THE CLINICAL STROKE DATA BANK study was initiated in 1978 by the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS) in order to amass information systematically on a large number of patients as a data source for clinical research. Dr. Donald Tower, then Director of NINCDS, saw the possibilities of a data bank to stimulate clinical research on neurovascular disorders. Dr. Murray Goldstein, Director of the Stroke and Trauma Program at that time, and William Weiss, Chief of the Office of Biometry and Field Studies, supported the development of data banks for stroke and for traumatic coma similar to those already under way in ischemic heart disease and rheumatology. It was anticipated that this research resource would aid in developing algorithms for diagnosis of various types of stroke, characterizing their clinical course and outcome, identifying prognostic variables, and planning clinical trials.

Despite the expanding interests in clinical trials for determining the value of specific medications for cerebrovascular disease, such trials continue to have methodologic problems as described by Spence and Donner. There is, for example, a need for precise diagnostic categories for various types of stroke, better descriptions of patient outcome, and more accurate estimates of patient accrual rates. These workers also point out the problems in comparing results across clinical trials caused by heterogeneous diagnostic criteria. In addition, Caplan has emphasized the need for utilization of CT and angiogram findings to establish etiologically meaningful definitions of stroke type. The Stroke Data Bank is intended to be a major aid in solving such definitional and methodological problems.

During the pilot phase of the study, which was completed in 1981, 938 patients with cerebral infarction or hemorrhage and 220 patients with transient ischemic attacks were entered into the data bank by four collaborating university centers. These clinical centers, chosen by competitive review, were Boston University, Duke University, University of Maryland, and the University of South Alabama. The objectives of the pilot phase were to determine whether neurologists at different institutions could collaborate in specifying research issues, agree on the clinical and laboratory data to be collected, develop uniform data collection forms that incorporate operational definitions, and collect data according to a common protocol. The pilot study was successful in each of these aspects and clarified the definitional and data collection issues for the main study, which began in late 1982. This paper briefly describes the methods and reports the preliminary findings of the pilot phase, and states the research issues for the main phase of the Stroke Data Bank.
Unlike a clinical trial, the data bank is not designed to evaluate specific forms of treatment.

The Stroke Data Bank is a collection of pertinent information about the onset, symptomatology, clinical course, therapy, and outcome of patients with various types of stroke. This information is collected prospectively beginning with the patient's initial hospitalization and continuing at specified follow-up intervals for the duration of the study. The data are based on direct patient examination and interviews rather than by abstracting medical records. The data are transmitted by micro-computers at each center and are stored in a central computer facility which provides timely access for all centers. The pilot Stroke Data Bank captured the medical history, clinical course and outcome of large numbers of patients, using standardized patient descriptors, and organized data collection procedures. The data were collected according to a common data collection protocol and provided the information needed for specific clinical research studies.

Methods

At each center, patients eligible for the study were identified by daily surveillance of admission records, neurology consultations and through contacts with nursing and medical staff personnel. Eligibility criteria specified patients aged 20 to 79 who had a stroke within the preceding month or who had been admitted to a hospital because of transient cerebral ischemia (TIA) within the past six months. One center also investigated stroke incidence in a local geographic area and included all area patients aged 20 or older in the data bank. Patients with strokes were excluded when the disorder was caused by unusual conditions, such as fibromuscular dysplasia, arteritis, venous lesions, blood dyscrasia, brain tumor, head trauma, radiation or congenital arterial lesions other than cerebral aneurysm or arteriovenous malformation. Also excluded were patients with a serious illness such as systemic cancer, liver disease, or renal failure, unlikely to survive for two year follow-up post-stroke.

Patients who met the eligibility criteria were usually examined by the principal investigator at each center soon after the onset of stroke, often within 48 hours of admission, and were monitored frequently during the acute phase. Examinations were performed at least weekly and, when possible, during worsening of the illness. After discharge from hospital, follow-up examinations were done at 3, 6, 12, and 24 month intervals. Data were entered and stored on micro-computers located within each participating hospital center, and then sent electronically to the Data Bank Maintenance Center (DBMC), a central storage facility at Stanford University Medical Center and maintained by the Databank Network. In the main phase, data are sent from the clinical centers to the DBMC utilizing an automated procedure which reduces the costs of data transmission.

