Current Status of a Medical Information System*)

UDC 681.3 : 616

(From The Permanente Medical Group and Kaiser Foundation Research Institute, Oakland, California)

E. VAN BRUNT, L. S. DAVIS, J. F. TEDDIMAN, S. SINGER, E. BESAG and M. F. COLLEN

A pilot medical information system is being implemented and currently is providing services for limited categories of patient data. In one year, physicians' diagnoses for 500,000 office visits, 300,000 drug prescriptions for outpatients, one million clinical laboratory tests, and 60,000 multiphasic screening examinations are being stored in and retrieved from integrated, direct access, patient computer medical records.

This medical information system is a part of a long-term research and development program. Its major objective is the development of a multifacility computer-based system which will support eventually the medical data requirements of a population of one million persons and one thousand physicians. The strategy employed provides for modular development. The central system, the computer-stored medical records which are therein maintained, and a satellite pilot medical data system in one medical facility are described.

DER GEGENWÄRTIGE STAND EINES MEDIZINISCHEN INFORMATIONSSYSTEMS

Ein im Aufbau befindliches medizinisches Informationssystem versieht zur Zeit Dienstleistungen für eine begrenzte Kategorie von Patienten-Daten. Innerhalb eines Jahres werden die ärztlichen Diagnosen von 500 000 Arztbesuchen, 300 000 Arzneimittelverschreibungen, 1 Million klinische Laboratoriums-Teste und 60 000 Ergebnisse von Vorsorgeuntersuchungen gespeichert und als integrierte Computer-Krankengeschichten im Direktzugriff abrufbar sein.

Dieses medizinische Informationssystem ist Teil eines Langzeit-Forschungs- und Entwicklungs-Programms. Sein Hauptgegenstand ist die Entwicklung eines Vielzweck-Computersystems, das den Bedarf an medizinischen Daten für eine Bevölkerung von einer Million Menschen und 1000 Ärzten befriedigen soll. Das zentrale System, die darin enthaltenen medizinischen Angaben und ein Satelliten-Versuchssystem für die Daten einer Behandlungsstelle werden beschrieben.

Introduction

The cost of information handling in a hospital is approximately one-third of a hospital's per diem costs (4, 5); as a result, many medical centers are now attempting to control these increasing costs by installing computerized information systems. The Kaiser-Permanente medical information system is in the process of being implemented, and this constitutes a report on its current status.

A medical information system (MIS), as herein defined, is one that utilizes electronic data processing and communications equipment to provide real time processing of patient data within one or more general medical centers, including both hospital and outpatient services.

*) Supported in part by Kaiser Foundation Research Institute, National Center for Health Services Research & Development (under Grant #HS00288), and Food and Drug Administration (under Contract FDA #68—38).

Our long-term objective is to develop a multifacility computer-based system that will support eventually the medical data requirements of one million people in the San Francisco area, one thousand physicians, and a large corps of professional and paramedical personnel involved in patient care. Our immediate objectives are: (a) to establish in one medical facility, a pilot data system subserving a representative spectrum of functions, and (b) to develop and implement an evaluation program, some of the results of which will be used to guide the design and implementation of data systems in the second and future medical facilities.

These objectives can be achieved by one of two basic strategies: (a) treat each hospital and clinic facility as an independent data processing entity; or (b) treat each hospital and clinic facility as a »terminal station«, serviced by a single »regional« data processing center. The latter plan has been adopted for this MIS.

Selection of those medical data to be collected, and of methods for doing so, was accomplished as a natural outgrowth of several years of experience with various data processing techniques employed in the development and operation of an automated multiphasic screening project (1). The techniques include on-line telecommunications operation and the use by physicians of optically read marked forms. Data collected in the multiphasic screening project include selected clinical laboratory tests (including chemistry, hematology, and urinalysis), x-ray (chest) examination, electrocardiographic findings, and physician diagnoses. In August of 1968 a new system went into operation with one application program to collect multiphasic data. In June, 1969 a second application program handling five on-line typewriter terminals in the San Francisco outpatient pharmacy was added to the system, and prescription data is currently collected on about 1200 patients a day. Since then, many application modules have been added. Diagnostic information is being collected on a daily basis, on optically read forms, for 13 specialties in the outpatient clinics in San Francisco. Also, multiphasic data is being captured for the Walnut Creek facility and for a pediatric multiphasic clinic in San Francisco. Finally, patient identification information has been assembled and stored in 1,135,000 directly accessible records.

The Pilot Medical Data System

The site of operation for the pilot medical data system is the San Francisco Kaiser Foundation Hospital and associated Permanente Medical Group Clinics. Together, these facilities provide acute and continuing care in essentially all medical and surgical specialties, and are available to a local population of approximately 125,000 Kaiser Foundation Health Plan members. There are about 2,000 physician office visits per day. Average occupancy of the 304-bed general hospital is 85 percent; an average of 40 patients daily are admitted to the hospital.

The specific goals of the San Francisco pilot data system are: (a) to acquire, and to store in the computer record of each hospital or clinic patient, on a continuous basis: all diagnoses, results of all laboratory tests, x-ray, pathology, and electrocardiographic examinations, and data concerning all drugs dispensed or administered; (b) to provide initially limited, but progressively expanding, services to the professional and technical staff, such as printed reports of test results and, in the hospital, visually displayed and printed sets of data (doctor's orders, medication schedules, laboratory specimen labels, etc.); (c) to provide a broad and flexible data base for clinical, epidemiological, and health services research.

The functional organization of the pilot data system is separated into outpatient and inpatient subsystems.

The Outpatient Data Subsystem

Patient registration data and physicians' diagnoses from the outpatient department are being recorded daily on specially designed standardized forms. Between 90 and 95 percent of the diagnoses are selected from a structured format. The remaining unstructured data is manually inscribed on the form; thereby, less than ten percent of this natural language (English word) data to be stored requires keypunching for input. Using, wherever possible, standardized forms with consistent terminology for data input to medical records, simplifies the storage and retrieval of medical data as well as the construction of dictionaries and thesauri. These forms are dispatched to the local hospital's record room where an optical scanner reads the structured data and generates punched cards; the latter are batched and transported to the computer center.

In the outpatient pharmacy prescription data, including patient and physician identification, drug name and dose, prescription refill and drug usage data are entered by pharmacists using on-line electric typewriters, directly into the appropriate patients' computer medical records. These records reside in direct-access mode in the central facility. Container labels are produced under program control, to be dispensed with the drug.

In the automated multiphasic screening laboratory, a wide variety of medical data is collected and »advice« rules are generated in real time to aid in follow-up patient care.

