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/pathophysiologic 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/pathophysiologic groupings. 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
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 1).

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 provided by 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 through 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 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 similarly 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 computerized

---

**Table 1. TOAST Diagnostic Classification**

<table>
<thead>
<tr>
<th>Diagnostic Group</th>
<th>Collapsed Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atherosclerosis, probable</td>
<td>Atherosclerosis</td>
</tr>
<tr>
<td>Atherosclerosis, possible</td>
<td>Cardioembolic</td>
</tr>
<tr>
<td>Cardioembolic, probable</td>
<td>Cardioembolic</td>
</tr>
<tr>
<td>Lacunar, probable</td>
<td>Lacunar</td>
</tr>
<tr>
<td>Lacunar, possible</td>
<td>Lacunar</td>
</tr>
<tr>
<td>Other determined etiology, probable</td>
<td>Other determined etiology</td>
</tr>
<tr>
<td>Other determined etiology, possible</td>
<td>Other determined etiology</td>
</tr>
<tr>
<td>Undetermined etiology, complete evaluation</td>
<td>Undetermined etiology</td>
</tr>
<tr>
<td>Undetermined etiology, incomplete evaluation</td>
<td>Undetermined etiology</td>
</tr>
<tr>
<td>Multiple possible etiologies</td>
<td></td>
</tr>
</tbody>
</table>

Details of the definitions of each category are provided in the original publications.
TABLE 2. Synthetic Cases

<table>
<thead>
<tr>
<th>Diagnostic Group</th>
<th>Case Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Atherosclerosis, probable</td>
<td>New onset left hemiparesis with sensory deficit affecting face and arm more than leg. Left homonymous hemianopsia. Left hemispatial neglect. CT shows loss area of ill-defined loss of gray-white junction in right parietotemporal region. Doppler shows &gt;95% stenosis in right ICA. Angiogram shows 80% stenosis right ICA with branch occlusion in right MCA. Patient in normal sinus rhythm. ECG normal. Echocardiogram normal. No coagulopathy.</td>
</tr>
<tr>
<td>2 Atherosclerosis, possible</td>
<td>New onset left hemiparesis with sensory deficit affecting face and arm more than leg. Left homonymous hemianopsia. Left hemispatial neglect. CT shows loss area of ill-defined loss of gray-white junction in right parietotemporal region. Doppler shows 60% stenosis in right ICA. Angiogram shows &lt;50% stenosis in right ICA with branch occlusion in right MCA. Patient in normal sinus rhythm. ECG normal. Echocardiogram normal. No coagulopathy.</td>
</tr>
<tr>
<td>3 Cardioembolic, probable</td>
<td>New onset left hemiparesis with sensory deficit affecting face and arm more than leg. Left homonymous hemianopsia. Left hemispatial neglect. CT shows loss area of ill-defined loss of gray-white junction in right parietotemporal region. Doppler shows &lt;50% stenosis in ICAs. Patient in atrial fibrillation. Echocardiogram dilated left atrium without clot. No coagulopathy.</td>
</tr>
<tr>
<td>4 Cardioembolic, possible</td>
<td>New onset left hemiparesis with sensory deficit affecting face and arm more than leg. Left homonymous hemianopsia. Left hemispatial neglect. CT shows loss area of ill-defined loss of gray-white junction in right parietotemporal region. Patient in atrial fibrillation. Echocardiogram dilated left atrium without clot. No coagulopathy.</td>
</tr>
<tr>
<td>5 Lacunar, probable</td>
<td>History of hypertension. New onset left hemiparesis affecting face, arm, and leg to same extent. No cognitive, visual, or sensory deficits. CT shows loss area of ill-defined decreased attenuation in right internal capsule. Doppler shows &lt;50% stenosis in ICAs. Patient in normal sinus rhythm. ECG normal. Echocardiogram normal. No coagulopathy.</td>
</tr>
<tr>
<td>6 Lacunar, possible</td>
<td>History of hypertension. New onset left hemiparesis affecting face, arm, and leg to same extent. No cognitive, visual, or sensory deficits. CT shows loss area of ill-defined decreased attenuation in right internal capsule. Doppler shows &lt;50% stenosis in ICAs. Patient in normal sinus rhythm. ECG normal. Echocardiogram patent foramen ovale. No coagulopathy.</td>
</tr>
<tr>
<td>7 Other determined etiology, possible</td>
<td>History of DVT and spontaneous abortion. New onset left hemiparesis affecting face, arm, and leg to same extent. No cognitive, visual, or sensory deficits. CT shows loss area of ill-defined decreased attenuation in right internal capsule. Doppler shows &lt;50% stenosis in ICAs. Patient in normal sinus rhythm. ECG normal. Echocardiogram normal. PTT prolonged without anticoagulants.</td>
</tr>
<tr>
<td>8 Other determined etiology, probable</td>
<td>History of DVT and spontaneous abortion. Prior workup showed protein C deficiency. New onset left hemiparesis affecting face, arm, and leg to same extent. No cognitive, visual, or sensory deficits. CT shows loss area of ill-defined decreased attenuation in right internal capsule. Doppler shows &lt;50% stenosis in ICAs. Patient in normal sinus rhythm. ECG normal. Echocardiogram normal.</td>
</tr>
<tr>
<td>9 Undetermined etiology, complete evaluation</td>
<td>New onset left hemiparesis with sensory deficit affecting face and arm more than leg. Left homonymous hemianopsia. Left hemispatial neglect. CT shows loss area of ill-defined loss of gray-white junction in right parietotemporal region. Doppler shows &lt;50% stenosis in ICAs. Angiogram normal. Patient in normal sinus rhythm. ECG normal. Echocardiogram normal. No coagulopathy.</td>
</tr>
<tr>
<td>10 Undetermined etiology, incomplete evaluation</td>
<td>New onset left hemiparesis with sensory deficit affecting face and arm more than leg. Left homonymous hemianopsia. Left hemispatial neglect. CT shows loss area of ill-defined loss of gray-white junction in right parietotemporal region. Patient in normal sinus rhythm. No coagulopathy.</td>
</tr>
<tr>
<td>11 Multiple possible etiologies</td>
<td>New onset left hemiparesis with sensory deficit affecting face and arm more than leg. Left homonymous hemianopsia. Left hemispatial neglect. CT shows loss area of ill-defined loss of gray-white junction in right parietotemporal region. Doppler shows &gt;70% stenosis in right ICA. Angiogram shows 80% stenosis in right ICA with branch occlusion in right MCA. Patient in atrial fibrillation. Echocardiogram dilated left atrium without clot. No coagulopathy.</td>
</tr>
</tbody>
</table>

