Operational Definitions for the NINDS-AIREN Criteria for Vascular Dementia
An Interobserver Study

Elisabeth C.W. van Straaten, MD; Philip Scheltens, MD, PhD; Dirk L. Knol, PhD; Mark A. van Buchem, MD, PhD; Ewout J. van Dijk, MD; Paul A.M. Hofman, MD, PhD; Giorgos Karas, MD; Olafur Kjartansson, MD; Frank-Erik de Leeuw, MD, PhD; Niels D. Prins, MD; Reinhold Schmidt, MD, PhD; Marieke C. Visser, MD, PhD; Henry C. Weinstein, MD, PhD; Frederik Barkhof, MD, PhD

Background and Purpose—Vascular dementia (VaD) is thought to be the most common cause of dementia after Alzheimer’s disease. The commonly used International Workshop of the National Institute of Neurological Disorders and Stroke (NINDS) and the Association Internationale pour la Recherche et l’Enseignement en Neurosciences (AIREN) criteria for VaD necessitate evidence of vascular disease on CT or MRI of the brain. The purposes of our study were to operationalize the radiological part of the NINDS-AIREN criteria and to assess the effect of this operationalization on interobserver agreement.

Methods—Six experienced and 4 inexperienced observers rated a set of 40 MRI studies of patients with clinically suspected VaD twice using the NINDS-AIREN set of radiological criteria. After the first reading session, operational definitions were conceived, which were subsequently used in the second reading session. Interobserver reproducibility was measured by Cohen’s κ.

Results—Overall agreement at the first reading session was poor (κ=0.29) and improved slightly after application of the additional definitions (κ=0.38). Raters in the experienced group improved their agreement from almost moderate (κ=0.39) to good (0.62). The inexperienced group started out with poor agreement (κ=0.17) and did not improve (κ=0.18). The experienced group improved in both the large- and small-vessel categories, whereas the inexperienced group improved generally in the extensive white matter hyperintensities categories.

Conclusions—Considerable interobserver variability exists for the assessment of the radiological part of the NINDS-AIREN criteria. Use of operational definitions improves agreement but only for already experienced observers. (Stroke. 2003;34:1907-1912.)

Key Words: dementia ■ magnetic resonance imaging ■ observer variation ■ vascular disorders
Recently, a vast interest in clinical trials on the efficacy of cholinesterase inhibitors and other drugs for VaD has emerged, and the NINDS-AIREN criteria with their radiological definitions are being used on a large scale in these trials. However, clear operational definitions on how to use and interpret the radiological criteria are lacking. Only a few interobserver studies of the NINDS-AIREN criteria have been published. In 2 of these studies, both clinical and radiological diagnoses were studied together. The agreement between raters was moderate to good ($\kappa=0.42$ in the first study mentioned, 0.46-$\kappa<0.72$ in the second study). It was suggested that the cause of the disappointing results could have been the difference in interpretation of the radiological criteria by the different raters.

In this study, we examined the interobserver agreement of the radiological part of the NINDS-AIREN criteria and the effect of subsequently formulated operational definitions on the level of agreement in patients with clinical signs of VaD. Second, we investigated whether experienced and inexperienced raters would benefit equally from such definitions.

### Methods

**MRI Studies**

For this study, we selected MRI studies of patients with dementia and clinical signs of cerebrovascular disease. The selection was done to get 10 cases of large-vessel disease and 30 cases of small-vessel disease, reflecting the distribution in a recently completed trial on VaD. Two authors who did not participate in the interobserver studies (E.C.W. van S., F.B.) performed the selection by applying the NINDS-AIREN criteria to the MRI scans. Based on the experience of an $\sim$50% rejection rate in the above-mentioned trial and to have a balanced distribution in the study sample, MRI studies were selected in a way that we expected half of the scans to be rated as having sufficient abnormalities to fulfill the NINDS-AIREN criteria. It should be noted that the percentage of cases fulfilling such criteria is not reflective of the general population of patients clinically suspected of having VaD.

In addition to the 40 scans, we selected 10 scans to be scored during the first assessment and to be used for consensus reading and formulation of definitions. All MRI studies consisted of axial T2, axial fluid-attenuated inversion recovery, and axial and coronal T1 series using 5-mm slices and 1×1-mm pixel size.