The Hospital Data Subsystem

The hospital data subsystem is in the final stages of development (6). The basic configuration consists of a satellite processor linked by duplex telephone lines with the central computer facility. This satellite computer system comprises Honeywell DDP 516 and 416 computers with a total storage capacity of 40K words of core memory (16 bits per word) and approximately four million characters on six disk drives. The potential exists for added core memory (limit of 64K) and added disk storage. This configuration represents only one-half (System »A«) of what is planned as the completed pilot hospital data system in which a duplication (System »B«) of equipment will provide »failsoft« reliability (Figure 1).

The System »A« processor drives 24 Sanders Associates display terminals. Each terminal consists of a cathode ray tube display device with associated »light-pen« sensor, typewriter keyboard, card-reader unit, and an electronic printer enclosed in a sound-proofing box. The cathode ray tube is capable of displaying approximately 1024 characters in a 2048 position matrix measuring approximately ten by seven inches. The display matrix is arranged in 40 lines of 52 characters.

Twenty-one terminals are to be deployed throughout patient care areas of the hospital, including all nursing stations, the nursery, intensive care unit, emergency room and admitting department. The average number of patients per terminal is 18. Three terminals are assigned to the clinical laboratory for logging-in of test specimens, test reporting, and production of work lists and labels.

The system of terminals (Figure 2) is that portion of the data system which will interface with medical care personnel including doctors, nurses, and selected technicians and clerks. It will serve for data input and output

of patient diagnoses, selected signs and symptoms, physicians' orders (including orders for drugs and general nursing orders), drug administration data, medication schedules, and clinical laboratory, x-ray, surgical pathology, and electrocardiogram results.

ing systems and applications programmers to write, enter, and debug programs directly via the visual display screen, and (b) relative ease of programming by allowing execution of numerous operations while using relatively few program statements (5).

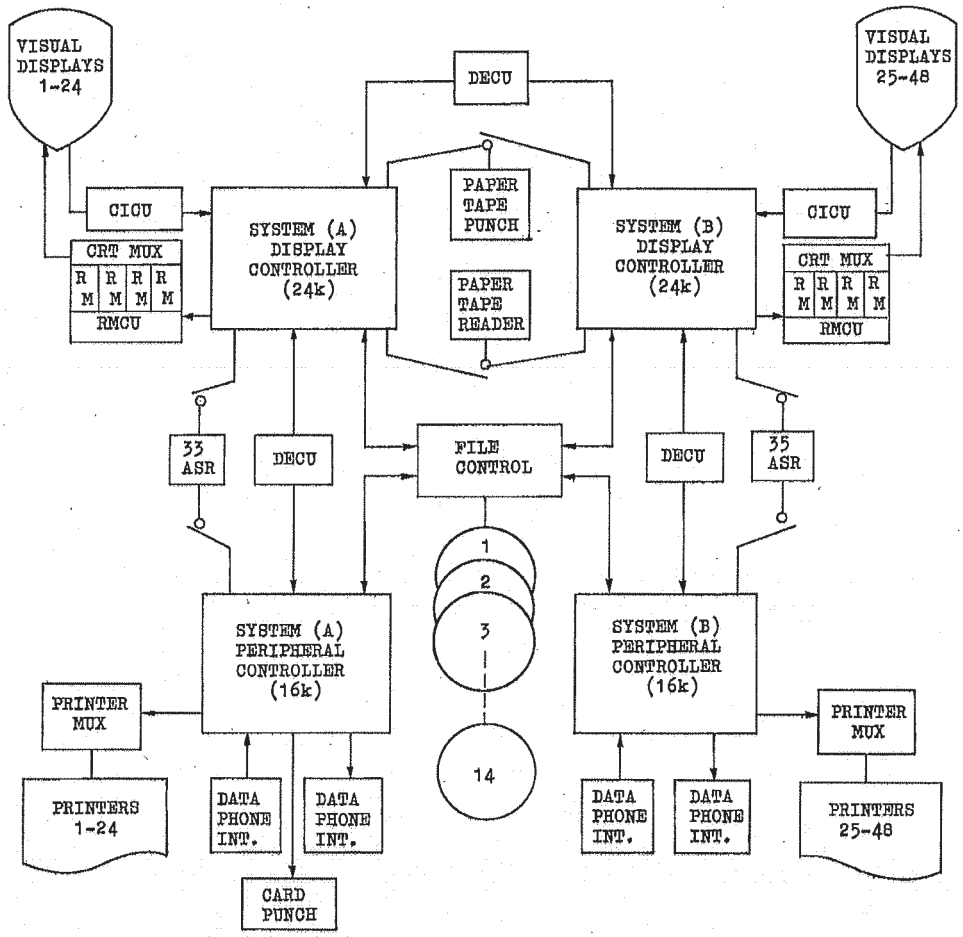

Fig. 1: One proposed configuration of integrated »fail-soft« computer systems (A+B) for a single hospital

At the time of admission of a patient to the hospital a file, composed of selected elements of the patient computer record, is called from the central facility and stored in real time in the memory of the satellite data system. Working with elements of this patient file and various applications programs, the authorized user (identified by his coded machine-readable identification card) may enter or retrieve data by interacting directly with the visual display terminals. The goal of high-quality data aquisition requires that responsibility for data input rest with the individual most capable of assuring its accuracy and completeness. Thus, physicians will enter orders and diagnoses, nurses will »chart« drug administration data and laboratory technologists will enter test data.

Displays seen at the terminals are under control of programs running in the small satellite computer. Programming is done in a file oriented, higher level language, FOPS (File Oriented Programming System). FOPS is a multiprogramming, list processing, virtual memory interpretive language system. Its capabilities include (a) allow-

Data is organized into files; each file may have any number of subfiles and each subfile may have many pages. A page corresponds to that which is seen on the screen of a visual display and each page has a corresponding form or overlay. The process of constructing a display consists of displaying a form, which may be blank or highly structured, and then filling in the corresponding blank parts of the form from the page.

The strategy of employing forms is useful in conserving storage since it is often the case that the same displays, with only minor changes, are used many times.

Several logical steps are required to input information into a hospital information system:

1.) Identification of the user to the system;
2.) System acknowledgement;
3.) Identification of one or more patients;
4.) Identification of information to be entered in the terminal;

5.) Verification that the selected information is correct;

6.) Printing of the information locally and transmission to the central system for storage in the patient's computer record.

Because the system is large, involves many people and privileged information, it is necessary to place restrictions on system response to a given user. An admission clerk may not »write« a medication order. A doctor will not enter laboratory test data. Appropriate constraints are placed on the user by identifying him as a member of a certain class. Identification is accomplished by inserting a machine-readable card in the reader on the terminal. The unique card code identifies the user as an individual member of a class; various displays are thus available only to certain classes of users.

