Stroke Classification
A Personal View
Louis R. Caplan, MD

Because all history is to some extent personal, and because I have been active throughout attempts at classification, I was asked to provide a personal account of stroke subtyping and stroke registries bringing the topic up to date.

Early Studies
Clinicopathological studies during the first half of the 20th century focused on clinical signs in patients who died after stroke. Brain hemorrhages and infarcts were recognized at necropsy, but could they be separated during life? In 1935, Aring and Merritt analyzed 245 stroke patients studied clinically and at necropsy at the Boston City Hospital. They reported demographic, epidemiological, historical, and clinical data to determine which features differentiated thrombosis from hemorrhage. Hemorrhages (intracerebral or subarachnoid) were found in 15% of patients and 82% had ischemic infarcts called “thrombotic;” however, only 3% were considered to have cardiogenic embolic brain infarcts. The main features favoring hemorrhage were headache, vomiting, impaired consciousness, progression of the clinical neurological deficit, bloody spinal fluid, and increased spinal fluid pressure. This series was biased toward patients with large fatal hemorrhages and infarcts. Later, Dalsgaard-Nielsen in Scandinavia compiled a series of 1,000 stroke patients, and clinicians at the Mayo Clinic analyzed series of patients seen during 1945 through 1954 and from 1955 through 1969.

These studies were based on retrospective chart reviews and all preceded CT. Infarcts were >4-times more common than hemorrhages. Infarcts were classified as embolic only if the patients had rheumatic heart disease or recent myocardial infarction. Rates of brain embolism using these criteria ranged from 3% to 8%. Nonembolic brain infarcts were assumed to be “thrombotic” and related to occlusion of brain supplying large arteries. Carotid and vertebral artery disease in the neck and lacunar infarction were not diagnoses included in any of these early studies.

The Harvard Stroke Registry
In 1971, Howard Bleich, a nephrologist at the Beth Israel Hospital in Boston, Massachusetts, published a computer-based diagnostic program guiding diagnosis and treatment of acid–base abnormalities encountered in hospitalized patients. This program required clinicians to enter patients’ clinical information into a computer. The computer used this data to identify an acid–base abnormality, eg, metabolic acidosis, and to calculate the content and amount of infusate needed to correct the imbalance. After using the program, clinicians learned the data needed and, by reading references appended to the program, learned how the computer made diagnoses and calculations. Doctors could then master the process without using the computer. The program used a large PDP-11 computer housed in a temperature-controlled room to ensure that the computer would not break and lose all data. Bleich asked me to construct a computer-based program that could diagnose all neurological disease similar to the program for renal and acid–base abnormalities. I found the task of confronting all neurological conditions to be much too formidable, but after cajoling I agreed to construct a program aimed only at stroke diagnosis that considered only stroke etiology subtyping and did not consider brain localization.

At that time, I was completely computer illiterate. A very bright medical student and Massachusetts Institutes of Technology graduate, Robert Goldstein, was assigned to work with me. Goldstein and Bleich explored with me potential strategies that could be used to construct a prospective computer-based diagnostic program. Branching logic and pattern-matching proved unsuitable because no large compilation of patients to pattern-match was available, and there were few branches that were absolutely indicative of 1 stroke subtype. The only feasible strategy was using a Bayesian analysis. Bayes theorem involved calculating the likelihood of a particular diagnosis based on the frequency of a condition in the population and the frequency of particular findings in that condition.

I buried myself in the library but could not find relevant data from the literature. Carotid artery disease, cardiac lesions other than rheumatic and acute myocardial infarction, and lacunar infarcts were recently described and there was no quantitative data about the various clinical findings in these conditions. Brain infarction was known to be present in ~80% of strokes, and subarachnoid hemorrhage and intracerebral hemorrhage each accounted for ~10% of strokes. So, in an emergency department, the diagnosis of brain infarction would be correct 4 times out of 5, but if one diagnosed intracerebral hemorrhage in all patients, the accuracy would be only 10%. The frequency of individual findings, various risk factors, headache preceding...
stroke and at onset, TIA preceding stroke, progression of the clinical deficit, vomiting, loss of consciousness, and so on, were not available from the literature. The next step was to ask experienced clinicians to estimate the various frequencies. Drs Miller Fisher, Raymond Adams, Richard Tyler, J.P. Mohr, and I made frequency guesses. To our surprise, the various estimates were all over the ballpark. The estimate of some single findings varied from 80% to 10%. A Massachusetts Institutes of Technology graduate averaged the estimates, and Goldstein, Bleich, and I constructed a computer-based diagnostic program that we used to separate subarachnoid hemorrhage, intracerebral hemorrhage, brain embolism, large artery-related brain infarction, and lacunar infarction using clinical data.7

The main outcome of this attempt at computer diagnosis was to make me aware that there were little data available about the frequency of various clinical findings in the various subtypes of stroke. So, I decided that we needed to prospectively collect that data. The Harvard Cooperative Stroke registry was born. I asked Dr J.P. Mohr, who was the chief of the Stroke Service at the Massachusetts General Hospital, to join me in collecting data. Mohr and I met with Bleich, Warner Slack, and John Melski, then a computer fellow at the Beth Israel Hospital, on a weekly basis to decide on classification criteria, entry items, and terms.