The initial eighteen months of the project were used to develop and refine data collection forms and definitions of the clinical manifestations and to develop test data entry methods. During the final eighteen months, the major focus of the project was to enroll patients and to determine adherence to the data collection protocol.

The set of data collected included demographic factors and medical histories, neurological examinations, medical treatment and details of surgical operations, and functional assessments. At discharge from hospital, a diagnostic sheet (fig. 1) contained the type of stroke, location of the brain and arterial lesions, and the basis for these diagnoses.

A key aspect in the implementation of the data bank was the involvement of a multidisciplinary team. Neurologists, nurse clinicians, epidemiologists, statisticians, and computer scientists worked on all aspects of the study, from research design through data analysis.

Quality Assurance

Methods for ensuring the quality of the data were directed at all aspects of the study. Data collection forms were designed to follow a logical sequence of patient examinations, and if enough space for comment was provided, all data items were included. Where space permitted, definitions and instructions were included on the form. The micro-computer was programmed to check all data as they were entered and to prohibit entry of out-of-range values or data forming
The data bank, as a study design, provides data appropriate for modelling prognosis and for identifying factors related to various outcomes. Through such descriptive studies, hypotheses are generated, and these may lead to the design of additional studies.

Although the data bank is based upon uniform methods of selecting patients and collecting data, certain inter-center differences cannot be avoided. Some differences are a reflection of diversity in local patient populations, catchment areas, and referral patterns. The appropriateness of pooling data across centers must be decided upon both substantive and statistical bases. A major consideration is the consistency of a relationship across centers. In the pilot Stroke Data Bank, for example, patients with lacunar infarcts had the lowest 30-day case-fatality rates among cerebral infarcts, at all centers.

**Results**

From January, 1980, through July, 1981, the pilot Stroke Data Bank enrolled 1158 patients: 708 (61%) with cerebral infarction, 220 (19%) with intracranial hemorrhage, and 220 (19%) patients with admission diagnosis of recent TIA. Ten patients had other types of stroke. The distribution of patients across centers for each type of stroke diagnosis is given in table 1. Center 2 is known to be a referral center for patients with TIA and, not surprisingly, reported the highest proportion of such diagnoses. Likewise, Center 3 is noted for the surgical management of intracranial hemorrhage and reported the highest proportions of subarachnoid hemorrhage and parenchymatous hemorrhage. Differences in the application of the diagnostic definitions for stroke types were, in part, responsible for the inter-center variation in proportions of cerebral infarction subtypes.

**Demographic Data**

The sex and race distribution of the cases in the data bank reflected the centers’ referral patterns. Center 1 included a Veteran’s Administration (VA) Hospital and, therefore, had a higher male:female ratio than the other centers. The number of patients within each race and sex group, categorized by type of stroke diagnosis,
is shown in Table 2. Among whites, there were almost twice as many men as women for every subtype of cerebral infarction. Although white women had the lowest proportion of lacunar infarctions, there was no association between sex and type of cerebral infarction for whites (excluding VA patients). Among blacks, excluding VA patients, there was a significant association between sex and type of cerebral infarction, and black women outnumbered black men in one category. Although the data bank combined hospital-based case series, some results corresponded to well-established demographic patterns of stroke distribution. For example, men accounted for only a third of the patients with subarachnoid hemorrhage, but they provided more than half the patients with parenchymatous bleeding. Overall, the median age at diagnosis was 63 years (Table 2). Patients with subarachnoid hemorrhages were youngest, with a median age of 49 years.

