized checklists to eliminate unfamiliarity as a source of misclassification, as follows:

1. NINCDS-ADRDA criteria for probable and possible AD.16
2. DSM-IV criteria for dementia of the Alzheimer type.16
3. HIS (original, ≥7; or modified score, ≥4)7
4. ADDTC criteria for probable and possible IVD.12
5. NINDS-AIREN criteria for probable and possible VaD.15
6. DSM-IV criteria for dementia of the vascular type.16

The 4 definitions for VaD differ significantly in their inclusion criteria, particularly regarding their requirement for focal neurologic signs, neuroimaging evidence of stroke or cerebrovascular disease, or causal relation between stroke and dementia (Table 1). Focal neurologic signs receive 2 points according to the HIS, but are not necessarily required for diagnosis of MID. Focal neurologic signs or evidence of significant cerebrovascular disease (by history, examination, or neuroimaging evidence) suffice for a DSM-IV diagnosis of dementia of the vascular type. Evidence of infarction on neuroimaging study is required for a diagnosis of probable (but not possible) IVD using ADDTC criteria. Focal neurologic signs, evidence of cerebrovascular disease on neuroimaging studies, and a causal relation between stroke and dementia are required for a diagnosis of NINDS-AIREN probable (but not possible) VaD.

Definitions of VaD also differ in their exclusion criteria. The HIS does not use exclusion criteria. Delirium is excluded by the ADDTC, NINDS-AIREN, and DSM-IV guidelines. In addition, severe aphasia, AD, or other brain disorder sufficient to cause dementia preclude an NINDS-AIREN diagnosis of VaD. Thus, the DSM-IV criteria and the HIS are the most lenient, the ADDTC criteria are intermediate, and the NINDS-AIREN criteria are the most stringent with regard to inclusion and exclusion criteria.

SAMPLE

The sample consisted of 25 case vignettes selected at the coordinating ADDTC, located at Rancho Los Amigos Medical Center–University of Southern California, Los Angeles. Subjects underwent a comprehensive interdisciplinary evaluation for memory loss. There were 13 men and 12 women (mean ± SD age, 67.8 ± 9.1 years). Of the 25 subjects, 4 were cognitively healthy; 4, cognitively impaired but not demented; 14, mildly demented; and 3, moderately demented. Six subjects had primary degenerative
dementia; 6, strokes; 3, parkinsonian features; 3, depressed mood; and 7, atypical presentations (including prominent personality change [n = 3], language disturbance [n = 1], apraxia [n = 1], visual disorientation [n = 1], and history of encephalitis [n = 1]). The initial mean (±SD) Folstein Mini-Mental State Examination (MMSE) score was 22.0 ± 6.6.

The vignettes included a narrative of the chief complaint, history of present illness, medical history, family history of dementia, results of physical and neurologic examination, MMSE,24 Blessed Memory Information Concentration Test,25 and laboratory test results (including complete blood cell count; sequential multiple analyzer for levels of sodium, potassium, carbon dioxide, chloride, glucose, serum urea nitrogen, and creatinine; thyroid functions; levels of vitamin B	extsubscript{12}; and testing for microhemagglutination—TREPONEMA PALLIDUM). The results of neuropsychological testing were included for 21 subjects. The results of neuroimaging studies (computed tomography or magnetic resonance imaging) were summarized for 24 of the 25 subjects (not available for 1 subject who was cognitively healthy but depressed). In 20 subjects, a composite of representative or key neuroimages (eg, showing an infarct or asymmetric pattern of atrophy) was also included.

RATERS

The participating ADDTCs are university affiliated and highly experienced in the diagnosis of dementia and its subtypes. The center responsible for preparing the case vignettes did not participate in rating them. At each of the remaining 7 centers, the diagnostic ratings were completed during a consensus team conference, which included a geriatric neurologist, psychiatrist, or internist. For the HIS, the teams were instructed to check a box if any of the 13 items was present. For the other criteria, the raters were asked to check yes, no, or don’t know regarding the presence or absence of each key element. A response of don’t know occurred infrequently (106/4725 [2.2%]) and usually indicated that despite the best available information, a determination could not be made (eg, unable to determine whether there was a temporal relation between the stroke and the onset of dementia). The coding sheets (7 centers × 25 vignettes = 175 rating sheets) were sent to the coordinating ADDTC for data entry and statistical analysis.

