COMPUTERS AND BIOMEDICAL RESEARCH 3, 464–476 (1970)

High-Density Medical Data Management by Computer

HOMER R. WARNER * AND JOHN D. MORGAN

University of Utah, Latter-day Saints Hospital, Salt Lake City, Utah 84103

Received June 8, 1970

The duplication of standard medical record format and content is not the intent of a computer-based medical record. Seven characteristics designed to improve patient care, efficiency of data entry, and research capability in a computerized system are enumerated. These characteristics are further amplified and demonstrated by reference to a system in use by the Biographics Department of the University of Utah. Portions of the system dealt with specifically are the file structure, patient screening, ICU monitoring, diagnosis entry and data retrieval.

The Biophysics Department of the University of Utah first became involved in patient care and clinical applications of the computer 4 years ago with the inception of a computerized heart catheterization procedure. Since that time more and more clinical departments within the hospital have begun to afford themselves the advantages of accuracy, calculation, time saving, and rapid data retrieval offered by the computer. Even though the primary goal was development of computerized clinical procedures and research on the clinical data (not the establishment of a complete patient record), it was soon discovered that sufficient data were being collected on each patient to form the nucleus of a respectable patient record.

The traditional medical record, however, is somewhat of an enigma. It contains a plethora of information about the patient's hospital stay (history and physical, progress notes and nurses notes, orders, reports of actions and findings, etc.) necessary to document the patient care received. This abundance of information, though necessary to provide a complete document, is at the same time a hinderance to rapid retrieval of vital and timely physiologic data needed in patient care (*1*). Further research and education using the standard medical record as a data base is a hopelessly time consuming task, as anyone who has spent hours pouring through these awkward files can readily testify.

A computer-based medical record *can* satisfy the immediate retrieval needs of a physician for patient care and the longer term retrieval needs of the researcher, if the data content is carefully chosen, obtained, and stored using

---

* Supported in part by Public Health Grant RR00012.

methods well-suited to computerization. To justify the use of such a computer system in the management of a patient record one or more of the following must characterize the computer-based system.

First, entry of some portion of the data into the system must be automatic, semiautomatic, or performed by nonprofessionals. Since some types of data (particularly narrative) are, at the present state of the art of man–machine interfaces, entered into the computer less efficiently than their corresponding manual methods, the automatic entry of the more quantitative data must offset this loss to provide a net gain in efficiency for the computer system (*2*).

Second, the data storage must be sufficiently structured to permit unequivocal definition of the data, and at the same time be sufficiently flexible to permit additions and modifications of the data structure to meet changing needs. There must also be enough structure to the data to allow rapid, efficient retrieval preferably with generalized retrieval programs.

Third, the amount of data in the file must be sufficiently small to be economically stored in machine accessible form and yet must contain enough redundancy to establish confidence in the deductions to be made from these data. For example, consider the monitoring of the electrocardiogram and the pressure pulse in an intensive care unit. Heart rate is a redundant variable available from each signal, providing a check on both the equipment and procedure.

Fourth, data necessary for the decisions made in the care of the patient must be immediately available for organization in summary form, while past details of only medical-legal interest are allocated to an inexpensively maintained archive (handwritten notes, etc.) That is, data must be structured and stored according to urgency of the need to access it.

Fifth, the data should be processed before storing to extract signal from noise and to reduce the data to a clinically significant form. For example, a variable such as forced vital capacity should be stored as a percent of predicted value based on parameters such as height, weight, sex and age.

Sixth, the data must be immediately available in readily interpretable graphic and narrative displays. These displays may be on hardcopy forms, or on display screens of remote terminals.

Seventh, computerized quality control should be available at each entry port to the system for narrative as well as numeric information (*3, 4*).

Examples of the preceding seven points will be illustrated throughout the remainder of this paper by reference to the computer-based medical records system currently operated by the Biophysics Department of the University of Utah.

The central processors of this system are a CDC 3200 and a CDC 3300. Both computers have access to six magnetic disks, each having a capacity of 2 million 24-bit words. Three tape units, a card reader, a high speed printer, and two A-D converters with associated communication links complete the system configuration (*5*). A PDP-8S is used to provide communication with remote teletypes. In the normal mode of operation research activities of the department are carried out on the CDC 3200 independently of the clinical work being done on the 3300, but communication between the central processors is provided through common access to the magnetic disks. In the event of failure of either central processor the other can act as backup.

This facility is located in the Latter-day Saints Hospital in Salt Lake City, a general hospital of 564 bed capacity, handling approximately 24,000 admissions annually. Figure 1 illustrates the types of clinical data currently comprising the computerized patient record. While the patients are in the hospital their records are stored on a magnetic disk to facilitate rapid access to their record, but when they are discharged their records are transferred to magnetic tape for long term storage.