The system acknowledges request for service by displaying the name of the user and the date and time. This name, date and time is printed on every document provided to the user by that terminal. Most transactions by physicians involve a single patient, or at least one patient at a time, during input of information. The next step is the identification of a patient, followed by input of information about that patient, such as a nursing order, a diagnostic statement, etc. Finally, the information is displayed back to the user for verification and correction of any errors. When correct, it is printed and placed in the hospital chart.

Examination of many different tasks has led to a fairly uniform functional layout for visual displays. Uniformity is useful because it simplifies the training process

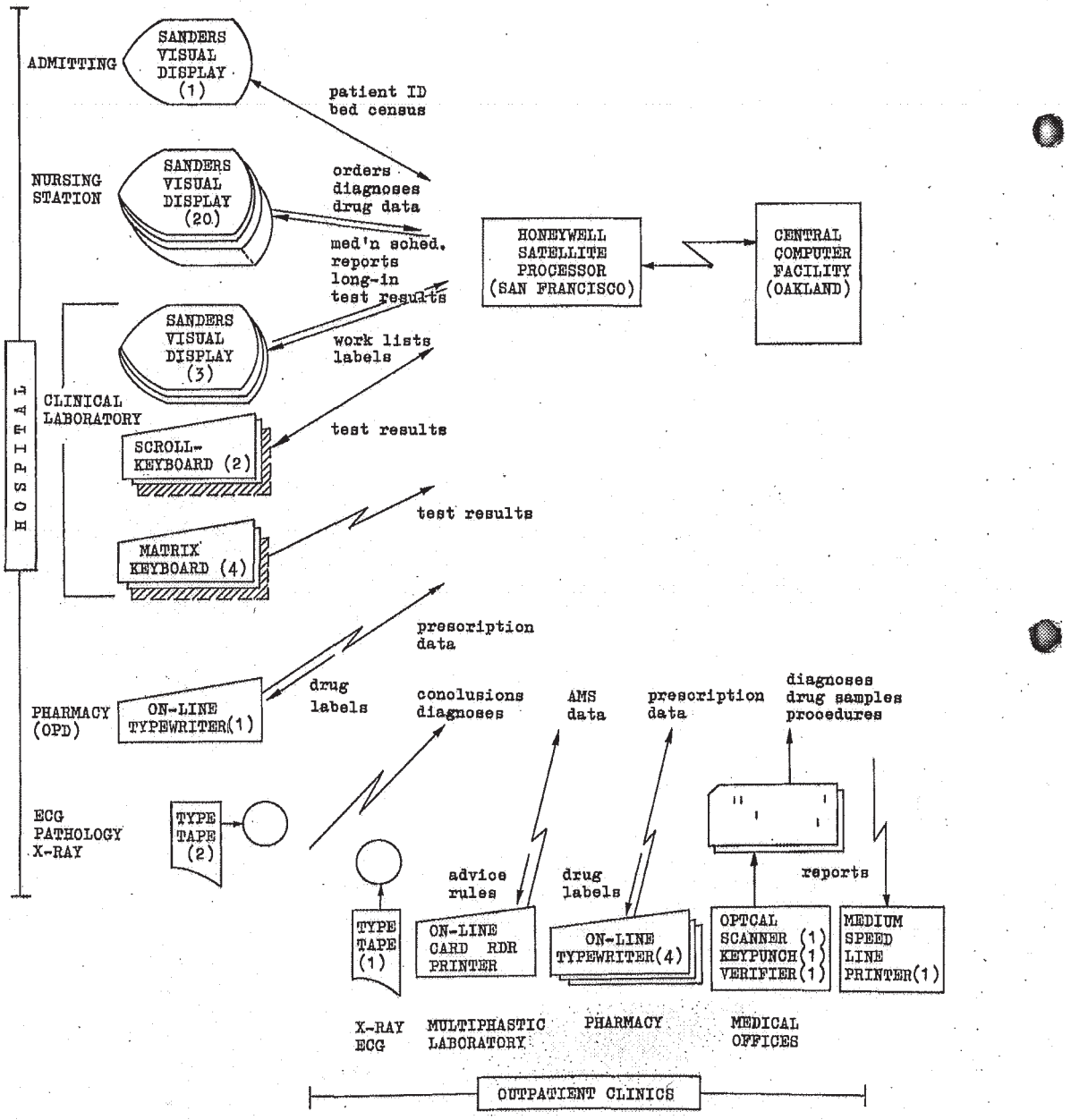

Fig. 2: Single facility pilot medical data system

required before the user becomes proficient in utilizing the system.

The basic display format contains a heading at its top which identifies the user, the patient, the date and the time. Below the heading is a »status area«, which always contains the last four lines of information entered. If more information is entered than can be contained in 4 lines, the oldest data is »rolled off« to a displayable page called a »work page«. Below the status area is a title that identifies the particular display. The »body« of the display contains those data to be manipulated by light-pen selection. Data may thus be transferred to the status area, or a given selection may simply lead to another display. The bottom area contains commonly used command functions, such as: return to an index, error correction, or page forward or back. It also contains certain common modifiers to be added to previously selected area.

The strategy developed for various tasks takes into account the functional limitations of existing terminal devices, but attempts to simulate as much as possible the techniques already familiar to the user. Displays are constructed with the intent that required information be located in the shortest number of steps and in the shortest possible time. This generally involves two familiar techniques: multilevel indexing and the use of prestructured commonly used information.

For example, the drug ordering sequence starts with identification of the doctor from his card. Upon reading the card, the terminal will display a list of patients logically associated with that terminal; this list will contain all the patients at the nursing station serviced by that terminal.

The doctor selects a patient by touching the patient's name with the light-pen. The name is transferred to the heading area and a general index is displayed. A class of orders may then be selected from the general index as, for example, drug orders-alpha index. Selection of an appropriate letter from the »drug alpha index« causes the display of a page of drugs starting with the given letter. If the drug desired is not on the page displayed, then selection of the function »forward page« at the bottom of the display will produce the next page in the alphabetic drug list.

The drug list is structured as follows: on the left of the display are drug names and forms arranged in alphabetic order, and on the right are the most common orders for selected drugs. One can generate an entire order by »light-penning« the common order in the right or, alternatively, simply select the drug name. Selection of the drug name causes display of a drug »sig page« with the drug name displayed in the status area. This page contains a large variety of options for selecting the dose, route, frequency of administration and special instructions.

The ordering sequence is completed by selecting the function »display workpage«. This results in a display of all the orders generated by the user during that period of interaction. The order sequence is ended by selecting one of two functions: »finish« (print and return to census), or »delete orders« (and return to census). Orders will generally be printed just as they appear on the display. The printout can then be initiated by the physician and placed in the hospital chart.

Although the system has not yet been installed in the hospital, it has been operated by a small number of physicians, most of whom are quite unfamiliar with computers. Certain tentative conclusions are possible at present: the overall strategy appears acceptable, and most people can learn to »write« orders with only a few minutes' instruction and a little experimentation. Summary displays are expected to be attractive to users since information such as »current active orders« and »status of outstanding laboratory requisitions« is not always easily available in the hospital at present.

