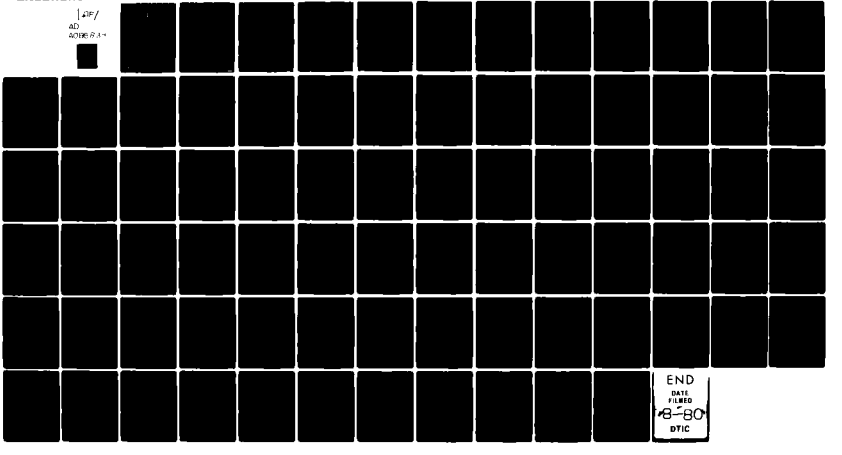
AD-A086 838 STANFORD UNIV CA DEPT OF COMPUTER SCIENCE F/G 9/2
DATABASES IN HEALTHCARE.(U)
MAR 80 G WIEDERHOLD N00039-80-G-0132
UNCLASSIFIED STAN-CS-80-790 NL



Stanford Computer Science Department
Report No. STAN-CS-80-790

March 1980

LEVEL II

DATABASES IN HEALTHCARE

by

Gio Wiederhold

AD A086838

Research sponsored by

National Institutes of Health

DTIC
ELECTE
JUL 18 1980
S D
C

COMPUTER SCIENCE DEPARTMENT
Stanford University

DDC FILE COPY

LELAND STANFORD JUNIOR UNIVERSITY
ORGANIZED 1891

This document has been approved for public release and sale; its distribution is unlimited.

80 7 16 009

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: STAN-CS-80-790

2. GOVT ACCESSION NO.: AD-A086 838

3. RECIPIENT'S CATALOG NUMBER:

4. TITLE (and Subtitle): Databases In Healthcare

5. TYPE OF REPORT & PERIOD COVERED: technical

6. PERFORMING ORG. REPORT NUMBER: STAN-CS-80-790

7. AUTHOR(s): Gio Wiederhold

8. CONTRACT OR GRANT NUMBER(s): ARPA/N00039-80-G-0132

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
Department of Computer Science
Stanford University
Stanford, California 94305 USA

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:

11. CONTROLLING OFFICE NAME AND ADDRESS:
Defense Advanced Research Projects Agency
Information Processing Techniques Office
1400 Wilson Avenue, Arlington, Virginia 22209

12. REPORT DATE: Mar 1980

13. NO. OF PAGES: 76 (12 79)

14. MONITORING AGENCY NAME & ADDRESS (if diff. from Controlling Office):
Mr. Philip Surra, Resident Representative
Office of Naval Research, Durand 165
Stanford University

15. SECURITY CLASS. (of this report): Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report defines database design and implementation technology as applicable to healthcare. The relationship of technology to various healthcare settings is explored, and the effectiveness on healthcare costs, quality and access is evaluated. A summary of relevant development directions is included.

Detailed examples of 5 typical applications (public health, clinical trials, clinical research, ambulatory care, and hospitals) are appended. There is an extended bibliography.

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

DATABASES IN HEALTHCARE

Gio Wiederhold

Stanford University
Computer Science Department

March 1, 1979

CS Report STAN CS80-790

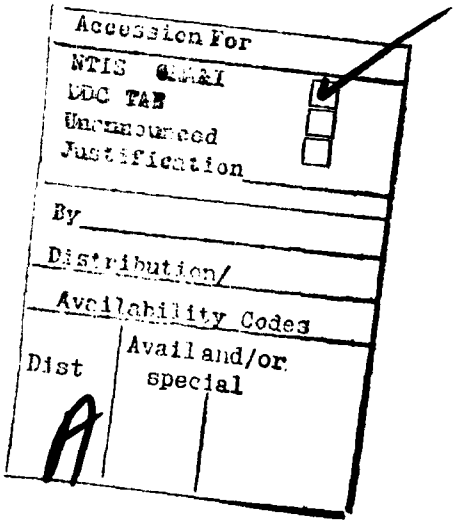
This paper was prepared for a Compendium on Computers in Health Care, Dr. D.A.B. Lindberg, Univ. of Missouri, Editor. The resources provided by the SUMEX Facility, supported by grant NIH RR-00785 were essential to the preparation of this document. Some of the work is based on research carried out as part of the RX project, sponsored by the National Center for Health Services Research under grant no. 1RO3HS93650. Basic research in Database Design leading to some of the concepts presented here is supported by ARPA through the KBMS project at Stanford University.