File Structure

To retrieve or add to the medical data on a given patient, the following file scheme is used: A library containing current inpatient hospital numbers is searched for the given patient number (Fig. 2). The position of the hospital number in the list serves as a pointer to a block of identifying information on the disk. This ID information consists of hospital number, name, sex, age, admit date, doctor number, etc., as well as pointers to the locations where the patient's data pages are stored. These pointers are accompanied by the date and time the data pages were initiated to serve as an aid to chronological recovery and storage of the data. There is no limit to the number of data pages a patient

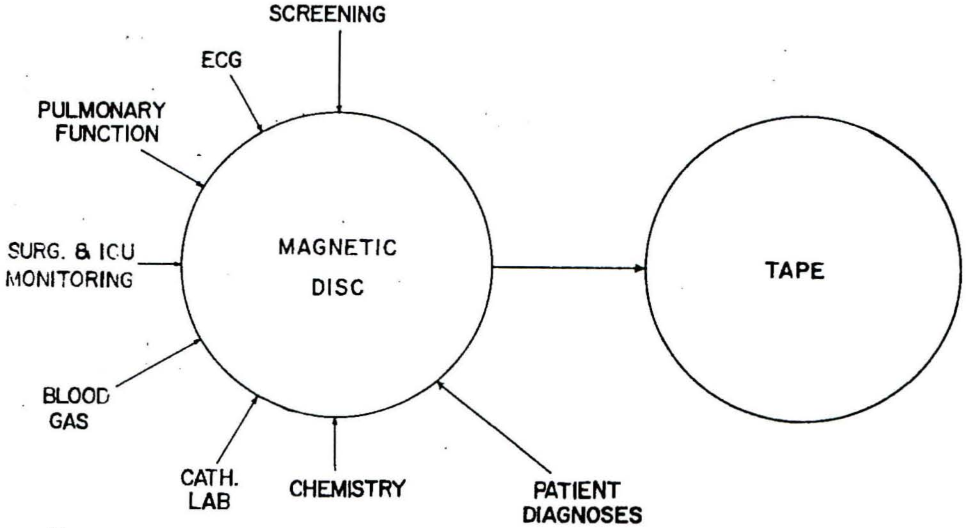


FIG. 1. Schematic diagram of computer based records system illustrating types of medical data being recorded. While patient is in hospital, records are stored on magnetic disk for rapid retrieval; but when patient is discharged, they are transferred to tape for long-term storage.

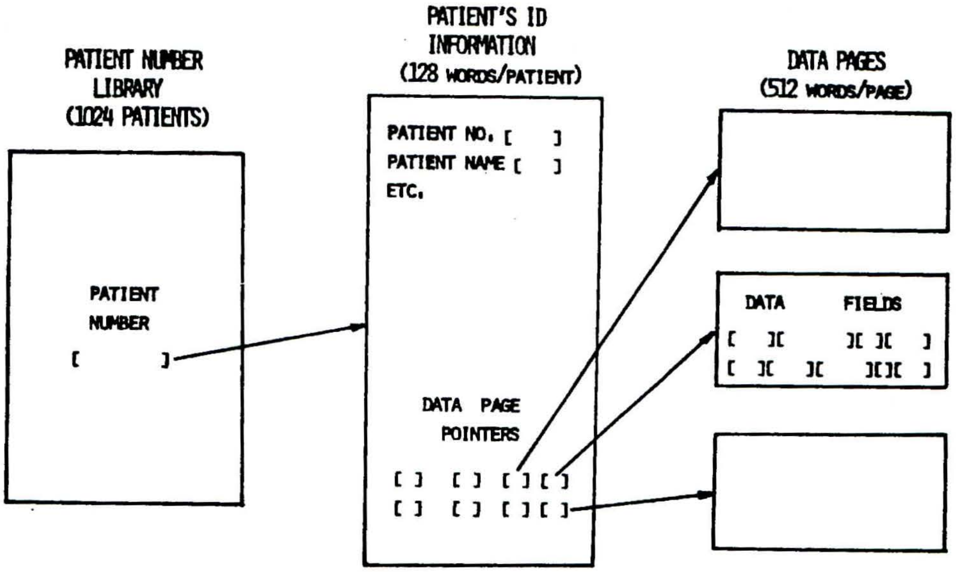


FIG. 2. Data storage schema. From the location of the patient's number in the library an algorithm determines the sector address of the patient's ID information. In addition to ID information stored here is a series of sector addresses pointing to the patient's data pages.

may be assigned (the average is 2, but some patients have occupied 90 or more). Each data page is 512 words long and is composed of a variable number of individual data entries formated as in Fig. 3.

