Improving the Reliability of Stroke Subgroup Classification Using the Trial of ORG 10172 in Acute Stroke Treatment (TOAST) Criteria

Larry B. Goldstein, MD; Michael R. Jones, MD; David B. Matchar, MD; Lloyd J. Edwards, PhD; Jennifer Hoff, MS; Vani Chilukuri, MD; S. Beth Armstrong, BA; Ronnie D. Horner, PhD

Background and Purpose—We sought to improve the reliability of the Trial of ORG 10172 in Acute Stroke Treatment (TOAST) classification of stroke subtype for retrospective use in clinical, health services, and quality of care outcome studies. The TOAST investigators devised a series of 11 definitions to classify patients with ischemic stroke into 5 major etiologic/pathophysiological groupings. Interrater agreement was reported to be substantial in a series of patients who were independently assessed by pairs of physicians. However, the investigators cautioned that disagreements in subtype assignment remain despite the use of these explicit criteria and that trials should include measures to ensure the most uniform diagnosis possible.

Methods—In preparation for a study of outcomes and management practices for patients with ischemic stroke within Department of Veterans Affairs hospitals, 2 neurologists and 2 internists first retrospectively classified a series of 14 randomly selected stroke patients on the basis of the TOAST definitions to provide a baseline assessment of interrater agreement. A 2-phase process was then used to improve the reliability of subtype assignment. In the first phase, a computerized algorithm was developed to assign the TOAST diagnostic category. The reliability of the computerized algorithm was tested with a series of synthetic cases designed to provide data fitting each of the 11 definitions. In the second phase, critical disagreements in the data abstraction process were identified and remaining variability was reduced by the development of standardized procedures for retrieving relevant information from the medical record.

Results—The 4 physicians agreed in subtype diagnosis for only 2 of the 14 baseline cases (14%) using all 11 TOAST definitions and for 4 of the 14 cases (29%) when the classifications were collapsed into the 5 major etiologic/pathophysiological groupings (κ=0.42; 95% CI, 0.32 to 0.53). There was 100% agreement between classifications generated by the computerized algorithm and the intended diagnostic groups for the 11 synthetic cases. The algorithm was then applied to the original 14 cases, and the diagnostic categorization was compared with each of the 4 physicians’ baseline assignments. For the 5 collapsed subtypes, the algorithm-based and physician-assigned diagnoses disagreed for 29% to 50% of the cases, reflecting variation in the abstracted data and/or its interpretation. The use of an operations manual designed to guide data abstraction improved the reliability subtype assignment (κ=0.54; 95% CI, 0.26 to 0.82). Critical disagreements in the abstracted data were identified, and the manual was revised accordingly. Reliability with the use of the 5 collapsed groupings then improved for both interrater (κ=0.68; 95% CI, 0.44 to 0.91) and intrarater (κ=0.74; 95% CI, 0.61 to 0.87) agreement. Examining each remaining disagreement revealed that half were due to ambiguities in the medical record and half were related to otherwise unexplained errors in data abstraction.

Conclusions—Ischemic stroke subtype based on published TOAST classification criteria can be reliably assigned with the use of a computerized algorithm with data obtained through standardized medical record abstraction procedures. Some variability in stroke subtype classification will remain because of inconsistencies in the medical record and errors in data abstraction. This residual variability can be addressed by having 2 raters classify each case and then identifying and resolving the reason(s) for the disagreement. (Stroke. 2001;32:1091-1097.)

Key Words: diagnosis ■ stroke classification

See Editorial Comment, page 1096

Received August 21, 2000; final revision received November 1, 2000; accepted February 6, 2001.


Correspondence to Larry B. Goldstein, MD, Duke Center for Cerebrovascular Disease, Stroke Policy Program, Center for Clinical Health Policy Research, Box 3651, Duke University Medical Center, Durham, NC 27710. E-mail golds004@mc.duke.edu

© 2001 American Heart Association, Inc.

