Stroke Data Banks

To the Editor:

Dr. Mohr is to be complimented for a thoughtful, balanced discussion of stroke data banks.[1] While there are pros and cons to this issue, in our view, the advisability of establishing stroke data banks remains questionable, given the predictably high cost of such endeavors and the predictably low-level, descriptive research that such enterprises can support.

The editorial acknowledged some of the problems of quality control in large data banks, where interobserver variability and differing standards between participating centers may play substantial roles. Transcription errors, omissions, and the like also account for serious quality control problems, particularly in large databases that are likely to be computerized.[2]

The desire to hoard data is a natural instinct for those of us involved in clinical research. However, unlike money in the bank, data in a data bank do not gain interest merely because they have been saved. On the contrary, data that are routinely collected and salted away for some ill-defined future purpose are likely to harbor many undetected errors. When discovered, many of these errors will be uncorrectable because nobody has looked at the data until long after the patients involved have disappeared from the scene.

As Dr. Mohr noted, the referral patterns, practice preferences, and catchment area of an institution or group of institutions that sponsor disease-specific databases may yield data that are difficult to apply in other settings. When it is possible to maintain a database from a well-defined population (e.g., a state tumor registry or mortality statistics), the quality limitations of the stored information are appreciated by knowledgeable investigators. Such records are rarely, if ever, comprehensive; for purposes of clinical inference, subjects may be identified from such registries, but individual patients or original documents are often located for collection of additional data specific to the studies being undertaken. No database can be all things to all researchers. It often transpires that a critical variable turns out to be the one item that was not collected in the "comprehensive" database.

Many quality control problems in stored databases can be overcome with sufficient money and effort, but even when the data collected are complete and accurate, the questions that a data bank can answer may be severely limited. Descriptive studies are widely regarded as the weakest evidence of causation, for example. Unless the database provides information on a control or comparison group — individuals without stroke, or patients receiving a different treatment from the one under investigation, for instance — it cannot answer fundamental clinical questions about the relative merits of different treatments in several groups of patients.

In the absence of a formal control group, to interpret descriptive data at all requires implicit comparisons with other patient populations, other centers, or past experience, with all the attendant perils of unknown comparability, biases, and confounding variables undermining the validity of comparisons between patients in the database and the reference population.

But the greatest danger in keeping expensive data banks that have been assembled in the absence of previously defined hypotheses is the almost irresistable temptation to delve into these presumed "gold mines," especially when data have been stored in computer systems where they can be perused with relative ease. Statisticians know that "if you dredge data sufficiently deep and sufficiently often, you will find something odd.[3]" Given enough variables, sampling error virtually ensures that some of them will lie outside the expected usual range or achieve statistical significance in comparison with data from a case-control group, because of chance alone.[4]

Dr. Mohr is correct in stating that a case series and other descriptive studies have been useful for hypothesis generation and as a stimulus for other, more powerful investigations.[5] If properly set up, descriptive studies make a strong contribution in studies of prognosis and comparisons of therapeutic maneuvers. Each of these activities assumes the application of fundamental principles of scientific investigation. Sifting through massive data banks in quest of some unexpected finding is no substitute for these principles.

Neil Harding McAlister, M.D., M.Sc., Ph.D.
Assistant Professor
Mary Chipman, B.Sc., M.A.
Associate Professor
Department of Preventive Medicine and Biostatistics
Faculty of Medicine, University of Toronto
Toronto, Ontario
M5S 1A8, Canada

References

1. Mohr JP: Editorial: Stroke data banks. *Stroke* 1986;17:171–172
2. Covvey HD, Craven NH, McAlister NH: *Computers in Health Care*, vol 1. *Concepts and Issues in Health Care Computing*. St Louis, CV Mosby Co, 1985, p 100
3. Armitage P: Importance of prognostic factors in the analysis of data from clinical trials. *Controlled Clin Trials* 1981;1:347–353
4. Fletcher RH, Fletcher SW, Wagner ED: *Clinical Epidemiology: The Essentials*. Baltimore, Williams & Wilkins, 1982, p 165
5. Sackett DL, Haynes RB, Tugwell P: *Clinical Epidemiology: A Basic Science for Clinical Medicine*. Boston, Little Brown & Co, 1985, p 229

The following letter is sent in response:

To the Editor:

The undersigned thanks the authors for their initial compliment. All the same, it is disappointing to find that they, and perhaps others, read so little between the lines that so heavy an amplification seemed called for.