A unique 18-bit field code is assigned to each procedure, test comment or diagnosis to be entered into the computerized record. For example, 11050 is the code for 12-channel chemistries, 11055 is pulmonary function, and 11057 indicates a diagnosis follows. This 18-bit code is followed in the same word by a 6-bit count (filling the 24 bits of the word) of the number of actual data words contained in this field. The date and time to be assigned to the subsequent data words is recorded in the next word.

Following this are the actual data words. These may be continuous variables, such as blood glucose values; binary variables, such as answers to patient history questions; or they may be codes signifying narrative phrases, as in diag-

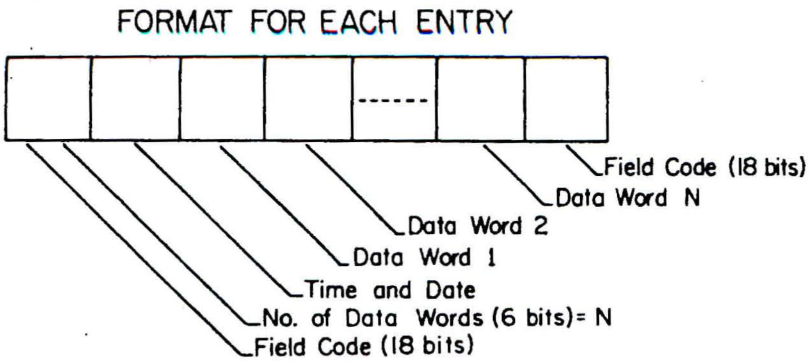


FIG. 3. Data field format.

noses. The word containing the field code and count of data words is again repeated at the end of the data string to aid in backward searching of the files.

Patient Screening

One portal of entry to the patient record is through a patient screening laboratory. The patient's first contact is with a receptionist who has a terminal to the computer through which she can enter such information as height, weight, age, sex, and social security number of the patient. A self administered patient history is then obtained by giving the patient a portable device about the size of a large book into which can be inserted a punched card whose holes are preperforated.[1] The patient's identification number is punched on the card. Overlying the card is a set of plastic covered pages on which are printed questions which can be answered "yes" or "no." There is a preperforated position on the card corresponding to each question, and the patient punches a hole opposite each question if his answer to that question is "yes." In about 20 min the patient can complete this set of 240 questions. At the end of the questioning the card is removed from the device and read into the computer from a small remote card reader. The computer prints the history immediately at the remote screening clinic on a printer.[2] The printed history consists of a formated set of statements corresponding to the "yes" answers given by the patient. These are organized according to family history, past history, allergies and system review. Since the questions are all binary, only 1 bit is required in the computer record to store each answer and the whole questionnaire can be stored in 10 words of this 24-bit machine.

At one station in the screening clinic a sample of venous blood is withdrawn and sent to the laboratory for hematological and chemical determinations. Upon receipt of the sample in the laboratory, cards are punched with the patient numbers and read into the computer which generates a sample tray loading list for the 12 channel autoanalyzer. The 12 chemical determinations from this blood are then automatically entered into each patient's computer file by sampling the output of the autoanalyzer at the appropriate time. Editing of this data by the medical technologist can be performed through a remote terminal in real time if this is necessary to correct machine errors due to such things as saturation of a reading or malfunction of the autoanalyzer. Thus, some quality control of the data is accomplished before any hard copy is generated from the patient's data file. Hematological measurements are entered automatically or manually depending upon the availability of automated counting equipment appropriately interfaced to the computer.

Another station in the clinic is manned by a technician who records electrocardiograms and performs spirometry (*6*). After stripping to the waist and put-