**Study Design**

Ten raters with different levels of experience evaluated the 40 selected MRI studies in 2 consecutive reading sessions. The decision to use the same data set twice (rather than having 2 independent data sets) was based on the expectation that this would preclude variability to be introduced by unbalanced matching in the distribution of cases over the various subcategories of the NINDS-AIREN criteria. On the other hand, we expected no bias from a learning effect when the same samples were rated twice because the second rating was done with a set of operational criteria developed from the additional training set of 10 scans; if any, this design would tend to maintain rather than to reduce interobserver variability and therefore is slightly conservative. The team of raters consisted of 10 physicians (2 radiologists, 4 neurologists, 3 research fellows, 1 neurology resident). Six had extensive experience in the evaluation of vascular lesions on MRI scans in clinical settings or in population-based studies on aging and dementia. The other 4 had experience in assessing MRI scans of the brain, but they had never assessed vascular lesions systematically on a large scale. The raters were blinded to all clinical and personal information. During the first reading session, all raters individually assessed the scans in random order with only the aid of the table of radiological findings of the NINDS-AIREN criteria for VaD as stated in the original article. All images were presented to the readers on identical personal computers using a digital viewing program, allowing window and level adjustment. The readers were able to browse through the scans as often as they wanted; no time limits were set. Scoring consisted of 2 stages. First, lesions had to be identified and classified topographically on a scoring form (Table 1), divided into a section on large-vessel disease (strategic infarcts in certain anterior, middle, or posterior cerebral artery territories) and a section on small-vessel disease (lacunes, white matter hyperintensities, bilateral thalamic lesions). Second, the topographical information had to be combined with severity criteria to decide whether the scan met the radiological criteria for VaD (final diagnosis). Subsequently, a joint consensus reading of the

### TABLE 1. Scoring Form of Brain Imaging Lesions Associated With VaD

<table>
<thead>
<tr>
<th>Large-vessel strokes in the following territories:</th>
<th>Left</th>
<th>Right</th>
<th>Both</th>
<th>None</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anterior cerebral artery</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posterior cerebral artery, including:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inferior thalamic infarctions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intermediate temporal lobe lesions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Association areas:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parietotemporal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temporo-occipital</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Angular gyrus</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Watershed carotid territories:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Superior frontal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parietal region</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Small-vessel disease</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple basal ganglia and frontal white matter lacunae</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extensive periventricular white matter lesions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bilateral thalamic lesions</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Bilateral large-vessel hemispheric strokes       |      |      |
| Leukoencephalopathy involving at least $1/4$ of the total white matter |      |      |

<table>
<thead>
<tr>
<th>Do the radiological findings fulfill the radiological criteria for VaD?</th>
<th>Yes</th>
<th>No</th>
<th>Uncertain</th>
</tr>
</thead>
</table>

The readers were able to browse through the scans as often as they wanted; no time limits were set. Scoring consisted of 2 stages. First, lesions had to be identified and classified topographically on a scoring form (Table 1), divided into a section on large-vessel disease (strategic infarcts in certain anterior, middle, or posterior cerebral artery territories) and a section on small-vessel disease (lacunes, white matter hyperintensities, bilateral thalamic lesions). Second, the topographical information had to be combined with severity criteria to decide whether the scan met the radiological criteria for VaD (final diagnosis). Subsequently, a joint consensus reading of the.
TABLE 2. Operational Definitions of the Imaging Guidelines of the NINDS-AIREN Criteria for VaD

Topography

Large-vessel stroke—Large-vessel stroke is an infarction defined as a parenchymal defect in an arterial territory involving the cortical gray matter. 

ACA—only bilateral ACA infarcts are sufficient to meet the NINDS-AIREN criteria 

PCA—Infarcts in the PCA territory can be included only when they involve the following regions: 

Parametermed thalamic infarction: the infarct involves the cortical gray matter of the temporal/occipital lobe and extends into the paramedian part (defined as extending to the third ventricle) of the thalamus; the extension may be limited to the gliotic rim of the infarct that surrounds the parenchymal defect. 

Inferior medial temporal lobe lesions 

Association areas—an MCA infarction needs to involve the following regions: 

Parietotemporal: the infarct involves both the parietal and temporal lobe (eg, angular gyrus) 

Temporal-occipital: the infarct involves both the temporal and occipital lobe 

Watershed carotid territories—a watershed infarction is defined as an infarct in the watershed area between the MCA and PCA or the MCA and ACA in the following regions: 

Superior frontal region 

Parietal region 

Small-vessel disease 

Ischemic pathology resulting from occlusion of small perforating arteries may manifest itself as lacunes or white matter lesions. Lacune is defined as a lesion with CSF-like intensity on all sequences on MR (water density on CT) surrounded by white matter or subcortical gray matter. Care should be taken not to include Virchow-Robin spaces, which typically occur at the vertex and around the anterior commissure near the substantia perforata. 

Ischemic white matter lesions are defined as circumscribed abnormalities with high signal on T2-weighted images not following CSF signal (mildly hypodense compared with surrounding tissue on CT) with a minimum diameter of 2 mm. 