User errors are frequent at first, but diminish with time. The commonest error is an incorrect light-pen selection which usually is noted immediately. Several techniques have been provided for error correction. Many displays have an »error« function at the bottom. Light-pen selection of »error« causes deletion of the prior selection. An entire entry may be deleted by displaying the work page and light-penning the entry.

Entries are presently checked for internal consistency and, when appropriate, an error message is displayed and blinked in the status area to call the error to the attention of the user. For example, the system will not allow only a modifier, such as »right« to be entered as a diagnosis. Error checks for many conditions are planned for early implementation, for example: drug doses, compatibility of drugs with each other, etc. Primary limitation for this type of checking is the processing time required and storage space for programs. The potential for this type of error checking emphasizes the desirability of a computerized data system.

Efficient utilization of visual terminals requires some structuring of data. While many forms used in hospitals today are already highly structured, such as an admission form, others are essentially blank pieces of paper, like the progress note form. While it is possible to handle unstructured natural language text in a computer the process is time consuming and error prone; here the visual terminal offers only modest advantage over a simple typewriter terminal. The diagnosis input sequences represent a first step toward structuring doctors' progress notes.

Since it is not possible to anticipate every possible entry for a pre-structured visual display, most sequences allow selection of a function such as »key«. This allows free text to be typed into the status area of the display.

Present visual displays, especially those using character generating units, are limited in resolution, character form, and color. A well designed paper form may contain five to six times as much information; that is, it takes five or six visual displays, of the capacity described above, to present the same information as a paper form of roughly the same size.

One of the critical factors in acceptance by the user of visual display units in a hospital system is the response time, or how long it takes to change displays. To be acceptable the response time, at a minimum, must be fast enough so the user cannot make a mistake such as light-penning another entry before the computer has finished processing a previous one. In practice, this means that the response time must be of the order of one second or less.

CENTRAL HARDWARE AND SOFTWARE

The central computer facility consists basically of two IBM 360/50 computers and associated equipment. One computer functions as an on-line device; the second subserves an essential »back-up« function by means of manual switching (approximately ten minutes switching time) and is used for off-line research, testing and »de-bugging«. Each computer has access to its own set of input-output devices and files; either can be connected to the telecommunications control units, which interface with all remote terminal devices (Figure 3).

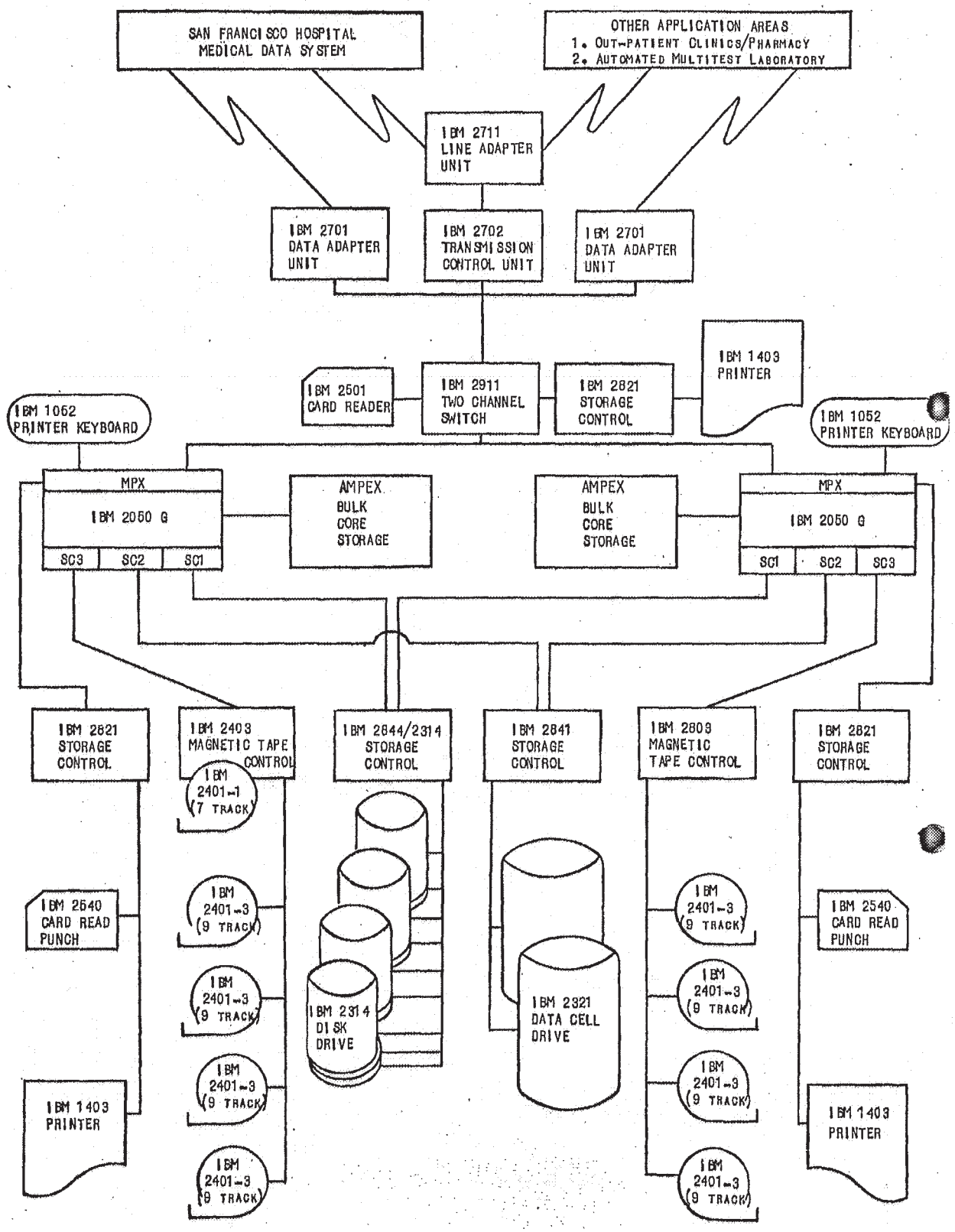

Fig. 3: Equipment configuration for central computer system

Computer storage of readily accessible patient records requires a random access mass storage device. Each *Patient Computer Medical Record* (PCMR), which contains identification, administrative and medical data, may consist of hundreds or thousands of characters (bytes). Therefore, to store all medical records of a million people, a storage device with a multi-billion bit capacity is required. The storage problem for this regional MIS is further complicated by the fact that the volume of medical data is steadily increasing. Even if the total patient population were stabilized at a fixed number, the volume of medical data would still continue to grow as patients continue to receive medical services. A current PCMR may contain identification and administrative data (120 bytes), medical data from multiphasic clinics in San Francisco and Oakland (1200 bytes per visit), and diagnoses from San Francisco outpatient clinics (30 bytes per visit). However, this data forms only a small fraction of the total volume of medical information that will be received when the full system is operational. The bulk of the data will come from medical services to hospital inpatients including drug orders, drug administrations, and laboratory test results. There is little doubt that within one to two years, additional storage will be needed.