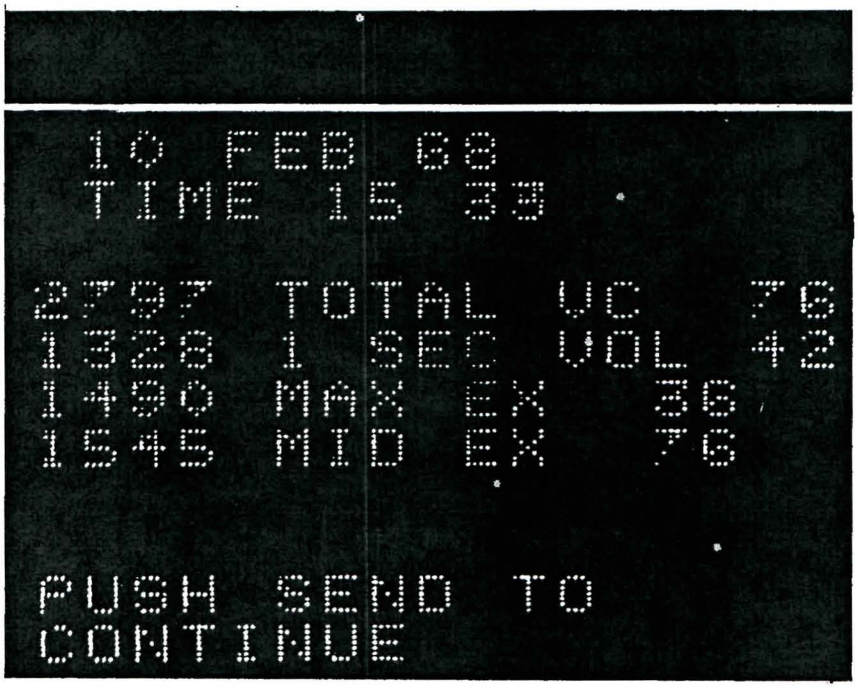
---

[1] IBM 3000 Inforecorder.

[2] Teletype Inktronics Terminal.

ting on a gown, the patient enters a room where the technician requests that the patient inspire maximally and expire into the spirometer. The movement of the spirometer is detected by a potentiometer which in turn is sampled through the A to D converter remotely by the computer. The computer recognizes the appropriate points on the waveform to determine forced vital capacity, 1-sec volume, maximum expiratory flow rate and midexpiratory flow rate. Calibration and correction for temperature and barometric pressure are carried out and the values determined are presented not only as absolute volumes and flow rates, but also as percent of predicted values based on height, weight, age and sex of the patient as shown in Fig. 4. These results are displayed on the oscilloscope to the technician and, if abnormal, the tests are repeated at least once. The best values are saved in the patient's record.

The technician then attaches the necessary ECG electrodes to the patient to record not only the 12 standard leads but also three Franck vector leads. Three leads at a time are recorded and when the standard leads are complete the machine is switched automatically to the vector position. At this time the technician interrupts the computer to start a sampling routine which samples each of the Franck X, Y, and Z leads 100 times per second. When five heart cycles have been recognized 10 parameters are identified and measured from each of these five waveforms. Each of the parameters is averaged among the five values



```
  10 FEB 68
  TIME 15 33

2797 TOTAL VC    76
1328 1 SEC VOL   42
1490 MAX EX    36
1545 MID EX    76


PUSH SEND TO
CONTINUE
```

FIG. 4. Computer storage scope display of results of pulmonary function test. On the left are the absolute values of forced vital capacity, one-sec volume, maximum expiratory flow rate, and maximum midexpiratory flow rate. On the right is the percentage of the predicted value for each test based on the patient's height, weight, sex, and age.

and the value differing furthest from the average is discarded and the remaining four averaged again. From these average parameter values a series of logical operations are performed, comparing parameter values at each step to empirically derived constants in order to classify the electrocardiogram. This classification, along with the parameters, are displayed immediately on the remote terminal storage oscilloscope in the screening laboratory. If the electrocardiogram is abnormal, it is repeated, and if confirmed, the values are stored in the patient's record. The patient now leaves the screening area and goes to his hospital bed. When the blood chemistry, hematology, and urine values have been measured a hard copy report is generated, including a summary of abnormal findings, and delivered to the Nursing Station for filing in the patient's chart.

Thus, all this information is available to the resident or staff doctor when he first sees the patient for work-up in the hospital. Any subsequent lab work, ECG's or spirometry performed on the patient during his hospital stay is also entered into the patient's record and logged according to the time the procedure was performed. An immediate review of any patient's computer record is available through any remote terminal. Fig. 4 is a typical page of such a record review.

ICU Monitoring

The Intensive Care Unit monitoring system has been well-documented in other publications (7), therefore, this resume will be concerned only with those portions of the system having more direct bearing on medical record organization.

Stroke volume, heart rate, cardiac output, systolic duration, total resistance, systolic pressure, diastolic pressure, and mean pressure are computer monitored on each ICU patient by means of an in-dwelling central arterial catheter. These measurements are made automatically every fifteeen minutes, but the computer will also make them at any time the ICU personnel request them (by pressing a button on the ICU "command console"). In addition to the eight parameters mentioned, nurses in the ICU enter into the record the patient's vital signs, fluid balance information (IV, blood, urine, etc.), IV fluid record, drugs administered, and other comments pertinent to the patient's physiologic state.