ICA indicates internal carotid artery; MCA, middle cerebral artery; DVT, deep vein thrombosis; and PTT, partial thromboplastin time.

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>0+</td>
</tr>
<tr>
<td>14</td>
<td>U−</td>
<td>AT+</td>
<td>AT+</td>
<td>M</td>
</tr>
</tbody>
</table>

Abbreviations are as defined in Table 3.

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

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.

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 interrater 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.
A. HISTORY/VENTRICULAR (Circle one response)

1. High Risk Source for Embol
   > Mechanical Prosthetic Heart Valve  [1]
   > Atrial Fibrillation  [2]
   > Hypertension  [1]
   > Heart Failure  [1]
   > Left Ventricular Thrombus  [2]
   > MI > 3 weeks ago but < 6 months ago  [2]
   > Congestive Heart Failure  [2]
   > Left Ventricular Aneurysm  [2]
   > Atrial Fibrillation  [2]
   > Hypertensive Heart Failure  [2]
   > Mitral Stenosis without Atrial Fibrillation  [2]
   > Atrial Valve Prolapse  [1]
   > Aortic Aneurysm  [1]
   > Atrial Septal Defect  [1]
   > Patent Foramen Ovale  [1]
   > Mitral Valve Prolapse  [1]
   > Intestinal Aneurysm  [1]
   > Nonbacterial Endocarditis  [1]
   3. No Source for Embol  [3]

B. HISTORY/ISCHEMIC VASCULAR DISEASE (Circle one response)

1. Known Extracranial Large Vessel Disease, Appropriate for Surgery  [1]
2. Negative Prior Extracranial Carotid Study  [2]
3. No Prior Extracranial Study Completed  [3]

C. HISTORY/PRIOR SPECIFIC TESTS (Circle one response)

1. Prior Specialized Tests (Hematologic, CSF, or Holostomy, Demonstrating Evidence of Underlying Cause  [1]
2. Negative Prior Specialized Test  [2]
3. No Prior Specialized Tests  [3]

ECHOCARDIOGRAM

INSTRUCTIONS: Check the TEE or TEE box (below) to indicate which type of procedure was performed. If only one type was performed, check the appropriate box, evaluate the results and circle one response below. If both types were performed, check the TEE box and evaluate TEE results below. If neither type was performed, circle “Not Done” [4] below.