STATISTICAL ANALYSES

Responses for each item were scored using computer algorithms to arrive at a diagnosis (yes or no) of dementia, AD, or VaD. In the algorithms, items rated as don’t know were considered to be absent (no). Agreement in diagnostic classification was assessed across 7 centers given the same set of criteria, using the χ statistic for 4 sets of criteria for dementia, 2 for AD, and 4 for VaD. The χ statistic is the rate of observed agreement between a pair of ratings (or all possible pairs of ratings when multiple raters are involved) adjusted for the proportion of the agreement that can be expected to occur by chance. The following guidelines have been suggested for interpreting χ scores: 0.00 to 0.20 indicates slight agreement; 0.21 to 0.40, fair agreement; 0.41 to 0.60, moderate agreement; 0.61 to 0.80, substantial agreement; and 0.81 to 1.00, almost perfect agreement. To determine whether any center was an outlier in the comparison of interrater reliability, the analyses were repeated after removing each center.

labeled Disorders Association (NINCDS-ADRDA), ADDTC, and NINDS-AIREN criteria. The χ scores ranged from 0.87 to 0.96 (Table 2). Substantial agreement was found between NINCDS-ADRDA criteria for probable AD and DSM-IV criteria for dementia of the Alzheimer type (κ = 0.79). Lesser (ie, fair to moderate) agreement was observed among the criteria for VaD. The χ scores ranged from 0.24 (between the ADDTC probable IVD and a modified HIS) to 0.60 (between the combined ADDTC probable and possible IVD and DSM-IV dementia of the vascular type).

Among the 4 criteria for VaD, significant differences occurred in diagnostic classifications (Table 3). The DSM-IV and the modified HIS were the most liberal, each yielding a diagnosis of VaD in 25% of the ratings. The NINDS-AIREN criteria were the most conservative, leading to a diagnosis of probable or possible VaD in only 6% of the ratings (ie, 1 case). Intermediate frequencies were noted for a diagnosis of VaD by ADDTC criteria and for an original HIS (20.6%, and 13.7%, respectively).

Substantial interrater agreement was noted among the 7 centers for the diagnosis of dementia. The χ values for each of the 4 definitions of dementia ranged from 0.62 to 0.74. Moderate interrater reliability was found for the diagnosis of AD, using DSM-IV or NINCDS-ADRDA criteria (κ = 0.56 and κ = 0.46, respectively). A DSM-IV diagnosis of dementia of the Alzheimer type and an NINCDS-ADRDA diagnosis of probable AD were assigned at similar rates (20.6% of all ratings in both cases). Possible AD by NINCDS-ADRDA criteria was diagnosed in 53.7% of ratings, with moderate interrater reliability (κ = 0.47).

Variable interrater reliability was encountered for the diagnosis of VaD. Substantial agreement was realized for the original or modified HIS (κ = 0.65 and κ = 0.59, respectively). Moderate interrater agreement was achieved for an ADDTC diagnosis of probable IVD (κ = 0.44), as well as the NINDS-AIREN criteria for probable or possible VaD (κ = 0.42) and the DSM-IV diagnosis of dementia of the vascular type (κ = 0.59). Lowest interrater reliability was found for the ADDTC diagnosis of possible IVD (κ = 0.15). One-by-one removal of each participating center, followed by a subsequent interrater reliability analysis, did not identify any particular center as an outlier.

COMMENT

Unlike criteria for dementia and AD, clinical criteria for a diagnosis of VaD are not interchangeable. In our study, excellent agreement was found among 4 diagnostic criteria for dementia (κ range, 0.87-0.96), despite differ-
ences in the number and type of cognitive impairments required. Similarly, the DSM-IV criteria for dementia of the Alzheimer type and the NINCDS-ADRDA criteria for probable AD exhibited high intercriterion agreement (κ = 0.81) and moderate interrater reliability (κ = 0.56 and κ = 0.49). These findings are comparable to those of previous reports in the literature for the diagnosis of dementia20,22 and AD.28-30

