Abstract

A properly implemented patient registry can greatly facilitate clinical research, serving to establish or negate commonly accepted manifestations of disease which are based on anecdotal evidence and experience, as well as uncover previously unrecognized correlations. It can provide lists of patients with any desired profile for use in new clinical studies. With the advent of inexpensive microcomputers, many clinicians now can effectively use patient registries. We have constructed a program to accumulate a registry of stroke patients. It stores 328 items of information about each patient, including history, physical findings, laboratory and radiological data and pathological findings. A search program will allow retrieval using Boolean combinatorial search on any of the keys. The program is implemented on an Apple II computer. At present, more than 500 patients have been entered into the registry.

Rationale

In medicine, as in other fields of human endeavor, we have the tendency to assume that what we have learned is true and there is an illusion that the knowledge is (almost) complete. Much of traditional medical wisdom is built upon years of careful observations by experienced clinicians. Those who have contributed most often have been burdened with the necessity to recall well the characteristics of a set of similar cases seen in the past and accumulate their “data” over a period of years. It requires a good memory as well as the perceptiveness to recognize the uniqueness of a previously undescribed syndrome in patients being seen over a period of time. It often comes as a surprise to experienced clinicians when they read the description of a new syndrome and finding in their own clinic such patients, they wonder why they had not made the connection.

Even in well-known clinical entities, much of the information regarding the frequency of the appearance of various manifestations, and the order in which they present is anecdotal. Studies containing large numbers of patients are rare except in the most common diseases. Many of the reasons for the paucity of data are related to the inconvenience of recording, storing, and searching the data.

The traditional method of chart review requires that either the physician remember which charts he is interested in (subject to fallability of memory), look at all the charts (impractical), or have sorted the charts in advance (know already what he is looking for). Although much of current clinical knowledge has been uncovered in this manner, it is so tedious that it is not commonly undertaken, and rarely outside academe.

The role of carefully designed patient registries has become recognized in recent years and several dealing with cerebrovascular disease have been implemented on mainframe computers [1,2,3]. The NIH is currently compiling a multi-center stroke registry and other registries dealing with rheumatic and cardiac disorders exist. If care is taken to record into the database systematic and reliable information about a large number of patients with a category of disease (for example, cerebrovascular disease), the data may be analyzed at a later time with regard to any of the clinical parameters entered into the registry. This avoids the problem of retrospective searching of those patients as remembered by the physician with its attendant bias, as well as providing a quick and convenient way of storing and searching the data.

Convenient and useful as computerized patient registries are, few clinicians have been able to take advantage of them because of the lack of funds for a large computer and the support staff necessary to run it. With the advent of microcomputers, this problem is obviated. For a few thousand dollars powerful computers with enough mass storage to maintain patient databases of significant size can be purchased, bringing the benefits of patient registries to any who wish to undertake this type of clinical research. It is our feeling that much valuable clinical knowledge could be discovered should the use of computerized databases coupled with careful collection of clinical data become widespread.

Implementation

The implementer of a patient registry may choose to use a database management system (DBMS) sometimes called a “universal database manager.” The programs are not written with any specific data in mind, but with the idea of allowing the user to specify the nature of the data he wishes to collect within as broad limits as possible. By nature this involves compromises of speed, mass storage efficiency, restraints the manner of data entry, and limits the sizes of the data base items to an extent that varies according to the cleverness of the DBMS implementation. It has the advantage of simplifying the programming task [4].

The Michael Reese Hospital Stroke Registry

A Microcomputer-implemented Data Base

Gordon Banks, Ph.D., M.D.
University of Pittsburgh
Decision Systems Laboratory and Department of Neurology
1300 Scaife Hall
Pittsburgh, PA 15261

Louis R. Caplan, M.D. and Daniel B. Hirt, M.D.
Department of Neurology, Michael Reese Hospital
Chicago, IL 60616
Another alternative is to design and program a database in a higher level language, such as Pascal, specifically tailored to the desired purpose, in this instance, a patient registry. This method has the advantage of optimization of efficiency of operation, and customization of the data base to fit the data, but requires programming knowledge, time, and effort.

The Michael Reese Hospital Stroke Registry was established to gather and correlate clinical and pathologic information regarding patients with stroke. Data on 328 different items regarding demographics, history, physical examination, diagnostic studies, treatment, and neuropathology were included. It is our hope to use this information to establish the value of various clinical data for diagnosis, localization, and prognosis of stroke. Currently over 500 patients have been entered into the registry.

The programs are written in UCSD Pascal and currently run on an Apple II microcomputer. They are also portable to any computer running UCSD Pascal (IBM-PC, or any CP/M computer). Several programs are needed to use and maintain the registry:

**EDITOR:** this program maintains the database. It is used to enter data, edit data already entered, and obtain reports on patients in the database. It maintains the file REGISTRY PACKED RECORDS (a Pascal data type [6]) containing all the information concerning one patient. The data is so compressed as to allow each record to occupy only about 512 bytes. Thus a single density 8-inch floppy disk can contain data for more than 500 patients.

The editor is frame-oriented. The user selects the registry number of the patient he wishes to edit (or simply tells the editor to add a new patient). The first frame displays on the CRT screen the demographic data (name, date of birth, hospital number, etc.). Subsequent frames present each of the 328 items of information regarding history, signs, tests and procedures, and pathologic data on the patient. They may be cycled through sequentially or the user may jump to any desired item (frame). Each frame presents one item (see figure 1) and a list of choices. The user must enter the number corresponding to the appropriate choice. Some frames contain numerical data which is entered directly (e.g. diastolic blood pressure).