Stroke is available at http://www.strokeaha.org

1091
tions to classify patients with ischemic stroke into 5 major etiologic/pathophysiological groupings (Table 1). Interrater reliability was moderate in a series of 18 patients who were independently assessed by 24 physician-investigators (overall \( \kappa = 0.54 \)). Since the description of the TOAST scheme, it has been used to classify patients according to ischemic stroke subtype in several studies. However, the TOAST investigators cautioned that disagreements in subtype assignment remain despite the use of these explicit criteria and that trials should include measures to ensure the most uniform diagnosis possible. In the final report of the TOAST study, all of the stroke subtype diagnoses were assigned by a central-blinded evaluator to minimize interrater variability.

In preparation for a study of outcomes and management practices for patients with ischemic stroke within Department of Veterans Affairs (VA) hospitals, we used the published TOAST criteria to retrospectively categorize a series of cases. Because only fair to moderate levels of interrater agreement were achieved, we engaged in a systematic effort to improve the reliability of assigning subtype diagnoses using the TOAST definitions.

**Methods**

All patient records used in this study were randomly selected from those of patients enrolled in the VA Acute Stroke Study. Patients had been admitted to any of 9 geographically dispersed VA medical centers and identified by onsite research assistants. The diagnosis of ischemic stroke was confirmed by medical record review and, when required, by consultation with the attending physician.

Two experienced neurologists and 2 internists first independently reviewed the medical records of each of 14 randomly selected patients and assigned stroke subtype on the basis of the TOAST definitions (Table 1). Each rater was provided with reference materials listing the published criteria used by the TOAST investigators in assigning patients to a given diagnostic category. Only fair to moderate levels of interrater agreement (see Results) led to a 2-phased approach aimed at improving reliability.

In the first phase, a standardized form was designed to record the abstracted data necessary to assign a TOAST subtype classification (Figure). A computerized SAS algorithm (SAS Institute Inc) was then developed and refined to classify cases according to the 11 described TOAST definitions. The reliability of the computerized algorithm was tested with synthetic cases that provided data fitting each of the 11 definitions (Table 2).

The computerized algorithm was then applied to the original 14 cases with the use of data from 1 set of abstractions. The resulting diagnostic categorizations were compared with each physician’s baseline assignment. With variability due to differences in the subjective interpretation of the data removed by use of the algorithm, remaining variability had to be due to differences in the physicians’ application of the TOAST criteria or differences in the abstracted data (ie, critical differences in data entered into the computerized algorithm could lead to differences in subtype diagnosis).

The data abstraction process was refined in the second phase. First, an operations manual designed to guide the abstraction process was developed. Using the operations manual, 3 experienced abstractors (a nurse with extensive stroke-related experience, a stroke neurologist, and a medical student with extensive training) independently recorded data on the standardized forms for a series of 17 patients. Systematically comparing the data abstracted by each rater identified areas of disagreement critical to the assignment of stroke subtype, and the abstraction manual was then revised though a series of iterations. Using the final revised operations manual (see the Appendix, which may be found online at http://stroke.ahajournals.org), 2 raters then independently abstracted a final set of 20 cases with the extracted data entered into the computerized algorithm. Intrarater reliability was assessed by having 1 observer abstract a set of 61 cases on 2 separate occasions 6 months apart.

The degrees of intrarater and interrater reliability were measured by simple percentage of agreement and with the unweighted \( \kappa \) statistic. The values of the \( \kappa \) statistic may be interpreted in a manner similar to the interpretation of correlation coefficients (\( \kappa = 0 \) to 0.20, slight; \( \kappa = 0.21 \) to 0.40, fair; \( \kappa = 0.41 \) to 0.60, moderate; \( \kappa = 0.61 \) to 0.80, substantial; and \( \kappa = 0.81 \) to 1.00, almost perfect agreement). Probabilities reflect the chances that the calculated \( \kappa \) values were statistically different from zero.

