Interobserver Agreement in the Classification of Stroke in the Women’s Health Study

Monika Atiya, MD; Tobias Kurth, MD, MSc; Klaus Berger, MD, MPH, MSc; Julie E. Buring, ScD; Carlos S. Kase, MD

Background and Purpose—Accurate classification of stroke events is essential in large cohort studies of risk factor assessment and in treatment trials such as the Women’s Health Study (WHS).

Methods—Based on medical record review, we assessed interrater reliability in stroke classification and disability status in the WHS using the $\kappa$ statistic.

Results—During 7.0 years of follow-up, 271 incident strokes occurred, of which 133 were reclassified. There was excellent interrater agreement in the diagnosis of major stroke types, hemorrhagic subtypes, and degree of disability, as well as substantial agreement in the definition of the vascular territory involved. Only moderate agreement was reached in the classification of ischemic subtypes.

Conclusions—Major stroke types, degree of disability, and vascular territory involved can be reliably classified on the basis of a review of medical records in the WHS, whereas classification of ischemic stroke subtypes needs further refinement of diagnostic criteria. (Stroke. 2003;34:565-567.)

Key Words: classification ■ diagnosis ■ interobserver variation ■ stroke

A valid and reproducible diagnosis of stroke is essential in the assessment of risk factors in observational studies. In addition, the differentiation of stroke subtypes and the determination of degree of disability are important factors in the evaluation of therapeutic agents to properly assess their effects.

Large epidemiological studies that have sufficient power to detect meaningful associations cannot include direct patient examinations and often have to rely solely on medical record information. We evaluated the interobserver reliability for stroke subtypes and degree of disability on the basis of medical record review in the Women’s Health Study (WHS).

Methods

Subjects were participants in the WHS, an ongoing clinical trial of low-dose aspirin and vitamin E in the primary prevention of cardiovascular disease and cancer. Study participants were 39,876 female health professionals ≥45 years at baseline. Information was collected by mailed questionnaire asking about newly diagnosed conditions, including stroke.

Stroke Classification

We asked participants who self-reported stroke for permission to obtain their medical records. Stroke was defined as a focal neurological deficit of vascular mechanism lasting >24 hours and confirmed by an end points committee. After extracting the information from the medical records, which included medical history, clinical information, and results of diagnostic tests, we classified strokes using a standardized stroke card. For stroke subtypes, the modified Trial of Org 10172 in Acute Stroke Treatment (TOAST) criteria were used. In addition, vascular territory was determined from clinical information or imaging results. Degree of disability was measured with the modified Rankin Scale from information from the hospital discharge summary. We randomly selected 134 of 271 stroke cases, of which 1 medical record was not available at the time of this study, leaving 133 strokes for this study. A second neurologist (M.A.), blinded to the first classification (by C.S.K.), independently reclassified those cases using the same medical records and standardized diagnostic card.

Statistical Analysis

We used Cohen’s $\kappa$ statistic to measure the agreement between the 2 raters. To measure the agreement of ordinal scales, we estimated quadratic-weighted $\kappa$. This measure accounts for differences in the classification by giving partial agreement weights for near but not exact agreement. Generally, $\kappa > 0.80$ represents excellent agreement; $0.60 < \kappa < 0.80$ is thought to be substantial agreement; $0.40 < \kappa < 0.60$, moderate agreement; $0.20 < \kappa < 0.40$, fair agreement; and $\kappa < 0.20$, slight or poor agreement.

Results

After a mean follow-up of 7.0 years, 271 incident strokes occurred, of which 133 were reclassified. Information found in the medical records was heterogeneous, ranging from...
self-reports only to extensive diagnostic work-up and clinical information. Of 133 cases, 111 (83%) had CT studies, 77 (58%) had MRI studies, 54 (41%) had contrast angiography, and 57 (43%) had echocardiograms.

The interrater agreement on major stroke subtypes (ischemic, hemorrhagic, unknown type) was excellent, with $\kappa=0.96$. Only 1 of 27 hemorrhagic strokes was reclassified as stroke of unknown type, whereas 1 of 2 strokes of unknown type was reclassified as hemorrhagic.

Table 1 shows the interobserver agreement in subtypes of ischemic stroke with only moderate overall agreement ($\kappa=0.49$). The $\kappa$ value varied widely with substantial agreement in cardioembolic subtype ($\kappa=0.74$) and moderate agreement in atherothrombotic and atheroembolic ($\kappa=0.59$) and unknown mechanism ($\kappa=0.48$) types. In contrast, agreement in hemorrhagic subtypes of stroke was excellent, with $\kappa=0.94$.