The most recent data from this large data base are essential to the improved patient care that can be delivered from the computerized ICU system, but data that are not current (say within the past 3 days) need not be kept in such detail. Consequently a program is available to replace the 15-min measurements of the aforementioned physiologic parameters with a single 24-hr average value. This compression applies only to data more than 3 days in the past.

As a companion to the data compression cited above is the simplification of the display of important parameter changes to the physician. Although detailed hardcopy reports are suitable for documentation and leisurely study, they do

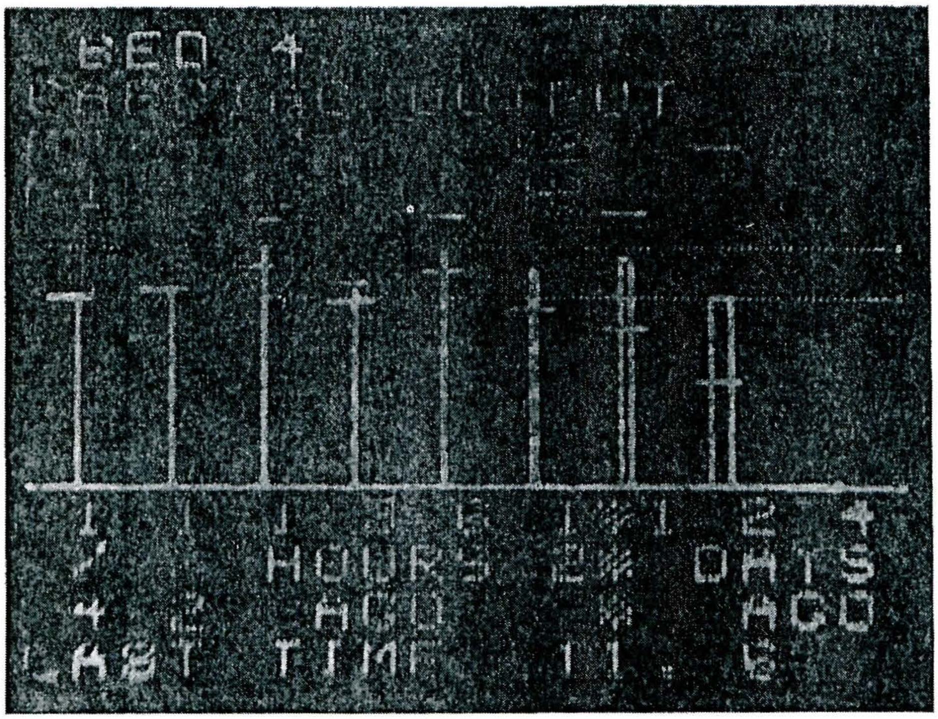


FIG. 5. ICU "patient course profile" display on a remote terminal. The patient's average cardiac output is displayed in a bar graph for time periods of ¼, ½, 1, 3, 6, and 12 hr and 1, 2, and 4 days in the past. The width of the graph is proportional to the number of measurements involved in the average. The horizontal lines associated with each bar are the maximum and minimum for the same time period.

not provide physicians with the immediate grasp of the patient's course needed in critical moments. The needed "patient course profile" is best observed graphically, as shown in Fig. 5. The profile can be produced immediately for any of the eight parameters the physician desires, however, in critical situations the first parameter to be displayed is that which differs most from its previously defined acceptable limits (*8*).

Diagnosis Entry

Diagnoses are entered in coded form to

1. Reduce number of storage cells required for each diagnosis (each narrative phrase requires only one word of storage).
2. Provide a standardized nomenclature. This provides quality control and eliminates vague and ambiguous diagnostic entries.
3. Improve retrieval efficiency. Retrieval using logical operators (And, Or, Not) on numerical codes that can be masked is very rapid.

The Systematized Nomenclature of Pathology (SNOP) developed by the College of American Pathologists (*9*) serves as a basis for the numeric coding scheme. Although some additions were necessary to the code (particulary in the areas of psychiatry and obstetrics) to make the code suitable for clinical use, the detail and logical arrangement of the code more than compensated those deficiencies. Each diagnostic entry in SNOP has components from four interdependent areas of information: (1) The part of the body affected (Topography); (2) the structural changes affected (Morphology); (3) the etiologic agent (Etiology); (4) the functional manifestations (Function). A diagnosis so specified is complete enough to be used for research purposes. Surgical procedures are entered using an expanded version of HICDA coding.

To provide for real-time entry of diagnoses from remote terminals, the entire multilevel SNOP code tree is placed on magnetic disk. In addition to the basic tree structure an alphabetic dictionary is stored on disk (dictionary size is currently 2200 words and phrases). Each word or phrase in the dictionary has associated with it a pointer to the disk address of a list in the tree at the proper level (see Fig. 6). Although it is possible to progress through the tree from level 1 on down, it is easier and more efficient to select a phrase which will enter the tree at as detailed a level as possible.