**Diagnosis**

The criteria for the diagnosis of stroke included a certainty component based on the availability of confirmatory laboratory findings. Confirmatory data were distinguished from findings consistent with, but not diagnostic of, a stroke type. For example, clear spinal fluid is consistent with cerebral infarction due to atherosclerosis, but is not the key diagnostic element for this stroke type. Table 3 displays stroke types and their confirmatory diagnostic support. The source listed is the one with the most definitive findings as stated in the stroke types definitions. CT visible lacunae were found in only 31% of the diagnosed lacunar infarctions, and only 24% (41 of 172) of diagnosed atherothrombotic infarction cases were based on angiographic (36 cases), surgical (one case) or post-mortem (four cases) evidence of atherosclerosis. Yet 90% of the lacunar cases had one or more CT scans and 46% of the

---

**Table 2** Stroke Diagnosis by Race and Sex and Median Ages

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Total Cases</th>
<th>Median Age</th>
<th>White</th>
<th>Black</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>White Men</td>
<td>White Women</td>
</tr>
<tr>
<td>Cerebral infarctions</td>
<td>708</td>
<td>65</td>
<td>243(35)</td>
<td>127(18)</td>
</tr>
<tr>
<td>atherosclerotic</td>
<td>172</td>
<td>64</td>
<td>61(36)</td>
<td>31(18)</td>
</tr>
<tr>
<td>embolic</td>
<td>200</td>
<td>64</td>
<td>79(40)</td>
<td>40(20)</td>
</tr>
<tr>
<td>lacunar</td>
<td>100</td>
<td>63</td>
<td>31(32)</td>
<td>9(9)</td>
</tr>
<tr>
<td>etiology unproved</td>
<td>236</td>
<td>67</td>
<td>72(30)</td>
<td>47(20)</td>
</tr>
<tr>
<td>Hemorrhages</td>
<td>220</td>
<td>53</td>
<td>45(21)</td>
<td>52(24)</td>
</tr>
<tr>
<td>parenchymatous</td>
<td>101</td>
<td>56</td>
<td>23(23)</td>
<td>15(15)</td>
</tr>
<tr>
<td>subarachnoid</td>
<td>119</td>
<td>49</td>
<td>22(19)</td>
<td>37(31)</td>
</tr>
<tr>
<td>TIA</td>
<td>220</td>
<td>62</td>
<td>103(48)</td>
<td>65(30)</td>
</tr>
<tr>
<td>Total</td>
<td>11484</td>
<td>63</td>
<td>391</td>
<td>244</td>
</tr>
</tbody>
</table>

*Fourteen patients were of other or unknown race: 7 infarction, 4 hemorrhage and 3 TIA patients.
†The numbers in parentheses are row percentages.
∥Patients with other types of stroke are omitted from this table.

---

**Table 3** Certainty of Stroke Diagnosis

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Total Cases*</th>
<th>Clinical Impression† Only</th>
<th>Non-Confirmatory Laboratory‡</th>
<th>Diagnostic source</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>CT Scan</td>
<td>Angiography</td>
<td>Surgery</td>
</tr>
<tr>
<td>Cerebral infarctions</td>
<td>708</td>
<td>316(45)</td>
<td>253(36)</td>
<td>67(9)</td>
</tr>
<tr>
<td>atherosclerotic</td>
<td>172</td>
<td>61(35)</td>
<td>62(36)</td>
<td>36(21)</td>
</tr>
<tr>
<td>embolic</td>
<td>200</td>
<td>71(35)</td>
<td>77(39)</td>
<td>31(15)</td>
</tr>
<tr>
<td>lacunar</td>
<td>100</td>
<td>62(62)</td>
<td>31(31)</td>
<td>0</td>
</tr>
<tr>
<td>etiology unproved</td>
<td>236</td>
<td>122(52)</td>
<td>83(35)</td>
<td>0</td>
</tr>
<tr>
<td>Hemorrhages</td>
<td>220</td>
<td>24(11)</td>
<td>150(68)</td>
<td>19(9)</td>
</tr>
<tr>
<td>parenchymatous</td>
<td>101</td>
<td>1(1)</td>
<td>90(89)</td>
<td>2(2)</td>
</tr>
<tr>
<td>subarachnoid</td>
<td>119</td>
<td>23(19)</td>
<td>60(50)</td>
<td>17(14)</td>
</tr>
<tr>
<td>Total</td>
<td>928</td>
<td>340</td>
<td>403</td>
<td>86</td>
</tr>
</tbody>
</table>