Multiple basal ganglia and frontal white matter lacunes—criteria are met when at least 2 lacunes in the basal ganglia region (including thalamus and internal capsule) and at least 2 lacunes in the frontal white matter are present. Extremely periventricular white matter lesions—lesions in the white matter abutting the ventricles and extending irregularly into the deep white matter, or deep/subcortical white matter lesions. Smooth caps and bands by themselves are not sufficient. Gliotic areas surrounding large-vessel strokes should not be included here. 

Bilateral thalamic lesions—to meet the criteria, at least 1 lesion in each thalamus should be present. 

Severity 

Large-vessel disease of the dominant hemisphere—if there is a large-vessel infarct as defined above, to meet the criteria it has to be in the dominant hemisphere. In the absence of clinical information, the left hemisphere is considered dominant. 

Bilateral large-vessel hemispheric strokes—1 of the infarcts should involve an area listed under topography but is in the nondominant hemisphere, while the infarct in the dominant hemisphere does not meet the topography criteria. 

Leukoencephalopathy involving at least ¼ of the total white matter—extensive white matter lesions are considered to involve ¼ of the total white matter when they are confluent (grade 3 in the ARWMC scale) in at least 2 regions of the ARWMC scale and beginning confluent (grade 2 in the ARWMC scale) in 2 other regions. A lesion is considered confluent when >20 mm or consists of ≥2 smaller lesions fused by more than connecting bridges. 

Fulfillment of radiological criteria for probable VaD 

Large-vessel disease—both the topography and severity criteria should be met (a lesion must be scored in at least 1 subsection of both topography and severity). 

Small-vessel disease—for white matter lesions, both the topography and severity criteria should be met (a lesion must be scored in at least 1 subsection of both topography and severity); for multiple lacunes and bilateral thalamic lesions, only the topography criterion is sufficient. 

ACA indicates anterior cerebral artery; PCA, posterior cerebral artery; MCA, middle cerebral artery; CSF, cerebrospinal fluid; and ARWMC, age-related white matter changes.

additional 10 scans was held, and operational definitions for scoring vascular lesions according to the NINDS-AIREN criteria were discussed. After consensus on a set of definitions was reached, a second reading of the 40 scans was performed the next day, again in random order, according to the newly formulated operational definitions (Table 2).

Statistical Analysis

We determined agreement between raters for the 2 reading sessions separately by Cohen’s κ for >2 raters.10,11 The weighted κ was not used because most scorings were dichotomous and the different categories were not ordered. We did this using AGREE software (ProGAMMA), which also calculated standard error values. We determined κ for presence of radiological evidence for probable VaD, presence of large-vessel disease, and presence of small-vessel disease. To test whether agreement between the first and second readings differed statistically, we determined z values for the difference in κ and used the corresponding probability value for testing. All scores were calculated for 3 groups: the whole group of raters (n=10), the group of experienced raters (n=6), and the group of inexperienced raters (n=4). A κ between 0 and 0.2 refers to poor agreement; 0.2 to 0.4, fair agreement; 0.41 to 0.60, moderate agreement; 0.61 to 0.80, good agreement; and 0.81 to 1.00, very good agreement.12

Results

Table 3 shows the results of the baseline readings. In 35.8% of all cases, a large-vessel infarction was scored; in 60.3% of cases, small-vessel disease was scored. This distribution was roughly as expected. The percentage of cases in which the raters found vascular lesions that met the radiological criteria of the NINDS-AIREN was 41.3%, which again is in line with...
what we had anticipated. In Table 4, $\kappa$ is given for the various sections of the scoring separately. At the first reading session, agreement in the group of inexperienced raters was generally less than the agreement in the group of experienced raters. This is also true for the assessment of the final diagnosis (Table 5). At the first reading session, mean $\kappa$ for the final diagnosis for all raters signifies fair agreement.

After the first scoring, operational definitions were formulated in consensus (Table 2). During this consensus meeting, we identified the problems that had risen with the interpretation of the criteria. The meaning, exact location, and borders of a paramedian thalamic infarction were uncertain in our opinion. We had trouble interpreting the term “multiple basal ganglia and frontal white matter lacunes.” Questions that arose included, Are lacunes needed in both areas to meet the criteria? How many lacunes is “multiple” exactly? How big should an extensive periventricular white matter lesion be, and is a lesion considered only when directly abutting the ventricles? Should strokes in any area be considered in the bilateral large-vessel hemispheric strokes category, or only those strokes that are scored previously in the topography characteristics in the severity section? How can we approximate one fourth of the total white matter? We tried to address these questions in the operational criteria, leaving the original set of criteria fundamentally intact. Definitions were laid out for the different radiological types of vascular pathology, different regions of relevant strokes were defined, and for small-vessel disease, numeric definitions were adopted. With respect to the leukoencephalopathy, we agreed on quantification with the use of the age-related white matter changes (ARWMC) rating scale.13 In the severity section, we discussed dominance of hemispheres, and for practical reasons, the left hemisphere was considered dominant. In addition to describing the different parts of the diagnostic criteria, rules on how to combine these parts were added because we noticed differences in opinion during consensus reading. We agreed that a scan would meet the final diagnosis of VaD if both severity and topography criteria were met, with the exception of the bilateral thalamic lesions and multiple lacunes subcategories, which have no related severity criterion.