It must be possible to access at random any PCMR in an acceptable period of time. This working principle is based on the premise that use of the MIS must not impede the delivery of medical services in any way. Two considerations in determining maximum PCMR access time were: 1.) the expected peak load of PCMR retrievals in a given period and 2.) the problems of man-machine interaction. The San Francisco pilot facility, on an average day, has 2000 outpatient visits and 1200 drugs dispensed. It is clear from these operating characteristics that access times in excess of one second would result in frequent queueing of retrieval requests and consequent retardation of operation of out-patient services. Furthermore, pilot studies of visual display terminals at hospital nursing stations indicated that, for many operations, physicians will only tolerate terminal response delays of the order of one second.

Two IBM 2321 Data Cells are used by the MIS for on-line PCMR storage. The system is designed in such a way that, the first time a patient's record is accessed in a given day, his PCMR is retrieved from the data cell in its entirety and transmitted to an intermediate device — an IBM 2314 disk storage unit. The reason behind this data transfer is that, if a PCMR is accessed once during a day, there is a high probability that it will be accessed again within a short period of time. Since the average access time for a record on disk is only 75 milliseconds, compared to 400 milliseconds for a data cell, placing a PCMR on disk greatly decreases access times for subsequent data retrievals. This procedure also serves to minimize use of the data cells by avoiding additional data cell accesses for the same PCMR on the same day. At the end of the day all PCMR's on disk have been updated and are transmitted back to the data cell. The updated PCMR's are stored at the end of the active file in the data cells while the storage areas formerly occupied, by the old PCMR's which are no longer valid, can be re-used by the system after execution of appropriate cleanup procedures.

There have been approximately three million accesses to two data cells since the beginning of on-line operation in August, 1968. We estimate that about one half of these PCMR retrievals were initiated by transactions in the pharmacy and multiphasic clinics and by the initial loading of PCMR data, while the remainder were due to the operation of data retrieval research programs and internal systems programs. At present an average of 3600 PCMR retrievals per day are made from the two data cells.

In view of the large number of retrievals the reliability of the data cells is of major importance for efficient system operation. When a data cell or its controller is disabled no PCMR is accessible unless it had been transmitted to disk prior to breakdown. Medical information destined for entry into a PCMR stored in a disabled data cell must be saved in temporary storage — on disk, tape, or off-line in the form of punched data cards or written documents — until the data cell is repaired. Of all our computer system hardware, the data cells have had the poorest operating record. The reason for their relatively frequent failure is that they contain a large number of moving parts, more than any other system component. Based on their past performance, we are concerned that the data cells may be inadequate for meeting the requirements of a regional MIS.

In order to support the large number of input-output and processing functions, the *Medical Function Control System* (MFCS) was developed. The essential function of MFCS is to maintain and coordinate the numerous data sets that constitute the medical record files. Five groups of programs comprise the MFCS: medical record manipulation routines, encoding and translation routines, medical language routines, medical function routines, and a medical function control program. These routines include direct-access storage and handling of remote terminal as well as the usual input and output devices.

A program routine called UPDATER is responsible for all updating of input information to the PCMR. UPDATER receives data in a standardized form from a Medical Input Description Language (MIDL) Compiler module and using MOVER/READER (see below) requests exclusive use of the PCMR. UPDATER processes the standardized form data against the PCMR and, when finished, releases the PCMR for use by other reading programs.

A medical record Directory »dataset« also exists on a disk drive. (A dataset is defined as a major unit of data storage and retrieval in the system, consisting of a collection of data in one of several prescribed arrangements and described by control information to which the system has access). The Directory is the location index of a PCMR in the data cell and in the disc transient storage areas. Input data is always identified by the patient's unique medical record number. That number is used to find the appropriate entry in the Directory, which in turn points to the exact address of the PCMR in storage. A program called MOVER/READER is solely responsible for reading the PCMR, updating the Directory and furnishing portions of the PCMR to requesting functions.

A program called MOVER/WRITER is solely responsible for removing a PCMR from the disc file and adding it to the end of the data cell file, updating the Directory at that time. Normally, MOVER/WRITER operates only at the end of a given time period, but can be called upon with increased priority under program control if the disc transient file datasets exceed predetermined limits. Periodically, MOVER/WRITER can be used to free superseded areas of the data cell file.

Encoding and Translation Routines

The need for compressing data to conserve storage space has resulted in two general concepts: 1.) encoding of text, and 2.) encoding of test and variable identifiers. A dictionary has been developed for the former and a catalog for the latter.

The primary use for an English word dictionary is to encode diagnoses. Several diagnostic classification schemes have been studied but none were considered satisfactory for use in the MFCS. It was therefore decided to store the diagnoses on a word-for-word basis, internally encoding each separate word so as to conserve storage space. This will allow for later transformation to a classification scheme, if desired. The dictionary consists of two datasets: English-to-code dataset and Code-to-English dataset. While decoding usually increases processing time, the Code-to-English decoding is direct (i.e., given a code the English word can be accessed directly) and is used mainly for output generation. The English word dictionary program is responsible for creating and updating these two datasets.

While all variable identifiers, standard medical questions, tests, etc., could be handled on a word-for-word basis using the encoded procedures in the English word dictionary, it was felt that significant storage savings could be achieved by having a second encoding scheme for those variables that have results attached to them. An »item« was therefore defined as an identifier for a value, or set of values, or a group of items; e.g., a test, a procedure, a check-off list, a question, a subset of questions, a questionnaire, etc. The Item Catalog dataset contains the code for each item, the English description for each item, key words and, in the case of quantitative items, upper and lower validity limits. The Item Catalog Program is responsible for creating and updating this dataset.

An item catalog was established as a master index to all items. The item catalog is a direct access dataset within MFCS. It serves not only as a master index to all possible variables (items) within the PCMR but, more importantly, becomes a means of standardization for all input. Before any program can be written to receive data for any medical application via any input medium (e.g., punched cards, optical mark read form, magnetic tape, or on-line terminal), the input variables must be checked against the current item catalog. Any variables or information not already listed must be defined. This definition includes: assignment of the next available catalog item number; a complete definition of the variable including identification of key words; identification of the part and summary bit numbers and what possible formats will be used in storing this data; a short description appropriate for printing item identification on various medical reports; assigning upper and lower machine (validity) limits if the value is quantitative; and a cross index to programs that process this data.

