Stroke
A Journal of Cerebral Circulation

MARCH-APRIL 1986
VOL. 17 NO. 2

Editorial

Stroke Data Banks

J.P. MOHR, M.D.

DATA BANKS ARE AN ATTRACTIVE MEANS of studying clinical aspects of stroke.[1] Population-based data banks have shown a declining incidence of stroke,[2] the natural history of infarcts in different vascular territories[3] and hemorrhage of different types,[4] the importance of atrial fibrillation,[5] and documented stroke rates by age, sex and race in communities world-wide.[6-10] Multi-institutional studies of low frequency conditions have clarified predictors for survival and morbidity in infarction[11, 12] and hemorrhage;[13] studied CT scanning in acute infarction;[14] shown that embolism is the most frequent cause of stroke[10, 15] and that many remain unexplained;[10, 16] correlated the lesions causing impaired eye movements,[17] hemiparesis,[18] and assessed the effects of metabolic variables.[19] Single institutions and communities have also found them a powerful tool for audits.[20, 21]

Despite all this effort, none of the data banks have yet achieved the ideal in which each case is studied with an intensity worthy of a single case report by a highly qualified principal investigator; where every case is part of a clearly defined population properly balanced for demographic factors; where each eligible case is entered, each is followed unfailingly at strict intervals, all laboratory tests are performed in a timely fashion; and with the whole effort yielding numbers so massive that each subset of cases can satisfy statistical criteria with contemptuous ease. This Moby Dick has proved a mighty adversary. The bent lances in his thick hide, including among them manuscripts in this issue,[26-28] testify to the many failed attempts at capture.

Data banks for stroke began as retrospective attempts to codify personal, institutional, or population material.[15, 22] Conflicting observations from what seem like similar studies have repeatedly shown how seemingly trivial variations in the population and study methodology can exert profound effects on the findings.[23] Efforts to check how each variable has been handled can tax the tenacity of the most battle-hardened reader of fine print. Although tedious, scrutiny of the methods sections is important because data from any source takes on a certain dignity when codified in tables in the results section. Compromise of some kind is common in attempts to approach the ideal data bank. Many multi-institutional studies have surveyed large populations for readily determined items such as demographic features, stroke incidence and the like, as shown in the studies in this issue.[25-28] Focussing on the fine points of individual clinical syndromes is another common gambit. A few population studies have attempted extensive detailed studies by limiting their efforts to communities of manageable size, but the subsets contain few cases. The literature is filled by those who avoided the problem entirely by reporting the exceptional case. The present issue contains reports from the multi-center community-based stroke program.[26-28] They provide a timely opportunity to review these points and to assess the general applicability of the results.

For a data bank to deserve its name, the data points should be like money in a commercial bank, well defined and of known value, dependable in the uses to which it is put and accessible to many users. Comparable data across centers can usually be assumed for readily defined items such as age, sex, date of hospital admission, length of stay, placement after discharge, death, etc, but questions that can be asked of such data are limited in scope. For the more contentious items, a common vocabulary must have been developed between the participants for each of the points in question. Review of the definitions provided by the authors for these items are a sign of the amount of effort to resolve the ambiguities of clinical features and diagnoses often glossed over in clinical practice. For final diagnoses, the extent of supporting laboratory data is a rough index of how much credence to give to the incidence data for any given subtype of stroke.

Although inter-observer reliability is easy to assume, something less has been demonstrated in the few studies when it has been tested, even among those who work closely together.[24, 25] (It is worth a moment's pause to ask how well many of the published randomized trials withstand such tests.) When data in a multicenter trial are to be pooled, evidence of comparability is needed.

For the studies to have much clinical interest, the expertise of the observer is as important as whether a tested vocabularly has been hammered out. The farther the source of the data gets from the investigators in direct contact with the patients, the more limited is the use to which the findings can be put. In descending order, the utility of clinical data declines when the investigator merely supervises others; personally reviews, or, worse, has others review charts created for another purpose by those not directly involved; merely has chart abstracts; or records diagnoses coded in rec-

From the Stroke Service, The New York Neurological Institute, New York, New York.