The system can best be further described by resorting to an example. Assume the diagnosis "acute myocardial infarct" is to be entered into the patient record. After entering the patient's hospital identification, and receiving the patient's name in reply from the computer as evidence the proper file has been found, the physician is presented with a display requesting him to choose to enter (1) the anatomic location of the lesion; (2) the morphology of the lesion; (3) the etiology; (4) the functional abnormality.

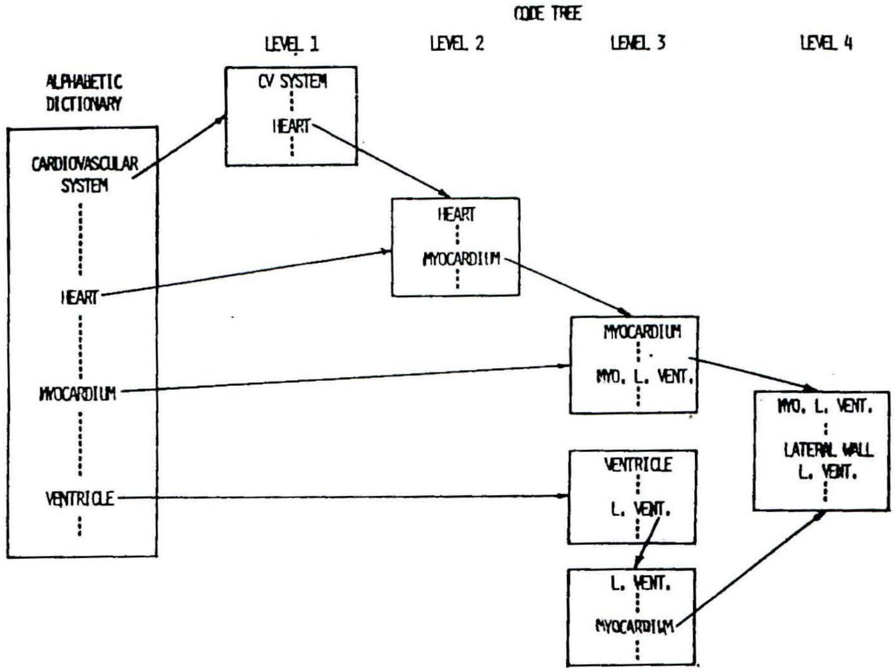


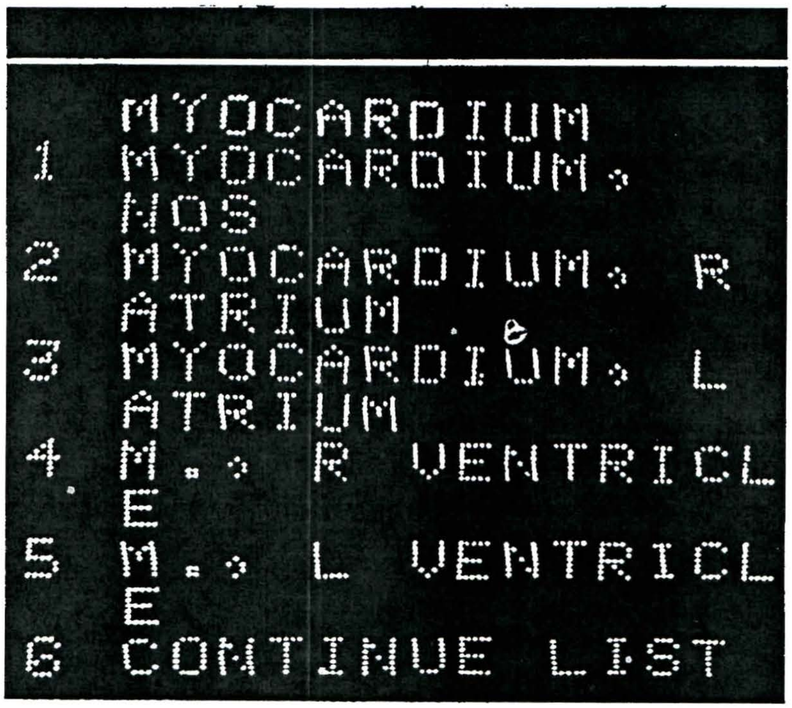
FIG. 6. Selection of more specific alphabetic term for diagnosis causes entry to code tree at a more specific level, thus eliminating wasted time choosing from less specific parts of the tree.