Medical Language Routines

Experience with our automated multiphasic screening project has shown that new input forms, revisions of old input forms, new output report formats, and revisions to old output formats occur with high frequency. In addition, requests for research data arise at a high rate.

It would be impossible to keep up with these requests if each had to be programmed at the symbolic machine language level. Simple interpretive languages were developed in the original multiphasic project to handle these demands. It was felt that more general languages could be developed for the MFCS and in addition could be compiled rather than interpreted at object time. Three language needs evolved. The routines to process two of these languages have been developed for 1.) a language to describe input forms and formats; and 2.) a research retrieval language.

Medical Input Description Language (MIDL): Any data that will appear in core memory from any device, card-reader, tape, telecommunication terminal, or another computer, can be specified by a format description. A high-level language has been developed to describe an input format. The language consists of a passive part and an active part. The passive part is used to describe the placement of the data on the form — location, field lengths, and related characteristics of the data as it is to be entered into the PCMR. The active part is used to express transformations and logical and algebraic relationships of data on the form. The active and passive statements for a given format are compiled by the MIDL compiler. The output (a MIDL module) is placed in the program load library. A given MIDL module is called when data in appropriate form appears in memory. The MIDL module translates the data into a standardized form which is then passed to the appropriate medical function and then to the UPDATER program for storage in the patient's record.

Medical Report Generator Language (MRGL): For reasons stated above, a high-level language was needed to describe the various types of medical reports. Further, various data on the reports must be flagged to indicate abnormalities, deviations from normal ranges, etc. It was hoped that a language could be developed in which the layout of the report could be expressed, as well as the logical and algebraic relationship between the variables. Once a given type of report was established, it could be described, using the MRGL language, and then compiled.

However, our experience has shown it to be more practical and easier to write a series of »macros« to handle the various editing, layout and other report generation needs. The programmer thus writes the program for a particular report, in assembly language, but is aided in that effort by the use of the various report macros.

Medical Data Retrieval: There are two general types of information requests: 1.) specific data from a given patient's record, and 2.) sets of data from the records of many patients who meet various specific criteria.

The first type of data request, usually made by a function program, is handled by MOVER/READER dynamically for specified portions of the PCMR in conjunction with certain system subroutines to retrieve specific items from the PCMR.

Two different types of programs are used for the second type of request: first, the Medical Data Retrieval Language (MDRL) fills the need for a higher language in which simple search requests can be made quickly and easily. It is designed so that a request written in the MDRL can produce output datasets on magnetic tape that are readable by Programming Language (PL/1).

A second type of retrieval allows an investigator to sit at a terminal on-line to the computer, make simple requests, and do analyses with a direct access data subset created by the program above.

Medical Function Routines

The »medical functions« are the application routines. Each medical requirement that can be identified as having independent rules, logic and needs, is called a medical function. However, all medical functions are related in one respect — each requires PCMR data.

To accomplish its needs, each medical function may require its own datasets: appointment lists, patient census, normal ranges, advice rules, etc. In some cases, multiple functions will need to access the same datasets. Some functions will require temporary subsets of patient data. However, the PCMR is always expected to have the most

current patient information; consequently provision is made for updating on a continuous basis.

The current multiphasic project, handling on-line analyses of patient data in the various automated multitest laboratories, then becomes one function. Examples of other functions are the pharmacy terminals, inpatient nurses' terminals, laboratory terminals, etc.

The various medical functions within the MFCS are controlled by the IBM 360 Operating System allowing Multiprogramming with a Variable number of Tasks (OS/MVT). The functions operate as tasks in a multitasking environment under a single job, the Medical Function Control Program (MFCP). Research and other non-function programs operate as independent lower priority jobs, under control of OS/MVT, in a multijob environment along with the higher priority MFCP.

The need for coordinated control of telecommunication input and output to the functions, queued access to the PCMR, sharing of some datasets, operator console activation and deactivation of functions, and function to function communication, requires the MFCP to interface as a submonitor under OS/MVT. The MFCP controls the interaction of the functions and their use of all the available MFCS software discussed above. Priority task-switching of the functions is handled by OS/MVT. The programming language used to write the MFCP is primarily the Assembler Language with FORTRAN IV where possible.

THE PATIENT COMPUTER MEDICAL RECORD

A computer-stored, integrated medical record is central to the functioning of a large-scale multifacility medical data system. Such a record, and the system that maintains it, must be so structured as to store all classes of patient-related data: identification, administrative, and medical. It must be capable of continual updating, accepting the random quantity and timing, and variable format of data input, responding to the need for real-time inquiry to individual patient records, and supporting broad-based research activities. It must be maintained by reliable error detection and recovery routines. An operational computer medical record system which meets these requirements has been brought to an advanced stage of development and is described elsewhere (2).

If a medical information system is to operate successfully, an underlying standard file structure, with system programs to support it, must be developed. The file structure must be sufficiently general so that all forms of medical data can be handled. The system programs should be efficiently coded and should include error recovery and backup routines which are as thorough as possible. The MIS must be capable of capturing and storing all essential medical information in an organized and structured way, such that all potential clinical and research oriented users will have access to useful and comprehensive data. The available medical information should include not only data about a current patient visit to the system, but also data from previous visits of that patient.

If such a machine readable file, containing variable length, variable format records, each directly accessible, is available along with the computer software to support it, then various medical applications can be modularly designed, programmed and implemented. Each application or research project will immediately have the benefit of access to all the other stored medical information about each patient.

A file structure fulfilling the above requirements was a major development of the MFCS. All PCMR's are individually and directly retrievable by each Plan member's medical record number. The PCMR is a continuing, integrated record designed to store all forms of essential medical data for all office and hospital visits for the lifetime of each patient. Each PCMR within the computer direct access storage facility is maintained and moved together as a continuous string of data; i.e., it is never split, fractionated, or otherwise separated for overflow or other reasons.

The basic form of a PCMR is a tree structure. The PCMR tree progresses through levels of storage beginning at Level 0. A maximum of twelve levels (Level 0 to Level 11) is allowed although only eight levels have been necessary for the various kinds of medical data stored to date.

There are two kinds of data in a PCMR: 1.) patient data received as input, and 2.) program-generated data relating to the tree structure of the record itself. The latter data includes branching, level, and various record length information to give the PCMR processing programs a trail through the tree.