*220 TIA patients were diagnosed by clinical impression only and are omitted from this table. The 10 patients with other types of stroke are excluded as well.
†Patients without CT scan or angiography.
‡CT scans or angiography were performed but diagnosis was not corroborated.
§Numbers in parentheses are row percentages.
atherosclerotic infarction cases had angiograms. Overall, 90% of all 938 stroke patients were studied by CT scanning, and 42% underwent angiography.

**Mortality**

Fourteen percent or 117 of the 828 patients with cerebral infarction or hemorrhage died within 30 days after onset of this illness. The 30-day case-fatality rates differed by stroke type (table 4). Since entry into this study was predicated on admission to a teaching hospital and direct patient examination by study neurologists, a number of early (1–2 days after admission) stroke deaths may have been excluded. In addition, the exclusion of seriously ill patients, e.g., cancer, served to reduce the 30-day case-fatality rates in this series. One hundred patients from the South Alabama geographical area admitted to hospitals other than the University of South Alabama Medical Center were excluded from table 4.

**Discussion**

This pilot project demonstrated that the data bank concept worked, and provided the methods for building a clinical research data resource. Based on the pilot methods, the Stroke Data Bank began in 1982, using improved data definitions and data collection procedures. The Stroke Data Bank clinical centers, again chosen by competitive review, are Boston University, Michael Reese Hospital, University of Maryland, and University of South Alabama (subcontract to Neurological Institute at Columbia University). The studies to be accomplished using the Stroke Data Bank include the characterization of evolving stroke, clinical course and outcome of subtypes of stroke, identification of the complication-prone patient, and predictors of outcome. Studies on depression, piloted by the University of Maryland, will be carried forward during the main phase. In addition, the diagnostic definitions used in the pilot have been refined, and an algorithm for determining stroke type created. The clinical diagnosis will be compared with the algorithm diagnosis to evaluate the ability of the latter method to predict patient course. Because interobserver variability is a key element in the reliability of multicentered study data, agreement among clinicians on clinical evaluations, CT scan readings and diagnosis will be assessed through a series of special studies.

It is anticipated that the Stroke Data Bank will generate recommendations for a standard diagnostic clinical evaluation, as well as planning information for clinical trials. Since there is no standard patient work-up to establish the diagnosis of stroke, future recommendations will include the most reliable time intervals for laboratory testing, such as CT scanning and angiography. This information will be of use not only in clinical care but also in stroke research.

The Stroke Data Bank also offers direct benefits for planning and implementing of clinical trials. In planning such trials, the data bank may provide experience-based estimates of accrual rates of well-defined groups of cerebrovascular cases at these and perhaps other university centers. It will also provide historical data on the success rate of current treatment as well as on the approximate time to collect end-points for comparing successful versus unsuccessful therapies. It will aid in predicting how changes in eligibility criteria will affect patient recruitment. The data will also describe the characteristics of patients receiving standard treatment, identify trends, and provide data on complications of surgical or medical treatment.

The computer technology utilized in the data bank can be directly applied to the needs of multicentered clinical trials. The computer can generate randomization procedures, track patients throughout the trial, generate forms, test the quality of data, provide routine reports on patient acquisition, missing data and other relevant information, and electronically transmit mail among centers. The electronic mail procedure has proved to be a time saving communication technique among the data bank clinical centers.