Forcing the user to enter one of a fixed number of choices limits his flexibility, but this is necessary if valid analysis is to be done on the data later. If complete freedom to fill in a word or phrase was given, how would one decide on the equivalence of similar words or phrases? Using fixed choices requires thoughtful preparation of the list of choices to cover all useful eventualities. Identical lists of choices should be used in comparing two different methods of assessing the same phenomena. Identical lists of choices should be used in asking the clinician to enter the results of the autopsy, thus assuring that any correlation can be detected and not obscured because of requirement for interpretation of what is equivalent.

**REGISTRY** which is a sequential file of records present each of the 328 items of information to establish the value of various clinical items and the strings associated with the multiple choice items. Figure 1 shows a segment of the text file for the diastolic blood pressure.

When EDITOR is executed, it begins by reading a text file containing the prompting strings for the data items. Figure 2 shows a segment of the text file for the stroke registry. The first line of an item is the "title" of the data item, followed by a line containing numbers indicating how many possible choices there are (for a multiple choice item), and other control parameters (for example to allow the program to skip over a certain number of items should a data item take on a predefined value or if a null entry is made). The titles of the items and the strings associated with the multiple choices are read into a set of linked lists at initialization. The disadvantage of this scheme is the time required for initialization. This is counterbalanced by the advantage of being able to use a text editor to change the items in the text file, thus modifying the registry with ease. In fact, the same programs could be used with many different registries, simply by using a different text file and changing a constant definition to specify the number of data items in the source code for the programs, then recompiling. Thus, setting up a dementia
registry, for example, merely requires writing a set of prompting strings for the new data items, substituting them for the stroke parameters, changing the constant 328 (number of items) in the Pascal programs to match the number of items in the dementia registry, and the programs are ready to run.

The EDITOR is also able to modify any record previously entered to add data that subsequently becomes available (e.g. autopsy) or to correct errors. It also can print a report on the data of any patient.

SEARCH: this program searches the data base. Two search methods are allowed. The first method allows sub-registries to be formed containing all the data on patients meeting given criteria. For example, suppose we want a registry composed of all black males who had TIAs. The search condition can be specified as the Boolean ((RACE = BLACK) AND (SEX = MALE) AND (TIA = YES)). Any Boolean expression formed from combinations of data items can be used as a selection condition. This feature is be useful in selecting patient populations having certain characteristics for clinical trials and other studies. Once selected, these sub-registries can be operated on by the same programs and in the same manner as the master registry.

The second search method provides statistical information on all patients in the registry meeting certain criteria. The search program allows the criteria to be constructed from any legal Boolean expression about any number of data items in the registry. For example, one could choose ((RACE = BLACK) AND (DIABETES = INSULIN-REQUIRING) OR (SYSTOLIC- BLOOD-PRESSURE > 150)) as one of the search conditions, making the condition as complicated as necessary. Figure 3 depicts the output of a search of 207 patients. The criteria given were sex and race vs. the risk factors of hypertension and diabetes. The results of the searches are plotted in tabular form as shown in figure 3. Interpretation of the table is as follows: SUMS show the total number of patients fitting each category. There were 95 females (46% of 207), 110 males (53%), 147 blacks (71%), and 57 whites (28%). (Obviously, in our sample registry, the sex question was not filled in for 2 patients, and 3 were neither black nor white, or the race question was not filled in. The SEARCH program could be used to find the patients with incomplete data, and the EDITOR could be used to correct the data.) Hypertensives made up 157 or 76%, and diabetics 42 or 20%. At the intersections of the rows and columns, the sets containing patients filling both criteria are reported. For instance, there were 78 hypertensive females, or 83% of all the females, and 50% of all the hypertensives were female. 38% of all patients in the stroke registry were hypertensive females.

Using these two search methods, most information that one would desire from the registry can be retrieved. Both searches are accomplished by reading the data file sequentially, and comparing the patient data with the search condition, then selecting or tabulating based on the results of the comparison. A search of 500 patients typically takes less than 5 minutes.

FORM: uses the text files mentioned previously to prepare and print a blank form to be used by the physician at the bedside to fill in the data, which is later entered into the computer by a clerk.

Other programs have been written to provide a alphabetic list of patients in the registry, as well as providing reports in a form suitable for mailing to referring primary physicians.

Costs.

Development time was about two months of spare-time programming by one neurologist and a few days of thoughtful consideration by two other neurologists to decide on appropriate data items. The equipment included an Apple II, two 8-inch floppy disk drives and two 5-inch floppy disk drives with controllers, a monitor, and a dot-matrix printer. The total cost to duplicate the system would be less than $4000 at current prices.

Results.

The first 500 stroke patients are currently being analyzed and results are being reported in the neurologic literature [6,7].

Conclusions.

Computerized databases can be a valuable aid to furthering knowledge of clinical disease states. It is no longer necessary to forgo their convenience and power due to the necessity for high capital outlay. Anyone seriously interested in maintaining clinical databases can do so with minimal expense on the current generation of microcomputers. Further technical developments in computer and peripheral hardware currently taking place will serve to increase the power and storage capability as well and decrease the price of such systems.
References