Assuming the physician desires to enter the anatomic location, he chooses No. 1 and receives a display requesting that he enter letters on the alphanumeric keyboard before him to spell the key word of the anatomic location (myocardium). After entering the letters MY the computer makes a search of the dictionary terms for those beginning with MY and pertaining to anatomic location. If five or more items are found satisfying these criteria, the computer requests another letter be entered, and once more a search is conducted looking for five items or less. In this case MY is sufficient and the following display appears:

1. myelocyte
2. myelopoietic
3. myocardium
4. myometrium

"Myocardium" is chosen and the program enters the code tree (Fig. 6) at level 3, and displays the myocardium list (Fig. 7). At this point the physician is asked for more information than was originally offered. He must specify where the infarct is in the myocardium. He chooses myocardium, left ventricle and then lateral wall of left ventricle at level 4 (Fig. 6). The anatomic location is now completely specified. A similar procedure is followed for the morphology, etiology, and functional abnormality associated with the lesion. The physician then dates the diagnosis and the computer stores the code numbers associated with the narrative expressions he has chosen in the patient record.

The proceeding real-time "conversational" method of entry of the diagnosis



```
  MYOCARDIUM
1 MYOCARDIUM,
  NOS
2 MYOCARDIUM, R
  ATRIUM
3 MYOCARDIUM, L
  ATRIUM
4 M., R VENTRICL
  E
5 M., L VENTRICL
  E
6 CONTINUE LIST
```

FIG. 7. Storage scope display of "myocardium" list at level 3 in Fig. 6.

at a display terminal obviates the need to use the code book, and provides the necessary feedback to the physician to upgrade the quality of the diagnosis. It also permits the researcher or clinical committee who is to eventually use the diagnostic data to determine the level of specificity acceptable in the diagnosis, rather than to accept a possible lower level of specificity originally offered by the clinician. As an example, consider the morphologic entry of "tumor" which is a fairly common entry in discharge diagnoses, although the expression is so ambiguous as to be virtually of no value. Upon receipt of the term "tumor" the computer replies with displays from which the physician makes choices to specify if the tumor is neoplastic or not, benign or malignant, what specific type of neoplasm, metastatic or not, etc.

Data Retrieval

As pointed out in the introduction, this system was begun primarily as a research tool. As such there is no provision for rapid access to past patient records (as noted in Fig. 1 when the patients are discharged, their records are stored on magnetic tape). The slower tape search is entirely acceptable for most research applications, however.

Programs are available to search the master tapes containing all patient records for particular disease combinations (i.e., patients with ischemic heart disease, but not with concurrent emphysema) and to transfer the records of those patients with the given combination to a subtape. This subtape, containing a small subset of the total number of records, can be kept in the possession of a researcher without his need of again searching all master tapes (Fig. 8).

Another program then enables the researcher to select only patients on the

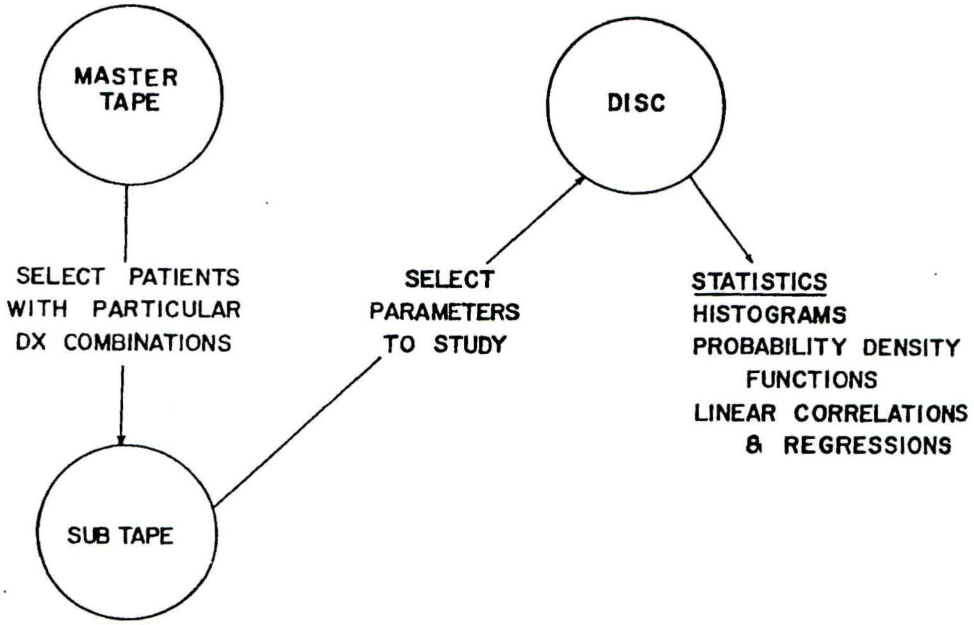


FIG. 8. Research oriented retrieval method.