Since the Stroke Data Bank clinical centers are uni-

### Table 4 Thirty-Day Case Fatality Rates by Diagnosis

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Total cases*</th>
<th>Case-fatality rates*</th>
<th>Deaths occurring within 30 days of onset</th>
<th>Patients lost to follow-up in less than 30 days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cerebral infarctions</td>
<td>620</td>
<td>8.5</td>
<td>52</td>
<td>20</td>
</tr>
<tr>
<td>atherosclerotic</td>
<td>170</td>
<td>11.9</td>
<td>20</td>
<td>5</td>
</tr>
<tr>
<td>embolic</td>
<td>180</td>
<td>7.3</td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td>lacunar</td>
<td>93</td>
<td>1.1</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>etiology unproved</td>
<td>177</td>
<td>10.3</td>
<td>18</td>
<td>5</td>
</tr>
<tr>
<td>Hemorrhages</td>
<td>208</td>
<td>31.3</td>
<td>65</td>
<td>0</td>
</tr>
<tr>
<td>parenchymatous</td>
<td>94</td>
<td>34.0</td>
<td>32</td>
<td>0</td>
</tr>
<tr>
<td>subarachnoid</td>
<td>114</td>
<td>28.9</td>
<td>33</td>
<td>0</td>
</tr>
</tbody>
</table>

*One hundred patients admitted to hospitals other than the University of South Alabama Medical Center and studied in conjunction with the South Alabama population study are excluded from this table, as are 220 TIA patients and 10 patients with other types of stroke.

Case-fatality rates = \( \frac{\text{Deaths}}{(\text{Total-lost}) + 0.5 \times \text{lost}} \times 100 \)
versity hospitals, they may attract a cohort of patients with more severe strokes than do community hospitals. However, these are the types of centers most likely to participate in therapeutic trials and the definitions and methods of this project can be applied to other clinical studies of stroke and to clinical trials. This information, on subtypes of stroke, on complications, and on outcome, should aid in the development of appropriate therapy, and improved management of stroke patients with various types of cerebrovascular disease.

Acknowledgements
We would like to thank the following persons for their efforts on behalf of the Stroke Data Bank.

Boston University: Russell Butler, M.D., Eloise Licata-Gehr, R.N., M.S., Margaret Kelley-Hayes, R.N. and Paula Murray.


University of Maryland: Maryann Banko, R.N., James Regges, M.D., Thomas Ducker, M.D., Aries Apostolides, Ph.D., Darrell Giles, Trudy Greene and Peggy Allen.

University of South Alabama: Sarah C. Cunningham, R.N. and Alecia G. Parra.

Stanford University: Denny McShane, M.D., James Stansich and Cathy Williams.

National Institutes of Health: James Dambrosia, Ph.D. and Margaret Meadows.

We would also like to thank Mr. William Weiss, NINCSD, for his support in the development of the data banks. Ms. Barbara Nichols for her extraordinary and tireless efforts as computer coordinator for the data bank, Ms. Irene Fishman for her editorial assistance and Deborah Trout for her typing.

This work was supported by the following research contracts from the National Institute for Neurological and Communicative Disorders and Stroke N01-NS-8-2396-2397-2398 and NOI-NS-9-2302.

References


4. Caplan LR: Are terms such as completed Stroke or RIND of continued usefulness? Stroke 14: 431–433, 1983


Appendix 1

1. Brain Infarction due to atherosclerosis and distal insufficiency

CT Scan
Distal infarction involving the superior fronto-central region of a cerebral hemisphere, with or without extension into the medial frontal and lower lateral convexity in an anatomic pattern reflecting the location of the “border zones” between the major cerebral arteries. Cases with full anterior to one hemisphere infarction are consistent with athero-thrombotic carotid occlusion or steno-occlusion. CT changes involve low density without high density components. Middle cerebral or basilar territory occlusions due to thrombus cannot be distinguished by CT scan from embolism when low density is the only abnormality. In such cases, although CT scan is done, the diagnostic source is “clinical choice.”

Angiogram
Oclusion or stenosis of the internal carotid at its origin or in the siphon with its intracranial branches patent. Multiple distal intracranial branch defects shall be considered embolic. Basilar stenosis or occlusion on angiogram will be considered the mechanism for brainstem stroke even if the syndrome is lacunar.