On the other hand, the 4 clinical criteria for VaD studied herein produced significant differences in case classification and interrater reliability. The frequency of a diagnosis of VaD was highest using the modified HIS or DSM-IV criteria and the lowest using the NINDS-AIREN criteria. The classification rate for VaD, using the original HIS and the ADDTC criteria, fell in an intermediate range. With a few exceptions, the rank order and relative frequencies noted in our study (ie, in descending order, DSM-IV, modified HIS, ADDTC, original HIS, and NINDS-AIREN) recapitulate the findings of previous investigators. Verhey et al19 noted the following classification fre-
quences in descending order: original and modified HIS, Erkinjuntti et al\textsuperscript{31} ADDTC, Diagnostic and Statistical Manual of Mental Disorders, Third Edition (DSM-III),\textsuperscript{32} and NINDS-AIREN. Wetterling et al\textsuperscript{10} reported the following frequencies in descending order: DSM-IV, ICD-10, ADDTC, and NINDS-AIREN. Thus, 2 of these 3 studies report the DSM-IV as the most liberal. The DSM-III\textsuperscript{32} rather than the DSM-IV criteria\textsuperscript{10} were examined in the third study\textsuperscript{10} and yielded more conservative results. The DSM-III criteria proved to be the most conservative, which is not surprising given the stringency of its inclusion and exclusion criteria.

Among the 4 criteria for VaD, the HIS showed the best interrater reliability ($\kappa = 0.65$ for the original HIS; $\kappa = 0.61$ for the modified HIS). This advantage may be related to the feature’s checkoff format. For the HIS, congruence in patient classification was greatest with the DSM-IV criteria ($\kappa = 0.61$).

The DSM-IV criteria also showed the best overall agreement with other criteria for VaD ($\kappa$ range, 0.32-0.60), yielding the greatest overlap with the combined ADDTC for probable and possible IVD. The DSM-IV criteria showed moderate interrater reliability (84% agreement; $\kappa = 0.59$) in our study. This is somewhat higher than the 46% to 60% agreement for the Diagnostic and Statistical Manual of Mental Disorders, Revised Third Edition\textsuperscript{33} diagnosis of MID noted among 4 sites in the Ni-Hon-Sea study.\textsuperscript{22}

Although moderate interrater reliability was found for ADDTC criteria for probable IVD ($\kappa = 0.44$), the level of agreement for possible IVD was extremely poor ($\kappa = 0.15$). A diagnosis of possible IVD can be made in the presence of dementia and one of the following: a single stroke without a clearly documented temporal relationship to the onset of dementia, orBinswanger disease (ie, early-onset urinary incontinence or gait disturbance, vascular risk factors, and extensive white matter changes on neuroimaging findings). Neither of these conditions was rated reliably among the different centers. We have not determined whether further training of raters might improve reliability of diagnosis ofBinswanger disease using ADDTC criteria. Favorable effects of training on the reliability of the diagnosis of dementia and dementia subtypes were achieved in the Ni-Hon-Sea study.\textsuperscript{22}

Although conservative, the NINDS-AIREN criteria showed moderate interrater reliability ($\kappa = 0.42$). This is somewhat lower than the agreement reported by Lopez et al,\textsuperscript{21} where $\kappa$ scores ranged from 0.46 to 0.72. However, in the latter study, case information was coded from the clinical record onto a standardized form, rather than reported as vignettes; coding might be expected to improve interrater reliability. A post-hoc item analysis was conducted for 12 cases that were classified by a participating center as VaD using DSM-IV but not NINDS-

### Table 3. Agreement Among 7 ADDTCs\textsuperscript{*}

<table>
<thead>
<tr>
<th></th>
<th>Dementia</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Present, %</td>
<td>% Agreement</td>
<td>$\kappa$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DSM-IV</td>
<td>72.0</td>
<td>84.8</td>
<td>0.62</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NINCDS-ADRDA</td>
<td>73.7</td>
<td>86.3</td>
<td>0.65</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADDTC</td>
<td>74.9</td>
<td>89.3</td>
<td>0.72</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NINDS-AIREN</td>
<td>72.0</td>
<td>89.3</td>
<td>0.74</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Alzheimer Disease</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Present, %</td>
<td>% Agreement</td>
<td>$\kappa$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DSM-IV</td>
<td>20.6</td>
<td>85.5</td>
<td>0.56</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NINCDS-ADRDA</td>
<td>20.6</td>
<td>83.2</td>
<td>0.49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NINCDS-ADRDA possible</td>
<td>53.7</td>
<td>73.3</td>
<td>0.47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NINCDS-ADRDA probable and possible</td>
<td>68.6</td>
<td>76.8</td>
<td>0.46</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Vascular Dementia</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Present, %</td>
<td>% Agreement</td>
<td>$\kappa$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DSM-IV</td>
<td>25.7</td>
<td>84.4</td>
<td>0.59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIS, original</td>
<td>13.7</td>
<td>91.6</td>
<td>0.65</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIS, modified</td>
<td>25.1</td>
<td>85.1</td>
<td>0.61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADDTC probable</td>
<td>10.3</td>
<td>89.7</td>
<td>0.44</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADDTC possible</td>
<td>14.3</td>
<td>79.0</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADDTC probable and possible</td>
<td>20.6</td>
<td>77.1</td>
<td>0.30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NINDS-AIREN</td>
<td>5.1</td>
<td>94.3</td>
<td>0.42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NINDS-AIREN probable and possible</td>
<td>6.3</td>
<td>93.1</td>
<td>0.42</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\*Abbreviations are given in the footnotes to Tables 1 and 2.