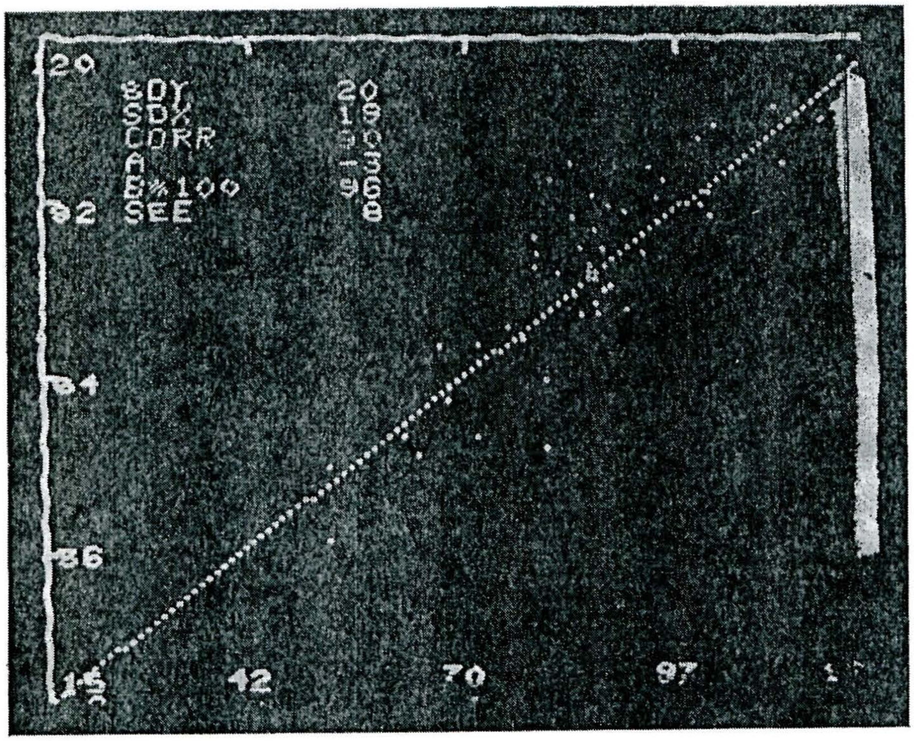


FIG. 9. Storage scope display of the regression of forced vital capacity as percent of predicted on the 1-sec volume as percent of predicted in patients over 20 years of age. Shown in tabular form are the standard deviations of the *Y* and *X* variable; the correlation coefficient as a percent; *A* and *B* for the linear regression equation; and the standard error of the estimate.

subtape having given parameters within selected limits (e.g., age greater than 50; LDH between 85 and 150, etc.) He then selects the variables to study (e.g., SGOT, cholesterol, forced vital capacity, etc.) Variables from the patients having the given parameters within the given limits are then stored on disk for rapid statistical analysis.

A final program then permits the researcher to plot histograms, fit three parameter density functions to the histograms, and perform linear correlation and regression on the data (Fig. 9).

All the above analysis programs are done in real time using remote terminals for display of plots, thus permitting the researcher an optimal medium for refinement and change in the experiment.

REFERENCES

1. HORTY, J. F. Patient care is the goal of medical records. *Mod. Hosp.* **100**, 112 (1963).
2. SINGER, S. J. Visual display terminals in a hospital information system (HIS). *Comput. Biomed. Res.* **3**, 510 (1970).
3. COLLEN, M. F. General requirements for a medical information system (MIS). *Comput. Biomed. Res.* **3**, 393 (1970).

4. Feinstein, A. R. Quality of data in the medical record. *Comput. Biomed. Res.* **3**, 426 (1970).

5. Pryor, T. A., Gardner, R. M., and Day, W. C. Computer system for research and clinical application to medicine. AFIPS Conf. Proc. Vol. 33, 1968.

6. Dickman, M. L., Schmidt, D. L., Gardner, R. M., Marshall, H., Day, W. C., and Warner, H. R. On-Line computerized spirometry in 738 normal adults. *Amer. Rev. Resp. Dis.* **100**, 780–89 (1969).

7. Warner, H. R. Experiences with computer-based patient monitoring. *Anes. Anal.* **47**, 453–62 (1968).

8. Warner, H. R. A data-trend log and bargraph display for physiological monitoring. *Comput. Biomed. Res.* **3**, 285–288 (1970).

9. Systematized Nomenclature of Pathology. Committee on Nomenclature and Classification of Disease, College of American Pathologists, Chicago, 1965.