Table 4 shows that agreement generally increases, especially in the small-vessel category. To calculate significance of change in $\kappa$, $z$ values were calculated. For appreciation of large-vessel disease, $z$ values indicated that none of these differences are statistically significant. For small-vessel disease, only the difference in scoring of the inexperienced raters in the small-vessel category showed statistically significant improvement ($P=0.04$).

The mean $\kappa$ for the final diagnosis for all raters at the second reading was slightly greater than at the first reading session (Table 5). For the experienced group, agreement rose to $\kappa=0.62$, but in the inexperienced group, it remained low. Only in the experienced group of raters did agreement improve significantly.

**Discussion**

We examined the interobserver agreement for the radiological assessment of the NINDS-AIREN criteria and quantified the added value of operational definitions. We found that overall agreement on the final diagnosis of VaD was only fair without guidelines, especially for inexperienced raters; this was true for the agreement in both the large- and small-vessel pathology categories. The large variability we found is in agreement with earlier studies on the total set of criteria for VaD of the NINDS-AIREN and is in part the result of a lack of operational definitions for the radiological criteria. Already at the first scoring, several problems with the interpretation of those criteria arose. The weakest parts of the radiological criteria are those related to small-vessel disease, for which the original publication provides no details. This is especially unfortunate because these are the most prevalent types of pathology in patients with VaD. The raters also experienced difficulties in combining the individual parts of the criteria to make the final decision of VaD or not. Confusingly, the original criteria by the NINDS-AIREN list some of the topography characteristics in the severity section and vice versa. Additional definitions will not solve this problem because they do not actually change the original criteria. Taking this into consideration, we can explain low agreement. Our results suggest that a revision of the original criteria might be needed in this respect.

After the application of operational definitions, agreement on the final diagnosis of VaD improved. However, stratified analysis showed that this improvement in agreement was confined to the group of experienced raters with a $\kappa$ of 0.62, indicating good agreement. This was due to improvements in both the large- and small-vessel categories. In the group of inexperienced raters, agreement worsened in the large-vessel
category but improved in the small-vessel category. The latter was due mainly to an increase in $\kappa$ by 0.35 to good agreement in the extensive white matter lesions subcategory.

The design of this study has some limitations. We did not have a gold standard. The operational definitions were not validated against pathology or clinical findings but had the sole purpose of being practical, usable, and able to improve standardization. In addition, the raters did not have clinical information that could have contributed to the final diagnosis. In large clinical trials in which the MRI scans are rated centrally, this information is also not available, but agreement can be expected to improve in a clinical setting because previous studies show higher $\kappa$ even when agreement is high. Finally, the operational definitions formulated are, of course, arbitrary and may be subject to further amendments. However, the raters who formulated the criteria were the same raters who were going to apply them in the second reading session. It can therefore be expected that they were optimal for use in this interobserver study.

In conclusion, we found that the radiological criteria for the NINDS-AIREN criteria for VaD are very complex. This makes these criteria less suitable for inexperienced raters and not appropriate for routine diagnosis on the basis of a standard radiological report only. The radiological criteria for the NINDS-AIREN criteria for VaD have suboptimal reproducibility. Use of operational criteria improves agreement to acceptable levels, but only in experienced readers. Because operational definitions essentially do not change the original criteria, a critical reappraisal of the NINDS-AIREN radiological criteria seems to be needed to further improve the quality of the criteria and interobserver agreement. We hope that our results set the stage for such an endeavor.
References

Operational Definitions for the NINDS-AIREN Criteria for Vascular Dementia: An Interobserver Study

Elisabeth C.W. van Straaten, Philip Scheltens, Dirk L. Knol, Mark A. van Buchem, Ewout J. van Dijk, Paul A.M. Hofman, Giorgos Karas, Olafur Kjartansson, Frank-Erik de Leeuw, Niels D. Prins, Reinhold Schmidt, Marieke C. Visser, Henry C. Weinstein and Frederik Barkhof

*Stroke.* 2003;34:1907-1912; originally published online July 10, 2003;
doi: 10.1161/01.STR.0000083050.44441.10

*Stroke* is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2003 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/34/8/1907

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in *Stroke* can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to *Stroke* is online at:
http://stroke.ahajournals.org//subscriptions/