Address correspondence to: J.P Mohr, M.D., Stroke Service, The New York Neurological Institute, New York, New York.

ord rooms. These concerns may be less important when the data being collected are items of low priority.

When a multicenter effort is undertaken, it is assumed the cases share a common method of recruitment, demographic variables, extent of workup, diagnostic subtype, or followup. Even when the items are defined and the investigators qualified, pooled data become difficult to comprehend when some of it was collected retrospectively and some prospectively; when the time frames of collection do not match; when some of the populations are sharply defined and others not; when the team gathering the data differs substantially according to institution; when the source of the data varies from patient interview shortly after admission, to interviews after 30 days, to record reviews after discharge; and when the services available for diagnosis of stroke, and of stroke subtype vary from the outmoded to most advanced. Perhaps these ambiguities are of little relevance where the purposes of the study are simply to assess the impact of variations in stroke care and to establish geographic comparisons. Basic observations of this sort may prove robust enough to survive even striking variations in data collection methods.

The types of findings reported in this issue may give comfort to government agencies charged with keeping tabs on the overall stroke frequency, especially those agencies which make extensive use of nurses to implement the DRG programs that are currently tightening their grip on medical practice nationwide. For those whose ears ring from remarks from bored colleagues like, "Just tell me: is it a stroke or not?", the implications can be discouraging. Properly organized, multi-center data banks offer the hope of finding those elusive sub-groups who are susceptible to certain therapies. Their identification could lead to more sharply focussed clinical trials. In the process, some therapies could be resurrected which were set aside after disappointing results from clinical trials conducted on too broad a group. Large data banks could even improve current misestimates of sample size for a given trial. Data banks can provide the closest approximation to the answer sought for those many clinical problems whose incidence may defy the organization of a therapeutic trial. Assuming their data collection procedures are organized with these goals in mind.

J.P. Mohr, M.D.