The undersigned believes their stern admonitions have already been long digested by experienced clinical investigators. It seems doubtful that the clumsily conceived and crudely executed "data banks" decried by the authors actually exist, although any such are clearly a waste of time and effort. Do any investigators "hoard data" . . . "routinely collected and salted away for some ill-defined future purpose," casually assume data is uniformly useful, assume data from any one study is "comprehensive," require being reminded of the need for control groups, or regard data that is retrievable as a "gold mine"? (The leftover data in most projects is more lead in the proverbial balloon than gold, to use the authors' metal metaphor.) Few clinical investigators will be found recoiling in horror like some Draculas before the cross of "data dredging" thrust at them by the tight-lipped authors. Statisticians and epidemiologists are more likely to have the leisure and expertise to undertake any data dredging that is done; after all, it is hard enough to keep easily distracted clinical investigators adhering to a data collection protocol, let alone churning through computerized data bases in search of some unpredicted correlations, and fewer still would see any point in such activity in the absence of a previously posed hypothesis.

Can there be any objection to the time-honored approach of organizing a study to pursue some one or more research hypotheses, carrying out the study with meticulous attention to these details, reporting its results, and being alert to observations during the study that may generate new hypotheses subject to yet another study? And if some of the subsidiary findings prove of interest, is it not useful to call to attention these preliminary observations, hoping others may find them useful in the creation of new hypotheses? Do such subsidiary reports differ in any way from "case series and other descriptive studies" approved by the authors?

Should the projected need for a large sample size bring investigators from several institutions together and should their joining together permit them to assemble a project that can pursue many hypotheses simultaneously, using a carefully validated common data collection protocol, and should the effort lead to a large data base, in what way is the multicenter–multihypothesis project different from the single investigator–single hypothesis in anything but size and complexity? The data collection will create a data base simultaneously to test the hypotheses, albeit at different rates, given the differences in sample size required.

And the same "case series" observations can be spun off, not later or through "data dredging," but at the same time, to create new hypotheses subject to subsequent data collection. And as new hypotheses are generated, the research consortium is in a position to collect the relevant data with few of the usual delays in forming a research team.

By contrast, should the authors' proposals for research be adopted in their extreme form, each study presumably would be pursued using only one hypothesis, and any data collected not directly bearing on the hypothesis would be rigorously set aside. Research of this type, while agreeably pristine, would be expensive indeed and would greatly delay the promulgation of new research hypotheses. Many of the consortia who have labored long and hard to generate the most productivity for the research dollar would be chagrined to learn that subsidiary efforts may be of so little interest.

Raising the issue of expense as a reason to avoid future projects may prove counterproductive. As computers fall in price, rise in power, and get more friendly, many investigators may be tempted into the field. They will find sophisticated programs available, their displays of frequencies and time-based curves beguiling to the unwary along exactly the lines that worry the authors.

If the hostility toward the term "data bank" as reflected in the authors' letter is widespread, existing research groups are well-advised to change their names lest they suffer the "Sambo's effect." That restaurant chain was said to have been forced under by minority groups protesting the name, misconstruing what was the portmanteau of the two owners' first names for the (minority member) hero of the children's story. Thanks to the authors for calling to general attention the ambiguities inherent in the term "data bank." They may have spared many computer-based clinical investigators unwanted future agonies.

J.P. Mohr, M.D.
Sciarra Professor of Clinical Neurology
College of Physicians and Surgeons of Columbia University
Neurological Institute
710 West 168th Street
New York, NY 10032

Dichloroacetate After Incomplete Ischemia

To the Editor:

We would like to comment on the article by Colohan and coworkers[1] in the May–June 1986 issue of *Stroke* regarding the effect of dichloroacetate (DCA) after incomplete ischemia in rats. Using the same model, our evidence that pretreatment with DCA does not decrease the amount of lactic acid build-up in the brain during ischemia[2] is in agreement with the findings of these authors. However, additional data of ours shows that treatment with DCA does result in lower lactates 30 minutes after the termination of ischemia.[2–5]

We agree with these authors that even though pyruvate dehydrogenase enzyme complex (PDHC) in the brain might be activated by DCA during ischemia, the lack of oxygen is probably what limits the entry of pyruvate (lactate) into the citric acid cycle.[1] Having made this assumption, we measured cerebral cortical lactic acid levels 30 minutes after reperfusion in rats that had been treated with DCA 15 minutes prior to, immediately after, or 15 minutes after ischemia. Our results showed that cerebral lactate is near control levels 30 minutes after reperfusion in fasted rats pretreated with 25 mg/kg DCA.[3] In untreated ischemic rats this resolution takes at least 60 minutes after the start of reperfusion.[6] When treated immediately or 15 minutes after ischemia, there likewise is a significantly faster amelioration of cerebral hyperlactatemia in DCA-treated ischemic rats when compared with untreated rats.[4] Since these effects were achieved with a small dose of DCA (25 mg/kg) that was not effective in resolving systemic acidosis,[7] we also agree with these authors[1] and Evans,[8] that the control of PDHC activity may be different in the brain than in other tissues.