**Results**

**Baseline Assessment**

Table 3 presents the baseline stroke subtype classifications for 14 patients with acute ischemic stroke derived by 2 experienced neurologists and 2 internists using the TOAST criteria. The internists were less likely to classify patients in the undetermined category than the neurologists. One of the neurologists was unsure whether one of the patients actually had a stroke. The 4 physicians agreed in subtype diagnosis for only 2 of the 14 cases (14%) using all 11 categories (\( \kappa = 0.29; 95\% \) CI, 0.21 to 0.37, \( P < 0.0001 \)). The 4 raters were concordant in diagnostic assignment for 4 of the 14 cases (29%) when the classifications were collapsed into the 5 major etiologic/pathophysiological groupings (overall \( \kappa \) statistic for the 4 evaluators was 0.42; 95% CI, 0.32 to 0.53; \( P < 0.0001 \)). The 2 neurologists’ classifications were concordant for 6 (43%) of the 14 cases using the full 11 TOAST categories and for 8 cases (57%) using the collapsed 5 categories. One of the internists arrived at the same diagnoses for 6 of the 8 patients (75%) for whom the 2 neurologists agreed. The second internist concurred with only 4 of these 8 classifications (50%).

**Development and Reliability of Computerized Diagnostic Algorithm**

Because of the relatively poor reliability found in this initial assessment, a standardized abstraction form was devised (Figure), and a computerized algorithm was developed to categorize patients according to the published TOAST criteria. This was accomplished through a series of iterations in which both the abstraction form and computer

---

**Table 1. TOAST Diagnostic Classification**

<table>
<thead>
<tr>
<th>Diagnostic Group</th>
<th>Collapsed Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Atherosclerosis, probable</td>
<td>Atherosclerosis</td>
</tr>
<tr>
<td>2 Atherosclerosis, possible</td>
<td>Cardiogenic</td>
</tr>
<tr>
<td>3 Cardioembolic, probable</td>
<td>Cardioembolic</td>
</tr>
<tr>
<td>4 Cardioembolic, possible</td>
<td>Lacunar</td>
</tr>
<tr>
<td>5 Lacunar, probable</td>
<td>Lacunar</td>
</tr>
<tr>
<td>6 Lacunar, possible</td>
<td>Undetermined etiology</td>
</tr>
<tr>
<td>7 Other determined etiology, probable</td>
<td>Other determined etiology</td>
</tr>
<tr>
<td>8 Other determined etiology, possible</td>
<td>Undetermined etiology</td>
</tr>
<tr>
<td>9 Undetermined etiology, complete evaluation</td>
<td>Undetermined etiology</td>
</tr>
<tr>
<td>10 Undetermined etiology, incomplete evaluation</td>
<td>Undetermined etiology</td>
</tr>
<tr>
<td>11 Multiple possible etiologies</td>
<td>Undetermined etiology</td>
</tr>
</tbody>
</table>

Details of the definitions of each category are provided in the original publications.
programming were tested and refined (data not shown). A group of 11 synthetic cases was then created to fit each of the TOAST categories (Table 2). There was 100% agreement between classifications determined by the computerized algorithm and the intended diagnostic groups for the synthetic cases.

The computerized algorithm was then applied to the 14 cases used in the baseline assessment with data from one of the neurologist’s abstractions. The algorithm-based diagnosis agreed with the 2 neurologists for 8 (57%) and 10 (71%) of the 14 cases and with the internists for 7 (50%) and 8 (57%) of the cases, respectively. Because the computerized algorithm yields consistent diagnostic categorizations in accord with the TOAST definitions, these discrepancies could only have been related to differences in the data as abstracted by the different raters, or differences in their interpretations of these data.

### Reliability of Data Abstraction

A manual to guide data abstraction was then developed by comparing disagreements among 3 experienced reviewers (data not shown). To test the revised abstraction methodology and to systematically explore remaining sources of variability, an additional set of 17 cases was independently reviewed by 2 raters with the extracted data entered into the comput-
TABLE 3. Baseline Interrater Reliability