Clinical Choice
Cases whose clinical features suggest atherothrombotic thrombosis, but in which the criteria for CT scan or angiogram diagnosis are not met Clinical features considered include prior TIA’s in the same territory, stepwise increment of deficit without fluctuation, increments separated by more than three days in time, or signs of progressive brainstem ischemia. Sudden onset of stroke attributable to internal carotid occlusion is inferred thrombotic.

2. Infarction, Etiology Unproven

Clinical Choice
No source for embolism, no bruit or prior TIA, normal CT scan beyond one week of stroke, or normal angiogram within two days of stroke, yet a clinical deficit that persisted beyond twenty-four hours even if eventual clinical resolution occurred.

3. Infarction Due to Embolism

CT Scan
Low density zone in the territory of a single cerebral surface branch of a major cerebral artery alone or in combination with infarctions in the distribution of branches of other divisions of major cerebral arteries. High density areas scattered in the infarct zone (“hemorrhagic infarction”) also suffice to make the diagnosis.

Angiogram
Cerebral surface branch occlusion unless the carotid artery is occluded or hemodynamically stenotic. Mere retrograde collateralization is not proof of embolism especially in occlusions involving the stems of the major cerebral arteries. Multiple sites of occlusion along the course of several branches of a major cerebral artery is sufficient for a diagnosis of embolism in the absence of spinal fluid changes or a clinical picture suggesting arteritis.

Clinical Choice
Clinical data suggestive of embolism include variables such as atrial fibrillation or flutter, bacterial endocarditis, previous myocardial infarction, right to left cardiac shunts, pulmonary vein thrombosis, ulcerative
plaque, atherostenosis or thrombosis of the internal carotid, sudden occlusion of a major cerebral artery stem without prior TIA's

4. Lacune

CT Scan
Focal deep site of infarction without involvement of the cerebral surface in the same parent major arterial territory at the same time.

Angiogram
Normal arterial anatomy of the major arteries of which the artery to the lacune is a branch. Local atherosclerosis at the site of origin of the artery to the lacune reclassifies the lacune as an atherosclerotic infarct.

Clinical Choice
Clinical characteristics conforming to a lacunar syndrome (pure motor hemiparesis, pure sensory stroke, ataxic hemiparesis, dysarthria clumsy-hand syndrome, pure sensory-motor stroke).

5. Parenchymal Hemorrhage

CT Scan
Scan performed within the first week: a deep seated mass of high-density with no associated appearance on contrast enhancement suggestive of aneurysm or arteriovenous malformation. Scans performed beyond one week: low or isodense appearance may exist (sometimes with "ring enhancement"). If the latter is the first scan, this diagnosis cannot be confirmed.

Angiogram
In combination with a positive CT scan, an angiogram showing no evidence of aneurysm or arteriovenous malformation.

Clinical Choice
Characteristic clinical symptoms showing smoothly evolving deficit over seconds to days without fluctuating signs.

6. Subarachnoid Hemorrhage

CT Scan
Focal or generalized collections of blood-density material in the basal cisterns and/or convexity subarachnoid space, with or without intraventricular blood, in combination with appropriate clinical syndromes.

Angiogram
No diagnostic features of subarachnoid hemorrhage itself but with appropriate clinical syndrome, demonstration of aneurysm or arteriovenous malformation.

Clinical Choice
Bloody spinal fluid with no sign of cause on CT or angiogram.

7. Transient Cerebral Ischemia (TIA)

Clinical Choice
Focal neurologic deficit of cerebral or retinal origin, lasting less than twenty-four hours. It must include either motor difficulty or visual or speech loss. If CT performed, must be normal in site attributable to symptoms or show only cavitation attributable to old infarction. Angiogram, if performed, may be normal in arteries overlying the symptomatic brain site but may show focal occlusion(s) in intracranial or extracranial arterial vasculature.

NOTE: Criteria for autopsy or surgery as the source for diagnosis are available on request.
The pilot Stroke Data Bank: definition, design, and data.
S C Kunitz, C R Gross, A Heyman, C S Kase, J P Mohr, T R Price and P A Wolf

Stroke. 1984;15:740-746
doi: 10.1161/01.STR.15.4.740

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/15/4/740

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/