### Table 4. Discrepant Within-Center Classifications of VaD\textsuperscript{†}

<table>
<thead>
<tr>
<th>Case Number</th>
<th>Total No. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>No. of centers with discrepant ratings†</td>
<td>1</td>
</tr>
<tr>
<td>No. of ratings that failed specific NINDS-AIREN criterion</td>
<td>Dementia</td>
</tr>
<tr>
<td></td>
<td>Rule out disturbance of consciousness, severe aphasia</td>
</tr>
<tr>
<td></td>
<td>Rule out AD</td>
</tr>
<tr>
<td></td>
<td>Focal signs</td>
</tr>
<tr>
<td></td>
<td>CVD on neuroimaging</td>
</tr>
<tr>
<td></td>
<td>Relation between dementia and CVD</td>
</tr>
</tbody>
</table>

\†Indicates diagnosis of VaD by DSM-IV but not NINDS-AIREN criteria.

©2000 American Medical Association. All rights reserved.
AIREN criteria (Table 4). The 3 most common items leading to discrepant diagnoses were exclusion of other systemic or brain diseases (including AD) that might cause the dementia (failed in 29/34 [85%] of ratings leading to discrepant diagnoses); a causal relationship between cerebrovascular brain injury and dementia (failed in 25/34 [74%]); and the presence of focal neurologic signs (failed in 15/34 [44%]). Among the criteria studied herein, the exclusion of AD and the requirement for focal neurologic signs were unique to the NINDS-AIREN criteria.

Given their lack of comparability, the choice of diagnostic criteria for VaD remains a critical but elusive methodological issue for clinical and epidemiological studies. Estimates of the incidence and prevalence of VaD will vary severalfold, depending on the criteria selected. Full evaluation of the relative merits of various diagnostic methods must include data from clinical-pathological correlation. The limitations of the HIS for the diagnosis of pathologically confirmed MII have been recognized previously. More recently, Gold et al reported the pathologic findings for cases retrospectively diagnosed using ADT and NINDS-AIREN criteria for VaD. Unfortunately, neuroimaging studies were not available in 80% of cases, thereby limiting their report to the clinical diagnosis of possible VaD. Also, neither subcortical ischemic vascular dementia norBinswanger disease was included in their pathological definition of VaD. Prospective longitudinal studies including clinical-pathological correlation are greatly needed. Meanwhile, the diagnostic approach that will enable accurate prediction of significant vascular brain injury, as well as the presence or absence of AD lesions, remains an enigmatic challenge.

Accepted for publication February 17, 1999.

From the Geriatric Neurobehavior and Alzheimer Center, Rancho Los Amigos National Rehabilitation Center, Downey (Dr Chui); the Departments of Neurology (Dr Chui) and Preventive Medicine (Dr Mack), University of Southern California, Los Angeles; Department of Medicine, University of California, San Diego (Dr Jackson); Department of Neurology, University of California, Davis (Drs Mungas, Reed, and Jagust); Martinez Veterans Affairs Medical Center, Martinez (Drs Reed and Jagust); Veterans Affairs Palo Alto Health Care Systems, and Department of Psychiatry, Stanford University, Palo Alto, Calif (Dr Tinklenberg); Departments of Neurology (Dr Chang) and Anatomy (Dr Skinner), University of California, San Francisco; and Department of Neurology, University of California, Irvine (Dr Tasaki).

This work was supported by grants 1PO1 AG12435, 1P50 AG05142, and 1P30 AG10129 from the National Institute on Aging, Bethesda, Md, and the State of California Department of Health Services, Sacramento.

Reprints: Helena C. Chui, MD, Geriatric Neurobehavior and Alzheimer Center, Rancho Los Amigos National Rehabilitation Center, 800 Annex W, 7601 E Imperial Hwy, Downey, CA 90242 (e-mail: chui@hsc.usc.edu).

REFERENCES


©2000 American Medical Association. All rights reserved.