The beginning zero level of a PCMR is the *Medical Record Index* which contains the patient's medical record number, certain program-generated data, and a field of 50 summary bits. The following summary bits are indicators which signal the existence of selected classes of medical data within the record:

Bit No.		Bit No.	
0	PART 0	17	Bacteriology
1	PART 1	18	Urinalyses
2	PART 2	19	Hematology
3	PART 3	20	PART 2 Other
4	PART 4	21	X-ray Contrast Study
5	PART 5	22	X-ray Chest
6	PART 6	23	X-ray Other
7	PART 7	24	ECG
8	PART 8	25	Surgical Pathology
9	PART 9	26	Autopsy
10	PART 10	27	Cytology
11	PART 11	28	PART 3 Other
12-15	(Reserved for Computer Use)	29	Drugs
		30	Operative Procedure
16	Blood Chemistry	31	PART 8 Other
		32	Anthropometry

In practice, for example, summary bit 16 would be »on« if any chemistry result, such as serum glucose or serum cholesterol, existed in the record. The summary bits are used to decrease retrieval time for research requests across many patient records. If a research request, for example, requires serum glucose results over a set of records, the retrieval program can quickly eliminate those records that do not have the Medical Record Index chemistry summary bit 16.

Branching out from the Medical Record Index, the PCMR divides into three major sections: the Identification Section, the Administrative Section, and the Medical Data Section. Data for these sections are stored in defined areas, or Parts, of the appropriate section.

The *Identification Section* contains five parts, as follows:

Part 0 contains information labeling the patient himself. Examples are: name, social security number,

Part 1 contains information about the locations of the patient in time and space. Examples are: birth year, birth

month, place of birth, street address, city, zip code, telephone number,

Part 2 contains linkage information with respect to the patient's blood relatives. Examples are: Medical Record Numbers (MR#s) of parents, MR#s of siblings, MR#s of children, mother's maiden name,

Part 3 contains linkage information for the patient's non-blood relatives. Examples are: spouse MR#, step-parent MR#, adopted children MR#,

Part 4 contains identifying information about the patient with respect to categorical classifications. For example, Multiphasic Study Group Codes.

The *Administrative Section* contains only two parts. Because the MFCS is operated primarily for medical care and research and not for business and accounting purposes, this section contains only that administrative information important for patient service and research purposes:

Part 0 contains information relating to Health Plan membership status. Examples are: effective dates in Plan, termination dates, type of Plan coverage, status (subscriber, spouse, dependent), Plan group and sub-group, Medicare,

Part 1 contains information relating to Health Plan generated information not specific to membership status. An example is residence code.

The *Medical Data Section* is first divided into »visits« and the data in each visit is subdivided into parts, analogous to the way data is stored in the Identification and Administrative Sections.

The basic strategy for grouping of data within the Medical Data Section was developed as a consequence of the clinic visit concept. All the data generated at and following the patient's appearance at one registration point are stored together. Specifically, data from: a location (Oakland, San Francisco, etc.), a category of service (office appointment, inpatient, house call, etc.), a department (medical, pharmacy, laboratory, etc.), a given date and a given physician, will be grouped with other input data having identical collection point values. These groupings are called »Computer Defined Visits«.

The Medical Data System may have multiple entries in its first level which is called the »Year Index«. An entry exists for each year in which a patient has had at least one visit.

Visit index entries are located at level two in the PCMR. Each level two entry (one for each visit) locates the information that uniquely defines a single visit, namely, department, location, day, category of service and doctor identification. The time of each visit is also stored here as well as a summary bit table (like the one in the Medical Record Index), but only for data that pertain to this single visit. The Medical Record Index summary bit table is the logical »OR« of all the separate visit summary bit tables in the record.

Under each visit index entry is a unique parts index which contains computer-generated information only, and is used exclusively by the processing programs to point to existing parts that are stored below level three within each visit. All medical data is then stored at level four or below.

Within the Medical Data Section, the parts represent categories of medical information for each visit, as follows:
Part 0 will contain any information reported by any means (word, phone, correspondence, etc.) from any source outside of the Kaiser-Permanente medical entities. Examples include the following: information from patients, referring doctors, schools, agencies, lawyers, etc.; all parts of the classical medical history such as chief complaint, present illness, past history, family history, occupational history, etc.; laboratory results when reported from an outside laboratory,

Part 1 includes any and all observations currently made by Permanente physicians on either the patient as a whole, or that portion of the patient which the physician customarily and usually examines. Examples of data in this part include: results of the physical examination; results of those 'tests' customarily done on the patient by physicians, such as Romberg and Weber tests; observations of MD physiatrists (but not physical therapists); observations of neurologists and the office tests they perform; blood pressures taken by physicians, observations of anesthesiologists,

Part 2 contains the results of all tests of body fluids (urine, blood, cerebrospinal fluid, bone marrow examinations, etc.) currently customarily performed in the laboratory, usually by technologists or automated equipment,

Part 3 contains the results of a specified list of observations generally comprising those tests done by physicians or technologists on the body as a whole for the purpose of testing a specific organ or organ system. This list will include, but not be limited to: radiological examinations, all observations of pathologists, EEG, ECG, EMG, photomotogram, thermography, anthropometry, etc.,

Part 4 will contain, but not be limited to, those observations not categorized above. Nurses' notes and other observations of paramedical personnel will be included here;

Part 5 will contain provisional diagnoses and impressions and other notes by physicians, intended as reminders, but not to be considered firm diagnoses at the time. These impressions are to be of the 'consider' or 'rule-out' variety and are not to be included in insurance reports. Impressions by any PMG physician, including radiologists, are to be stored here;

Part 6 includes 'firm' diagnoses at the time, made by physicians, including pathologists' diagnoses;

Part 7 will contain prognoses and information consisting of estimates of future events. This includes rehabilitation potential;

Part 8 includes information concerning therapeutic procedures, and diagnostic studies involving drug administration, either ordered and/or performed, by physicians, nurses and other personnel. It does not include referral for consultation. (Includes drugs, operative procedures, diet, appliances, occupational therapy, physical therapy, etc.);

Part 9 includes all recommendations not primarily therapeutic in themselves and not included in the treatment section. For example, return appointments, referral to another physician or optometrist, ordering of tests and physician suggestions.

Parts 10 through 14 are reserved for future medical uses.

Because the PCMR is a variable-length, variable-format record, it is necessary to store the identification and format of the data as well as the value of the data itself. Therefore, each datum is considered to be composed of three parameters: item identification, format, and value. An item is any variable such as a test, a procedure, a medical question, etc. The value expresses the result of the item; e.g., numeric value such as 250, English value such as »yes«, »hiatus hernia«, »test not done«, etc. The format specifies how the value is expressed; i.e., field length and type of element (integer, floating point, characters, text, etc.).

In January, 1970 a retrieval search of the entire data file was made. The current status of the PCMR file is summarized in Tables 1 and 2.