<table>
<thead>
<tr>
<th>Case</th>
<th>Neurologist 1</th>
<th>Neurologist 2</th>
<th>Internist 1</th>
<th>Internist 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>O+</td>
<td>O+</td>
<td>CE−</td>
<td>O+</td>
</tr>
<tr>
<td>2</td>
<td>U−</td>
<td>AT−</td>
<td>L−</td>
<td>AT−</td>
</tr>
<tr>
<td>3</td>
<td>U−</td>
<td>U−</td>
<td>AT+</td>
<td>AT−</td>
</tr>
<tr>
<td>4</td>
<td>AT+</td>
<td>AT+</td>
<td>AT+</td>
<td>AT−</td>
</tr>
<tr>
<td>5</td>
<td>AT+</td>
<td>U−</td>
<td>AT+</td>
<td>AT+</td>
</tr>
<tr>
<td>6</td>
<td>U−</td>
<td>CE−</td>
<td>CE+</td>
<td>CE−</td>
</tr>
<tr>
<td>7</td>
<td>L−</td>
<td>L+</td>
<td>L+</td>
<td>L+</td>
</tr>
<tr>
<td>8</td>
<td>U−</td>
<td>L−</td>
<td>L+</td>
<td>L−</td>
</tr>
<tr>
<td>9</td>
<td>M</td>
<td>M</td>
<td>L+</td>
<td>L−</td>
</tr>
<tr>
<td>10</td>
<td>L+</td>
<td>L+</td>
<td>L+</td>
<td>L+</td>
</tr>
<tr>
<td>11</td>
<td>? Stroke</td>
<td>U+</td>
<td>U+</td>
<td>L+</td>
</tr>
<tr>
<td>12</td>
<td>AT+</td>
<td>AT+</td>
<td>AT+</td>
<td>AT+</td>
</tr>
<tr>
<td>13</td>
<td>O+</td>
<td>O−</td>
<td>U+</td>
<td>O+</td>
</tr>
<tr>
<td>14</td>
<td>U−</td>
<td>AT+</td>
<td>AT+</td>
<td>M</td>
</tr>
</tbody>
</table>

O indicates other determined etiology; CE, cardioembolic; U, undetermined etiology; AT, atherosclerosis; L, lacunar; M, multiple possible etiologies; +, probable; and −, possible.

erized algorithm. Using the 5 collapsed categories, the 2 raters agreed in subtype diagnosis for 11 of the 17 cases (65%; \( \kappa=0.54; 95\% \) CI, 0.26 to 0.82; \( P<0.05 \)).

Examination of the raters’ data abstraction forms revealed that disagreements in subtype diagnoses were primarily due to differences in the interpretations of CT and MRI scans, cardiovascular evaluations, and carotid ultrasound results. Many of these disagreements occurred because raters relied on different reports in the medical record. For example, one rater used official radiology reports, whereas the other used interpretations of the studies as reflected in physicians’ notes. As a result, the operations manual was revised to specify a hierarchy of test reports to be used for abstraction of the results of the radiological tests (Appendix).

Final Interrater and Intrarater Reliability

The abstraction process was repeated for an additional set of 20 cases with the use of the final operations manual (Appendix) with patients categorized into the 5 collapsed major etiologic/pathophysiological groupings (Table 4). Reliability further improved, with the 2 raters agreeing in subtype diagnosis for 17 of the 20 cases (85%; \( \kappa=0.68; 95\% \) CI, 0.44 to 0.91; \( P<0.05 \)). All of the differences in diagnostic assignment were due to differences in abstracted data. Examining each disagreement revealed that half were due to ambiguities in the medical record (eg, in one case medical notes indicated results of an MRI without other evidence in the record that the test was actually performed; in another case a carotid duplex evaluation indicated mild to moderate stenosis), and half were due to otherwise unexplained errors in data abstraction.

Intrarater reliability was assessed by having one observer abstract a set of 61 cases on 2 separate occasions 6 months apart, with the data entered into the computerized algorithm for diagnostic categorization. Diagnoses agreed for 50 cases (82%; \( \kappa=0.74; 95\% \) CI, 0.61 to 0.87). Again, discrepancies were largely due to differences in classification of CT and MRI scans, cardiovascular evaluations, and carotid ultrasound results.