Since most patients with cerebral ischemia will be fed and since high blood glucose correlates with poor physiological and neurological outcome from cerebral ischemia,[9–11] we examined the effect of postischemic treatment with DCA in fed rats. Untreated rats exhibited mean cerebral lactates of 23 μM/g 30 minutes after ischemia. In contrast, the mean lactate level in DCA-treated rats 30 minutes after ischemia was 13 μM/g,[5] significantly less than 18 μM/g that has been shown by other investigators to promote irreversible cell damage.[6,12] Since brain ischemia commonly will affect nonfasted patients and since treatment will occur after, not prior to, ischemia, we believe these results have important clinical significance. We therefore urge the continued investigation of the use of DCA for the treatment of cerebral hyperlactatemia in ischemic events.

Sincerely,

Ruth V.W. Dimlich, Ph.D.
William G. Barsan, M.D.
University of Cincinnati
Cincinnati, Ohio

References

1. Colohan ART, Welsh FA, Miller ED, Kassel NF: The effect of dichloroacetate on brain lactate levels following incomplete ischemia in the hyperglycemic rat. *Stroke* 1986;17:525–528
2. Biros MH, Dimlich RVW: Brain lactate dynamics during partial global ischemia and reperfusion in the rat. *Ann Emerg Med* 1986;15:636
3. Dimlich RVW, Biros MH, Barsan WG: Treatment of lactic acidosis following partial global ischemia in the rat. *J Cereb Blood Flow Metab* 1985;5:S227–S238
4. Biros MH, Dimlich RVW, Barsan WG: Post insult treatment of ischemia-induced cerebral lactic acidosis in the rat. *Ann Emerg Med* 1986;15:397–404
5. Kaplan J, Dimlich RVW, Biros MH: Dichloroacetate treatment of cerebral ischemia lactic acidosis in the fed rat. *Ann Emerg Med* (in press)
6. Rehncrona S, Rosen I, Siesjö BK: Brain lactic acidosis and ischemic cell damage: 1. Biochemistry and neurophysiology. *J Cereb Blood Flow Metab* 1981;1:297–311
7. Dimlich RVW, Biros MH, Widman D, Kaplan J. Comparison of sodium bicarbonate with dichloroacetate treatment of lactic acidosis in rats. *Ann Emerg Med* 1986;15:631–632
8. Evans OB: Effects of dichloroacetate on brain tissue pyruvate dehydrogenase. *J Neurochem* 1983;41:1052–1056
9. Siemkowitz E, Hansen AJ: Clinical restitution following cerebral ischemia in hypo-, normo- and hyperglycemic rats. *Acta Neurol Scand* 1978;58:1–8
10. Myers RE: A unitary theory of causation of anoxic and hypoxic brain pathology, in Fahn S, David JN, Rowland LP (eds): *Cerebral Hypoxia and its Consequences. Advances in Neurology,* vol 26. New York, Raven Press, 1979, pp 195–213
11. Welsh FA, Ginsberg MD, Rieder W, Budd WW: Deleterious effect of glucose pretreatment on recovery from diffuse cerebral ischemia in the cat. II. Regional metabolite levels. *Stroke* 1980;11:355–363
12. Kalimo H, Rehncrona S, Söderfeldt B, Siesjö BK: Brain lactic acidosis and ischemic cell damage. 2. Histopathology. *J Cereb Blood Flow Metab* 1981;1:313–327

The following letter is sent in response.

To the Editor:

Dichloroacetate is a theoretically attractive agent for treatment of cerebral ischemia. We were quite disappointed when the results of our studies failed to suggest a beneficial action. However, the additional studies cited by Drs. Dimlich and Barsan indicate that dichloroacetate may indeed be useful in the management of ischemic stroke. Our enthusiasm has been rekindled, and we agree completely that further studies should be conducted.

Very truly yours,

Neal F. Kassell, M.D.
Professor and Vice Chairman
Austin R.T. Colohan, M.D.
Department of Neurosurgery
University of Virginia
School of Medicine
Charlottesville, Virginia