In the Harvard Stroke Registry, the diagnosis of embolism was defined as blockage of a distal vessel by material generated at a distance. The evidence used related mostly to the recipient artery. The source could be cardiac, aortic, intra-arterial, or unknown. At that time (early 1970s), the major diagnostic tool was cerebral angiography and echocardiography was not available.

The frequencies of diagnosis among the 694 patients in the Harvard Stroke Registry were: thrombosis, 53% (34% large artery, 19% lacunar); embolism, 31%; intracerebral hemorrhage, 10%; and subarachnoid hemorrhage, 6%.5 Some cases were classified as unsettled cause and some as unusual cause, and these were not included in the frequencies cited. Also noted were the basis of the diagnosis (clinical, angiography, necropsy) and confidence that the diagnosis was correct, high or low. The basis for the diagnoses in approximately half of the patients was only clinical; 45% had cerebral angiography, only 3% had CT scans, and 4% had necropsies.

The Harvard Stroke Registry was the first prospective published database on any medical condition. The results were submitted in a manuscript to the New England Journal of Medicine. The rejection letter commented that computers would never take hold in medicine. The registry ended when Dr Mohr and I both left Boston. Knowledge that the advent of CT scanning would forever change diagnosis led us to begin new data collections based more on brain imaging.

The Stroke Data Bank

Although the National Institute of Neurological Disease and Stroke had rejected the original Harvard Stroke Registry application for funding, they became impressed with the idea of a government-sponsored prospective stroke registry and issued a request for proposals for a multicenter Stroke Registry. The patients studied in the Harvard Stroke registry were from one institution and were predominantly white. More geographical and racial diversity was important. The Stroke Data Bank was created and funded by the National Institute of Neurological Disease and Stroke. The institutions and principle investigators were J.P. Mohr at Columbia University, Tom Price at the University of Maryland, Phil Wolfe at Boston University, and I at Michael Reese Hospital in Chicago. During the collection period, Dan Hier became the principle investigator in Chicago.

During the first year of the grant, the principle investigators and their assistants (including Ralph Sacco and Stan Tuhrim) met often in Bethesda with National Institute of Neurological Disease and Stroke designates to generate the terms, data entry items, main queries to be answered, and statistical methods. During 1983 to 1986, 1 805 patients (54% black) were prospectively examined and the results were published during the late 1980s and early 1990s.9–15

In the interval between patient enrollment in the Harvard Stroke Registry and the Stroke Data Bank, technology had changed. Nearly all patients (97%) in the Stroke Data Bank had brain CT scans and many had echocardiography. Few had cerebral angiography; MRI, MRA, and CTA were not available. Cardiac conditions were characterized as high risk or medium risk. Cortical or cortical/subcortical infarcts in the Stroke Data Bank were labeled infarct of undetermined cause, unless there was a high-risk cardiac source. If a large neck artery was occluded and studies showed that a distal branch was also occluded, then the diagnostic category was tandem arterial pathology. Only 6% of patients qualified for large artery stenosis or occlusion. Reacting to the criticism that the Harvard Stroke Registry had overdia- gnosed embolism, the Stroke Data Bank used different criteria that emphasized the embolic source rather than the recipient artery. The frequencies were: infarcts, 70% (large artery 6%, tandem pathology 4%, lacunes 19%, cardiac source 14%, undetermined cause 28%); hemorrhages, 27% (subarachnoid hemorrhage and intracerebral hemorrhage each 13%); and other, 3%. The sum of those diagnosed as cardiac source embolism and infarct unknown cause (lacunar infarcts without large artery occlusions or tandem arterial pathology) totaled 42%. These patients were classified as embolism in the Harvard Stroke Registry.

The presence of CT scan data now allowed study of the distribution of infarcts. The infarct lesions were sketched on standard grid template grids that covered 10 CT sections. Correlations could be made with neurological signs and stroke mechanisms.11–15

New England Medical Center Posterior Circulation Registry17–25

After moving to Boston in 1984, I began a registry limited to patients studied at the New England Medical Center (NEMC) who had ischemic disease of the vertebral basilar arterial system. The aims and rationale of the NEMC Posterior Circulation Registry (NEMC-PCR) were quite different than those in the Harvard Stroke Registry and Stroke Data Bank. The focus was narrowed to ischemia (rather than all strokes) and to the posterior circulation rather than the entire brain. By the end of the 1980s, CT scans became generally available and brain hemorrhages were readily diagnosed by imaging. Attention turned to differentiating the various subtypes of brain ischemia. By then, the TOAST trial investigators had published criteria for diagnosing...
mechanisms. Outcomes also could be analyzed in relation to locations and symptoms and signs to vascular lesions and stroke mechanisms, and to relate brain imaging (predominantly MRI).