Table 1: Distribution of PCMRs by Record Length (Bytes)

Length Interval	No. of Records
0 — 499	974195
500 — 999	18966
1000 — 1499	30801
1500 — 1999	26400
2000 — 2499	3481
2500 — 2999	3481
3000 — 3499	982
3500 — 3999	207
4000 — 4499	89
4500 — 4999	33
5000 — 5499	18
5500 — 5999	3
6000 — 6499	4
6500 — 6999	1

RELIABILITY AND CONTROL REQUIREMENTS

Reliability

It is an important goal of our MIS that it must perform with a reliability approaching 100 percent. That is, a physician must be able to enter and retrieve patient data virtually any time of day, seven days a week. It is therefore essential to have a proper mix of backup equipment, modules, parts and corrective maintenance capability to maintain operations despite failure and breakdown. A preventive maintenance program is also required. Alternative backup procedures are being implemented in order to maximize continual system operation.

It is required that the system provide backup capabilities for both hardware and software. Thus the central computer equipment is duplicated and, by manual switching, one unit provides backup in case of failure of the other. Similarly, it is planned that the peripheral terminal equipment will be duplicated so as eventually to provide »fail-soft« operation. Remote pharmacy terminals currently use two independent telephone lines to the central computer so that, if one telephone line is interrupted, at least one-half of the terminals remain active at each user station. Visual display terminals are »plug-in« movable modules, so that any one can be replaced by a standby terminal when necessary. A duplex system also provides a means for programming and testing new applications without materially compromising service to the professional user. While such redundancy is costly, it is considered essential at this time.

Software reliability appears to be a more formidable problem at present. Periodic degradation of system and/or applications software performance can be expected. Provisions for rapid error recovery, once faulty performance is recognized, and reasonably stress-free backup support for the human users of the medical data system are vital for successful implementation and operation.

Backup records of data transactions are logged on magnetic tape and/or disc storage at critical points in the

Table 2: Distribution of PCMRs by Number of Visits

No. of Visits	No. of PCMRs	Total Lenght (Bytes)
0	911118	113634304
1	82762	73662656
2	23297	19287296
3	12503	9082080
4	8294	6940812
5	5491	5169164
6	3874	4064000
7	2758	3153560
8	2052	2580972
9	1497	2006956
10	1159	1694852
11	848	1312432
12	670	1124500
13	472	862628
14	368	704792
15	287	562924
16	261	536644
17	164	354260
18	138	331632
19	112	264628
20	113	290400
21	62	173868
22	71	201452
23	48	129848
24	40	114336
25	31	79392
26	23	73820
27	19	61076
28	22	70468
29	19	66640
30	11	37524
31	4	14644
32	14	49076
33	10	35188
34	7	29688
35	1	3596
36	6	23528
37	2	8888
38	2	6840
39	5	19936
40	6	27328
41	4	20172
42	3	13408
44	1	2836
47	5	17232
48	1	5400
49	1	5140
50	1	5380
54	1	6432
55	1	5716
63	1	5448
67	1	6660

flow of data so that, in case of failure, data are not lost and the patient computer record can be updated, or even reconstructed if necessary. It is a common experience that most down-time occurs when equipment or programs are replaced or modified; at such times one must be especially alert to the potential need for the backup and recovery systems.

A non-interruptible power supply is employed in case of electric power failure. The system contains four major components: a constant voltage battery charger; storage batteries; static inverter, and a diesel engine generator. The alternating current (AC) line supplies power to the static battery charger which »float« charges the battery, and at the same time supplies direct current power to the static inverter; this in turn supplies AC power to the AC load. If normal AC power failure occurs, the battery continues to supply power to the inverter in order to sustain the AC load without interruption. Because of the limited time the battery can power the total hospital data system, the battery is supplemented by the diesel engine generator. When normal AC supply is restored, the load is automatically transferred back to normal power, again without interruption.

It is planned that the air conditioning system for the central and peripheral processors will comprise at least two independent units. Thus, if one unit fails, the second is capable of fulfilling minimal cooling to maintain temperature control at a level that will prevent computer failure.

Data Quality Control

Much of the quality control for MIS involves editing of incoming data for validity at the input level (e.g., automatically checking a patient's identifying data against his existing computer record at time of data input or hospital admission, checking validity limits of laboratory tests against standard definition tables at the time of entry of test value, etc.).

On-line data entry permits immediate error detection and correction at source. An error detected at a later time is usually difficult to correct when the professional user who generated the input no longer has readily available the relevant source data. Programmed error surveillance procedures are being developed to scan data at input for missing or logically invalid data and notify the operator and/or user when such errors are found.

A prime concept of our MIS, therefore, is that all data should be entered directly into the computer from source. Avoiding the use of intermediary personnel should decrease errors and information loss. Physicians will enter their medical orders directly into the computer; the use of clerk technicians to enter into the system the medical orders handwritten by doctors, is merely to mechanize a traditional manual operation.

Similarly, pharmacists enter prescriptions directly into the computer and verify the label printout; they do not work through clerk typists. Physicians will enter the great majority of their diagnoses directly into the computer by selecting the best fit from a structured terminology; this avoids coding by medical record librarians.

Patient Data Confidentiality

All data within the patient computer record are subject to the same regulations governing their privacy and confidentiality as are data within our hospital record room. This requires controls for protection of the computer records to specified degrees of user-imposed privacy; such controls require that a physician identify himself, so that a specific patient's data will be released only to the physicians responsible for the care of that patient. Psychiatric data require additional special controls so as to be released only to or on the order of the specific psychiatrist responsible for the patient. Research data require identification of the principal investigator; epidemiologic research on groups of patients require maintenance of the privacy of individual patients; thus medical data is distributed only in aggregate form, without individual patient identification.

References

(1) Collen, M. F.: The multitest laboratory in health care of the future. Hospitals *41*: 119—123, 1967.

(2) Davis, L. S., Collen, M. F., Rubin, L., Van Brunt, E. E.: Computer stored medical record. Comp. Biomed. Res. *1*: 452—469, 1968.

(3) Greenes, R. A., Pappalardo, A. N., Marble, C. W., Barnett, G. O.: Design and implementation of a clinical data management system. Comp. Biomed. Res. *2*: 469—485, 1969.

(4) Jydstrup, R. A. and Gross, M. J.: Cost of information handling in hospitals. Health Serv. Res. *1*: 235—271, 1966.

(5) Richart, R.: Evaluation of hospital expenditures in patient care and communication. Comp. Biomed. Res. (in press).

(6) Van Brunt, E. E., Collen, M. F., Davis, L. S., Besag, E., Singer, S. J.: A pilot data system for a medical center. Proc. IEEE *57*: 1934—1940, 1969.

Addresses of the authors: Edmund E. Van Brunt, M.D.; Lou S. Davis, M.A.; Joseph F. Terdiman, M.D., Ph.D.; Sam J. Singer, B. Chem., M.D.; Eric Besag, B.E.E.; Morris F. Collen, B.E.E., M.D.; The Permanente Medical Group and Kaiser Foundation Research Institute, 3779 Piedmont Avenue, Oakland/Calif., 94611.