TEE  [2]

High Risk Source for Embol  [1]
   > Left Ventricular Thrombus  [2]
   > Dilated Cardiomyopathy  [2]
   > Atrial Aneurysm  [2]
   > Infective Endocarditis  [2]
   > Medium Risk Source for Embol  [2]
   > Congestive Heart Failure  [2]
   > Mitral Stenosis without Atrial Fibrillation  [2]
   > Mitral Valve Prolapse  [2]
   > Mitral Annular Calcification  [2]
   > Aortic Septal Defect  [2]
   > Patent Foramen Ovale  [2]
   > Intestinal Aneurysm  [2]
   > Nonbacterial Endocarditis  [2]
   Nonbacterial Endocarditis  [2]
   Not Done  [4]

EXOGON RHYTHM STRIP

INSTRUCTIONS: Check the TEE or TEE box (below) to indicate which type of procedure was performed. If only one type was performed, check the appropriate box, evaluate the results and circle one response below. If both types were performed, check the TEE box and evaluate TEE results below. If neither type was performed, circle “Not Done” [4] below.

4. POSTMORTEM (Circle one response)

A. Blunt or Hemorrhagic Infarct, an Atheromatous Lesion (Possibly with Adherent Thrombus) of an Appropriate, Large Extracranial Artery or Possibly Occlusion of Superficial or Cortical Arteries  [1]

B. Blunt or Hemorrhagic Infarct(s), Arterial Occlusion by an Embolus and Underlying Heart Disease  [2]
   > Small Deep Infarct in the Territory of a Penetrating Artery  [3]
   > Blunt or Hemorrhagic Infarct without Significant Atherosclerosis, Normal Cardiac Pathology and Arterial Occlusion of Specific Etiology  [4]
   > Infarction but no Specific Vascular Pathology  [5]
   > Not Done  [6]

TOAST subtype classification.

A. ARTERIAL VASCULAR ETIOLOGY (Circle one response)

1. Atrial Fibrillation  [2]
2. Evidence of Systemic Involvement  [2]

B. CEREBRAL ARTERIOGRAPH (Circle one response)

1. Occlusion of a 50% Stenosis or 2mm Occlusion of an Appropriate Large Intracranial or Extracranial Artery  [1]
2. < 50% Stenosis or < 2mm Occlusion of an Appropriate Intracranial or Extracranial Artery with no Occlusion of Appropriate Stem, Dives or Branch Artery  [2]
3. 50% Stenosis or > 2mm Occlusion of an Appropriate Intracranial or Extracranial Artery with Occlusion of Appropriate Stem, Dives or Branch Artery  [3]
4. Specific, Nonthrombotic, Intracranial or Extracranial Vascular Pathology  [4]

E. CARDIOVASCULAR EVALUATION (Circle one response)

INSTRUCTIONS: The following 3 sections (ECHO-CARDIOGRAM, EXOGON RHYTHM STRIP, and HOLTER MONITOR/TENSITY) are included to indicate the scoring for this section. If there is a high or medium risk source for embol, the highest level of risk indicated in ANY of the 3 sections is “High Risk Source for Embol” [1], then this section should be scored [1]. If the highest level of risk indicated in ANY of the 3 sections is “Medium Risk Source for Embol” [2], then this section should be scored [2]. If Echocardiogram and EXOG are scored “Normal” [3], and Holter/Incompetent Workup or Workup Not Done [4].

1. High Risk Source for Embol  [1]
3. No High or Medium Risk Source for Embol  [3]
4. Incompetent Workup or Workup Not Done  [4]
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 over the years the definitions have 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.”

One suspects that, outside the centers of stroke research, such aspirations were considerably in advance of their time, and that for the majority of stroke patients worldwide, meaningful (if fairly basic) subclassification became a reality only with the advent of CT and ultrasound scanning. Most of the early research that used mechanistic classifications was observational epidemiology, most notably that from the Mayo Clinic2 and later from the Stroke Data Bank collaborators3 and the Lausanne group.4 When the original classification was reviewed some 17 years later, at a time when the growth in stroke research in general, and clinical trials in particular, was beginning to expand dramatically, Millikan5 wrote: “It continues to be evident that in such a complex set of clinical-pathophysiological phenomena some standard reference language or set of definitions should be used or the literature of investigation will be uninterpretable.”

The point about the need for a common language of communication continues to be of paramount importance in an era when the uses of a classification have broadened from observational epidemiology to clinical trials and, more recently, to the purchasing of healthcare. Perhaps most importantly, there are the individual clinicians caring for stroke
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."7

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 al8 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,9,10 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,11 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 al12 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

*Stroke* is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2001 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/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/