Discussion

Presumed stroke subtype diagnosis guides both clinical evaluations and treatment decisions and may be important for understanding differences in the impact of a given intervention in the setting of clinical trials. The Oxfordshire criteria categorize subtypes of ischemic stroke primarily on the basis of vascular territory. Although this classification is both reliable and valid and provides information relevant to prognosis, it does not classify stroke patients with regard to pathophysiology or etiology. The TOAST classification scheme for ischemic stroke subtype is being used in both prospective clinical trials and retrospective studies of patterns of care and stroke-related outcomes for this purpose. Even when applied prospectively in a clinical setting, the TOAST investigators found that initial stroke subtype diagnosis should be made cautiously. Initial clinical impression of stroke subtype with the use of the TOAST criteria agreed with final diagnosis in only 62% of patients.

We found that the reliability of the TOAST classification was only fair to moderate when the published definitions were retrospectively applied to a randomly selected series of cases. Unified central assessment of stroke subtype was used in the TOAST trial itself to minimize intersubject variability. The presence of this variability confirms that the published TOAST criteria should be used with caution unless the investigators can demonstrate acceptable levels of agreement within the context of an individual study.

TABLE 4. Final Interrater Reliability

<table>
<thead>
<tr>
<th>Case</th>
<th>Rater 1</th>
<th>Rater 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AT</td>
<td>AT</td>
</tr>
<tr>
<td>2</td>
<td>AT</td>
<td>AT</td>
</tr>
<tr>
<td>3</td>
<td>CE</td>
<td>CE</td>
</tr>
<tr>
<td>4</td>
<td>CE</td>
<td>CE</td>
</tr>
<tr>
<td>5</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>6</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>7</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>8</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>9</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>10</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>11</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>12</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td>13</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td>14</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td>15</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td>16</td>
<td>AT</td>
<td>L</td>
</tr>
<tr>
<td>17</td>
<td>AT</td>
<td>U</td>
</tr>
<tr>
<td>18</td>
<td>CE</td>
<td>U</td>
</tr>
<tr>
<td>19</td>
<td>CE</td>
<td>U</td>
</tr>
<tr>
<td>20</td>
<td>O</td>
<td>U</td>
</tr>
</tbody>
</table>

Abbreviations are as defined in Table 3.
### TOAST subtype classification.

#### 1. History/Vitals

- **High Risk Source for Emboli [1]**
  - Mechanical Prosthetic Heart Valve
  - Atrial Fibrillation
  - Sick Sinus Syndrome
  - Myocardial Infarction within 4 weeks
  - Dilated Cardiomyopathy
  - Hypertension
  - Infective Endocarditis
  - Atrial Left Ventricular Thrombus
  - Left Ventricular Thrombus

- **Medium Risk Source for Emboli [2]**
  - M > 4 weeks ago but < 6 months ago
  - Congestive Heart Failure
  - Left Ventricular Aneurysm
  - Atrial Fibrillation
  - Bipedal Heart without Atrial Fibrillation
  - Mitral Valve Prolapse
  - Atrial Rhythm Disruption
  - Atrial Septal Defect
  - Patent Foramen Ovale
  - Intracranial Aneurysm
  - Nonbacterial Endocarditis