Table 2 shows the agreement in the classification of the involved vascular territory. The initial category of stem and branches of the main arteries was collapsed into 1 category, which slightly increased the $\kappa$ value from 0.68 to 0.73. The highest disagreement was found in the basilar artery category, for which 3 of 7 strokes originally diagnosed in the basilar artery territory were attributed to the posterior cerebral artery territory.

Table 3 represents the interrater agreement on the degree of disability. Agreement was found in 58% of cases with a weighted $\kappa=0.89$. Collapsing the modified Rankin Scale into categories of independent (0 to 1), moderately dependent (2 to 3), and severely dependent (4 to 5) yielded a weighted $\kappa$ of 0.86.

**Discussion**

We have shown that stroke reclassification based on medical record information in the WHS can be done reliably with an excellent interrater agreement in major stroke types, hemorrhagic stroke subtypes, and degree of disability. Additionally, we found substantial agreement in the vascular territory involved. In contrast, the agreement for the various ischemic stroke subtypes was only moderate.

Diagnosis of hemorrhagic stroke is greatly facilitated by the generally unequivocal CT features of its subtypes.² The difficulty in classifying ischemic stroke might be explained by several factors. The classification of lacunar strokes depends heavily on interpretation of clinical features because CT neuroimaging may be negative. The lower reliability of clinical data alone is well known. The 2 physicians might have weighted discrepant information differently, a problem also encountered by others.¹⁰ Some variability will always remain because of inconsistencies in medical records and errors in data abstractions, even after reliability is improved through the use of a computerized algorithm to guide data abstraction.¹⁰ Considering the results of 2 prospective studies that showed different interrater agreement for neurological signs,¹¹,¹² our findings are not surprising.

Categorization of ischemic stroke is difficult because it still has to rely on poorly defined data such as stroke pathogenesis and vascular territory involved.²,³ Our findings are in line with those of the Physician’s Health Study⁷ and others that showed variable interrater agreement and emphasize the need to refine the definitions. Because misclassification is likely, studies investigating ischemic stroke subtypes should be interpreted with caution.

Classification of stroke according to vascular territory involved showed substantial agreement. This is likely due to the high frequency of neuroimaging-based diagnoses in this group of subjects. The diagnostic value of neuroimaging data has been shown when both a vascular territory classification and the modified TOAST criteria are used.¹⁴ Our findings of interrater agreement in degree of disability confirm previous results.⁷ Patients with severe deficits and those with minor impairment can be classified more accurately than those in the intermediate range of disability.
Similar results with the use of a 6-item modified Rankin Scale have been reported by other investigators.\textsuperscript{3,9}

In summary, our study demonstrates excellent interrater agreement on major stroke types and degree of disability, and substantial agreement on vascular territory involved using a classification system based on review of medical records. Classification of ischemic stroke subtypes is more difficult, leading to moderate levels of interobserver reliability. Further refinement of diagnostic criteria and widespread use of modern neuroimaging data are required for a reliable classification of ischemic stroke subtypes in observational studies and controlled clinical trials.

### Acknowledgments

The study is supported by the National Institutes of Health (CA-47988 from the National Cancer Institute and HL-43851 from the National Heart, Lung, and Blood Institute). We are indebted to the participants in the WHS for their outstanding commitment and cooperation and to the entire WHS staff for their expert and unfailing assistance.

### References

5. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull. 1968;70:213–220.

### Table 3: Interrater Agreement on Degree of Disability Using the Modified Rankin Scale

<table>
<thead>
<tr>
<th>Second Classification</th>
<th>None</th>
<th>Significant</th>
<th>Slight</th>
<th>Moderate</th>
<th>Severe to Severe</th>
<th>Severe</th>
<th>Death</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>First classification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>4</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>No significant</td>
<td>7</td>
<td>33</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>49</td>
</tr>
<tr>
<td>Slight</td>
<td>0</td>
<td>9</td>
<td>12</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td>Moderate</td>
<td>0</td>
<td>5</td>
<td>9</td>
<td>6</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>23</td>
</tr>
<tr>
<td>Moderate to severe</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>Severe</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Death</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td>53</td>
<td>29</td>
<td>12</td>
<td>12</td>
<td>3</td>
<td>13</td>
<td>133</td>
</tr>
</tbody>
</table>

Weighted $\kappa = 0.89$. 

Downloaded from http://stroke.ahajournals.org/ by guest on May 21, 2017
Interobserver Agreement in the Classification of Stroke in the Women's Health Study
Monika Atiya, Tobias Kurth, Klaus Berger, Julie E. Buring and Carlos S. Kase

Stroke. 2003;34:565-567; originally published online January 16, 2003;
doi: 10.1161/01.STR.0000054159.21017.7C
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/34/2/565