By the 1990s, it was possible to analyze the presence of vascular lesions, brain locations, stroke mechanisms, and neurological symptoms and signs to vascular lesions and stroke mechanisms, and to relate brain locations and symptoms and signs to vascular lesions and stroke mechanisms. Outcomes also could be analyzed in relation to brain location, vascular lesions, and stroke mechanisms. Because designation of stroke mechanism was controversial in the Harvard Stroke Registry and the Stroke Data Bank, we chose a different strategy in the NEMC-PCR. After review and review of each patient’s studies, Drs Caplan, Pessin, DeWitt, Tapia, and the Stroke fellows present at that time decided on the most likely stroke mechanism but also recognized any other potential stroke mechanisms. Tabulation of stroke mechanisms consisted of 2 columns, one column that contained the consensus selection of the most likely stroke mechanism and a second column of ranges of potential mechanisms. The lowest percentage would be if the minimum number with that attribute was responsible, and the highest frequency in the range would be if all were culpable, for example, if a patient with a bifid stenotic aortic valve had occlusion of the right intracranial vertebral artery and had experienced multiple TIA referable to territory supplied by that artery, followed by a lateral medullary infarct. The single most likely mechanism would be large artery occlusion disease. However, because there was a potential cardiac source (although unlikely to be culpable in this case), it would be included in the cardiac embolism source second column range.

Similarly, in a patient with atrial fibrillation with sudden onset of an occipital infarct and also 50% stenosis of the left vertebral artery, the most likely mechanism would be cardiac-source embolism, but the large artery lesion would be included within the range of large artery disease. In this way, all potential mechanisms would be tabulated.

In the NEMC-PCR, brain lesions were categorized as involving proximal, middle, and distal intracranial posterior circulation territories. Proximal posterior circulation territory included intracranial vertebral arteries-supplied regions, the medulla and posterior inferior cerebellar artery territory of the cerebellum. The middle intracranial posterior circulation territory included brain regions supplied by the basilar artery up to its superior cerebellar artery branches, the pons and anterior inferior cerebellar artery-supplied cerebellum. The distal territory included regions supplied by the rostral basilar artery, superior cerebellar artery, posterior cerebral artery, and their penetrating artery branches—midbrain, thalamus, superior cerebellar artery cerebellum, and posterior cerebral artery territories.

Brain imaging (CT and/or MRI) was performed on all patients (>80% had MRI). Vascular imaging was also performed in all patients (80% had contrast catheter angiography). Ultrasound was widely used (>80% of patients had transcranial Doppler). Echocardiography and heart rhythm monitoring were performed when clinically indicated.

Our major aims in the registry were to clarify outcomes and the frequency of various stroke mechanisms and vascular lesions, and to understand the relationship of these mechanisms and vascular lesions to the topography of brain infarcts. Figure 1 shows the distribution of infarcts within the proximal territory, and Figure 2 shows the distribution of lesions according to stroke mechanisms.

A-S-C-O (Phenotypic) Stroke Classification

During the past decade, I and a group of international stroke leaders (Amarenco [France], Donnan [Australia], Bogousslavsky [Switzerland], and Hennerici [Germany]) met and devised a new type of Ischemic Stroke Registry. We called it the A-S-C-O (Phenotypic) Stroke Classification, in which A indicated atherosclerosis, S indicated small vessel disease, C indicated cardiac disease, and O indicated other. The registry, like
Table. Grades of Likelihood Within Each Phenotype

| Grade 1: Definitely a potential cause of the index stroke |
| Grade 2: Causality uncertain |
| Grade 3: Unlikely a direct cause of the index stroke but disease is present |
| Grade 0: No disease |

Levels of evidence according to diagnostic testing

A: Direct demonstration by gold standard diagnostic tests or criteria

B: Indirect evidence or less sensitive or specific tests or criteria

C: Weak evidence in the absence of specific tests or criteria

the NEMC-PCR, was limited to brain ischemia. Data entry made information useful for a variety of purposes, such as describing patient characteristics in therapeutic trials, grouping patients in epidemiological studies, phenotyping in genetic studies, and classifying patients for therapeutic decision-making in daily practice. We defined each subtype clearly. The system was designed to include the potential likelihood of the presence of each stroke subtype and the probability that it caused the index event. Definitions and the levels of evidence were agreed on. The likelihood of the presence of each phenotype was assigned grades and levels of evidence (Table).

Disclosures

None.

References

Stroke Classification: A Personal View
Louis R. Caplan

Stroke. 2011;42:S3-S6; originally published online December 16, 2010;
doi: 10.1161/STRÓKEAHA.110.594630
Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2010 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/42/1_suppl_1/S3

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/