- **No Source for Emboli [3]**

#### 2. EXAM

**A. Plain Film**
- Soft Signs
- Hard Signs

**B. MRI**
- Sagittal plane
- Axial plane
- Coronal plane

**C. Carotid Ultrasound**
- Internal carotid artery
- Common carotid artery
- Extracranial carotid bypass

**D. Doppler Ultrasound**
- Transcranial Doppler
- Intra cranial Doppler

**E. Cardiac Catheterization**
- Right Heart Catheterization
- Angiography

**F. Holter Monitor/Telemetry**
- 24-hour Holter monitor
- 3-channel Holter monitor

**G. Electrocardiogram**
- Standard 12-lead ECG
- 24-hour ambulatory ECG

**H. Echocardiogram**
- Transthoracic ECHO
- Transesophageal ECHO

**I. CT or MRI Scan**
- Brain CT Scan
- Brain MRI Scan

**J. MRI of Whole Brain**
- Brain MR
- MRI of the spine

**K. MRA of the Brain**
- MRA of the Circle of Willis
- MRA of the intracranial arteries

**L. Carotid Doppler**
- Transcranial Doppler
- Carotid Doppler

**M. Carotid Ultrasound**
- Carotid duplex ultrasound
- Carotid artery ultrasound

**N. Carotid Bypass**
- Carotid artery bypass
- Carotid endarterectomy

**O. Cerebral Angiography**
- Cerebral angiography
- Diagnostic cerebral angiography

**P. CT or MR Angiography**
- CT angiography
- MR angiography

**Q. Transcranial Doppler**
- Transcranial Doppler
- Transcranial Doppler ultrasound

**R. Brainstem Stimulation**
- Brainstem stimulation
- Brainstem evoked potentials

**S. Cerebral Blood Flow**
- Cerebral blood flow
- Cerebral blood flow measurement

**T. Carotid Stenting**
- Carotid stenting
- Carotid artery stenting

**U. Carotid Endarterectomy**
- Carotid endarterectomy
- Carotid artery endarterectomy

**V. Carotid Artery**
- Carotid artery
- Carotid artery plaque

**W. Carotid Bypass**
- Carotid artery bypass
- Carotid artery stenting

**X. Intracranial Aneurysm**
- Intracranial aneurysm
- Intracranial aneurysm rupture

**Y. Intracranial Hemorrhage**
- Intracranial hemorrhage
- Intracranial hemorrhage subarachnoid

**Z. Intracranial Mass**
- Intracranial mass
- Intracranial mass lesion

#### 3. DIAGNOSTIC TESTS

- **CT Scan**
- **MRI**
- **MRA**
- **Doppler**

**A. TOAST subtype classification**

#### 4. POSTMORTEM

- **Brainstem Stimulation**
- **Cerebral Blood Flow**
- **Carotid Stenting**
- **Carotid Endarterectomy**

**END**
We used a 2-phase process aimed at improving the reliability of the TOAST classification scheme. Creation of a computerized algorithm (available at http://hsrd.durham.med.va.gov/) eliminated variability due to differences in the interpretation of stroke-related characteristics for a given patient. Remaining discrepancies were related to differences in the abstracted data, prompting the development of a standardized manual and procedures for extracting relevant information from the medical record. This improved reliability in the classification of stroke subtype to the substantial to almost-perfect level for both intraobserver and interobserver agreement. Residual differences in diagnostic categorization were related to simple errors in abstraction or ambiguities in the medical record, occurring in 25% of cases.

In practice, having each medical record abstracted by 2 raters with the data entered into the computerized algorithm could identify this remaining variability. Abstraction forms for cases in which there is a difference in subtype diagnosis could then be reviewed (focused on CT and MRI scan, cardiovascular, and carotid ultrasound results) and the reason(s) for the discrepancies identified and resolved. Our data show that variability in stroke subtype diagnosis can be reduced to a minimum through the use of this rigorous methodology. The generalizability of these results will need to be confirmed in other settings.

**Acknowledgments**

This work was supported by grants from the Department of Veterans Affairs (Health Services Research and Development Service [SDR 93-003] and Cooperative Studies/Epidemiologic Research and Information Center Programs [CSP/ERIC 602]). The authors also wish to acknowledge Dr David Good (Wake Forest University) and Dr John R. Feussner (Chief Research Development Officer, Department of Veterans Affairs) for efforts on initial phases of the study and Maren Olsen, PhD (Duke University), for statistical advice. Kathleen Hoffmann, RN, Denice Wood, RN, and Diedra Coney, RN (Duke University), performed some of the data abstractions used in refining the procedures developed as part of this study.

**References**


**Editorial Comment**

**Classifying the Mechanisms of Ischemic Stroke**

Schemes of classification, in one form or another, continue to be used widely in many areas of stroke care. Their origins lie in the classic descriptions of parenchymal and arterial pathology, but by the 1950s they were beginning to incorporate clinically based anatomic and mechanistic subdivisions that could be used in vivo. Initially, the diagnoses of the underlying mechanisms of stroke were based mainly on clinical patterns derived retrospectively from autopsy studies, but by the 1970s the definitions had been refined to incorporate the results of frequently performed, and increasingly complex, investigations. Nevertheless, the newer classifications have continued to use a number of core mechanistic groupings (eg, large-vessel atherosclerosis, cardioembolism, small-vessel disease) that were present in earlier classifications.