Stroke Service
The New York Neurological Institute
New York City

References

1. Heyman A, Wilkinson WE, Pryor DB, et al: Data banks for the solution of clinical problems. In: F Plum, W Pulsinelli (Eds), Cerebrovascular Disorders, New York, Raven Press, pp 55–62, 1985
2. Garraway WM, Whisnant JP, Furlan AJ, et al: The declining incidence of stroke. New Eng J Med **300:** 449–452, 1979
3. Turney TM, Garraway WM, Whisnant JP: The natural history of hemispheric and brainstem infarction in Rochester, Minnesota. Stroke **15:** 790–794, 1984
4. Sacco RL, Wolf PA, Bharucha NE, et al: Subarachnoid and intracerebral hemorrhage. Neurology (Cleveland) **34:** 847–854, 1984
5. Wolf PA, Dawber TR, Thomas HE, et al: Epidemiologic assessment of chronic atrial fibrillation and risk of stroke: the Framingham Study. Neurology (Minneap.) **28:** 973–977, 1978
6. Eckstrom PT, Brand ER, Edlavitch SA: Epidemiology of stroke in a rural area: second year of the mid-Missouri survey. Public Health Reports **84:** 878–882, 1969
7. Zupping R, Roose M: Epidemiology of cerebrovascular disease in Tartu, Estonia, USSR, 1970 through 1973. Stroke **7:** 187–190, 1976
8. Wolf PA, Kannel WB, Dawber TR, et al: Prospective investigations: The Framingham study and the epidemiology of stroke. Adv Neurol **19:** 107–120, 1978
9. Mettinger LK, Söderström CE, Allander E: Epidemiology of acute cerebrovascular disease before the age of 55 in the Stockholm County 1973–1977: I. Stroke **15:** 795–801, 1984
10. Gross CR, Kase CS, Mohr JP, Cunningham SC: Stroke in south Alabama: incidence and diagnostic features. Stroke **15:** 249–255, 1984.
11. Levy DE, Scherer PB, Lapinski RH, et al: Predicting recovery from acute ischemic stroke using multiple clinical variables. In: F Plum, W Pulsinelli (Eds), Cerebrovascular Disorders, New York, Raven Press, pp 69–77, 1985
12. Price TR, Dambrosia J, Wolf PA, et al: Early determination of prognosis for 30-day and 1-year survival in ischemic infarction: The NINCDS Stroke Data Bank. Neurology **35** [suppl]: 212, 1985
13. Price TR, Dambrosia J, Wolf PA, et al: Predictors of 30-day mortality in intracerebral hemorrhage: the NINCDS stroke data bank. Ann Neurol **18:** 107, 1985
14. Tatemichi TK, Mohr JP, Rubinstein LV, et al: CT findings and clinical course in acute stroke: the NINCDS pilot stroke data bank. Stroke **16:** 138, 1985
15. Mohr JP, Caplan LR, Melski JW, et al: The Harvard cooperative stroke registry: a prospective registry of cases hospitalized with stroke. Neurology **28:** 754–762, 1978
16. Kunitz S, Gross CR, Heyman A, et al: The pilot stroke data bank: definition, design, data. Stroke **15:** 740–746, 1984
17. Mohr JP, Rubinstein LV, Kase CS, et al: Gaze palsy in hemispheral stroke: the NINCDS Stroke Data Bank. Neurology **34**[Suppl]: 199, 1984
18. Mohr JP, Rubinstein LV, Kase CS, et al: Hemiparesis profiles in stroke: the NINCDS Stroke Data Bank. Ann Neurol **16:** 156, 1984
19. Mohr JP, Rubenstein L, Edelstein SZ, et al: Approaches to pathophysiology of stroke through the NINCDS Data Bank. In: F Plum, W Pulsinelli (Eds), Cerebrovascular Disorders, New York, Raven Press, pp 63–68, 1985
20. Brott T, Thalinger K The practice of carotid endarterectomy in a large metropolitan area. Stroke **15:** 950–955, 1984
21. Whisnant JP, Sandok BA, Sundet Jr TM: Carotid endarterectomy for unilateral carotid system transient cerebral ischemia. Mayo Clin Proc **58:** 171–175, 1983
22. Steinberg FU: A pilot stroke registry in a general hospital. Gerontologist **12:** 148–150, 1972
23. Kurtzke JF: Epidemiology. In HJM Barnett, JP Mohr, BM Stein, FM Yatsu (Eds), Stroke, New York, Churchill-Livingstone, pp 3–18, 1986
24. Tomasello F, Mariani F, Fieschi C, et al: Assessment of inter-observer differences in the Italian multicenter study of reversible cerebral ischemia. Stroke **13:** 32–35, 1982
25. Shinar D, Gross CR, Mohr JP, et al: Inter-observer variability in the assessment of neurologic history and examination in the Stroke Data Bank. Arch Neurol **42:** 557–565, 1985
26. Yatsu FM, Becker C, McLeroy KR, Coull B, Feibel J, Howard G, Toole JF, Walker MD: Community hospital-based stroke programs: North Carolina, Oregon, and New York. I: Goals, objectives, and data collection procedures. Stroke **17:** 276–284, 1986
27. Becker C, Howard G, McLeroy KR, Yatsu FM, Tooles JF, Coull B, Feibel J, Walker MD: Community hospital-based stroke programs: North Carolina, Oregon, and New York. II: Description of study population. Stroke **17:** , 1986
28. Howard G, Walker MD, Becker C, Coull B, Fiebel J, McLeroy K, Toole JF, Yatsu FM: Community hospital-based programs: North Carolina, Oregon, and New York. III: Factors influencing survival after stroke. Proportional hazards analysis of 4219 patients. Stroke **17:** 1986