In the introduction to one of the earliest attempts to synthesize the various strands of classification, Millikan1 wrote, “Our ultimate objectives are to obtain greater clarity of thinking in regard to cerebrovascular diseases, to compose a generally acceptable classification, to establish reliable criteria for diagnosis, and to promote further research in this field.”
patients who use the classifications to put the results from the research centers into the context of their daily practice.

Clearly, any scheme of classification that is used needs to be as reliable as possible, and the article by Goldstein et al describes their experience using a computer algorithm to improve the reliability of the widely used TOAST (Trial of Org 10172 in Acute Stroke Treatment) classification, a scheme that had its origins in the Stroke Data Bank classification of the 1980s and whose originators recognized that “Interobserver agreement is essential to the reliability of clinical data from cooperative studies and provides the foundation for applying research results to clinical practice.”

However, as Goldstein et al stress, their objective was only to try to standardize retrospective data collection. That is quite a different proposition from using the classification prospectively either in clinical research or to manage individual patients. Here, it is important to remember that diagnostic reliability (ie, interrater or intrarater agreement) should not be equated with diagnostic accuracy, something that requires a gold standard against which it can be judged, and which is lacking in vivo for many stroke mechanisms. Indeed, Johnson et al noted that in the absence of such a gold standard, “the merit of a classification system depends on its clarity, utility and reproducibility.”

It seems likely that the relatively modest interrater and intrarater agreement of the TOAST classification, when used prospectively in clinical practice, is in part a consequence of that rather nebulous, but extremely important, entity of clinical acumen, a complex interaction of pattern recognition and experience-influenced, repeated testing of a hypothesis against available evidence. Of course, such behavior does not sit easily alongside an administrative “bean-counting mentality,” in which it is more important to have everything in a category, regardless of the accuracy of the categorization!

So have the various classifications of stroke mechanism served us well over the last 40 years? It seems to me that they have been rather blunt tools. Even at the population level, we know relatively little about the natural history of the groupings. Furthermore, they have failed to identify subgroups of patients who would benefit from acute interventions (the original raison d’être of the TOAST classification), and where secondary prevention treatments have been more successful, they have been targeted at much more specific groups, eg, patients with atrial fibrillation or carotid stenosis. One suspects that the multiple failures of acute intervention trials will prompt a thorough review of this whole area, and although it has been shown that advances such as multimodal MR can improve the reliability of the TOAST classification, perhaps we should also consider other schemes that may have fewer links with the established clinicopathological paradigm. On the other hand, I think the current classifications do contribute to individual patient management, and harking back to Millikan’s original aspirations, I am sure that many of us will continue to use the basic skeleton of the classification to bring greater “clarity of thinking” to our clinical practice. Indeed, Gross et al observed that clinicians were able to use the classification to synthesize a number of basic clinical and investigative findings with relatively poor interrater and intrarater reliability to form a much more reliable overall diagnosis. However, I do not envisage sitting in the outpatient clinic using the algorithm of Goldstein et al on my Palm or Psion for diagnostic purposes!

Dr John Bamford, MD, FRCP, Guest Editor
St James’s University Hospital
Beckett Street
Leeds, UK
j.m.bamford@leeds.ac.uk

References
Improving the Reliability of Stroke Subgroup Classification Using the Trial of ORG 10172 in Acute Stroke Treatment (TOAST) Criteria
Larry B. Goldstein, Michael R. Jones, David B. Matchar, Lloyd J. Edwards, Jennifer Hoff, Vani Chilukuri, S. Beth Armstrong and Ronnie D. Horner

Stroke. 2001;32:1091-1097
doi: 10.1161/01.STR.32.5.1091

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/32/5/1091

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/