DATABASES IN HEALTHCARE

ABSTRACT

This report defines database design and implementation technology as applicable to healthcare. The relationship of technology to various healthcare settings is explored, and the effectiveness on healthcare costs, quality and access is evaluated. A summary of relevant development directions is included.

Detailed examples of 5 typical applications (public health, clinical trials, clinical research, ambulatory care, and hospitals) are appended. There is an extended bibliography.



CONTENTS

	page
I. Definition of the Technology	4
A. Databases and Their Objectives	4
B. Terminology in the Area of Data Bases.	5
1. Components of databases	5
2. File management systems versus database management systems	6
3. Related Systems	8
C. Scientific Basis for Database Technology	9
1. The schema	9
2. The data model	11
3. Types of data models	12
D. Database Operation	13
1. Entering data into the database	13
2. Data storage	15
3. Data organization for retrieval	15
4. Data presentation	17
5. Database administration	18
II. Use of Databases in Health Care	19
A. Health Care Settings and the Relevancy of Database Technology	19
1. Private solo practice	19
2. Group practice	19
3. Specialty practice	19
4. Hospitals	20
5. Clinical research	20
6. Non-patient databases	21
B. Current Health Care Applications of Databases	22
1. Reimbursement databases	22
2. Disease-specific shared databases	22
3. Databases used in HMO's	24
4. Surveillance databases	25
5. Speciality clinical databases	26
6. General clinical databases	27
7. Databases in research	28
C. The Future Use of Databases in Health Care	30
1. Cost-effectiveness issues	30
2. Initiatives and innovation due to technology push	30
3. The human element	31
4. Sharing of information	32
5. Privacy in Databases	33
6. Missing data	34
7. Problems of current interest	34
D. The Effect of Databases on Health Care Costs, Quality, and Access	36

III. State of the Art of Database Technology in Health Care 38

A. Systems in Research or Development Status 38

B. Industrial Status 39

1. Medical database systems 39
2. General database systems that are applicable to health care 48

C. Current Directions of Development 41

IV. Appendix : Examples of Current Databases in Health Care 42

A. Public Health :
CCPDS at the Fred Hutchinson Cancer Center in Seattle 42

B. Randomized Controlled Clinical Trials :
ECOG and RTOG at the Harvard Univ. Dept. of Public Health 45

C. Clinical Research :
TOD and ARAMIS at Stanford University 48

D. HMO Support:
COSTAR and the Harvard Community Health Service 52

E. Hospital Systems :
POMCS at Coral Gables Variety Childrens 54

V. References 63

DATABASES IN HEALTH CARE

I. DEFINITION OF THE TECHNOLOGY

In this chapter we will introduce the concepts of database technology in a way that will make it easy to relate the terminology to problems in health care. After the objectives have been defined the major components of databases and their function will be discussed. The remainder of this chapter will present the scientific and the operational issues associated with databases.

I.A Databases and Their Objectives

A database is a collection of related data, which are organized so that useful information may be extracted. The effectiveness of databases derives from the fact that from one single, comprehensive database much of the information relevant to a variety of organizational purposes may be obtained. In health care the same database may be used by medical personnel for patient care recording, for surveillance of patient status, and for treatment advice; it may be used by researchers in assessing the effectiveness of drugs and clinical procedures; and it can be used by administrative personnel in cost accounting and by management for the planning of service facilities.

The fact that data are shared promotes consistency of information for decision-making and reduces duplicate data collection. A major benefit of databases in health care is due to the application of the information to the management of services and the allocation of resources needed for those services, but communication through the shared information among health care providers, and the validation of medical care hypotheses from observations on patients are also significant.

The contents and the description of a database has to be carefully managed in order to provide for this wide range of services, so that some degree of formal data management is implied when we speak of databases. The formalization, and the large data quantity implied in effective database operations make computerization of the database function essential; in fact, much of the incentive for early [Bush45] and current computing technology [Barsam79] is due to the demands made by information processing needs.

Hence, the notion of a database encompasses the data themselves, the hardware used to store the data, and the software used to manipulate the data. When the database is used for multiple purposes we find also an administration which controls and assigns the resources needed to maintain the data collection and permit the generation of information.

We will in the next section define the technical scope of databases. The remaining sections in this chapter will deal specifically with current and future applications of databases in health care.

I.B Terminology in the Area of Data Bases.

Within the scope of databases are a number of concepts, which are easily confused with each other. The objective of a database is to provide information, but not all systems that provide information are databases. We will first define the term 'database', and then some terms that describe aspects of database technology. In the section which follows we will present types of systems which are related or similar to databases, but are not considerd databases within this review.

> A database is a collection of related data, with facilities that process these data to yield information.

A database system facilitates the collection, organization, storage, and processing of data. The processing of data from many sources can provide information that would not have been available before the data were combined into a database. Hence, a collection of data is not by itself a database, a system that supports data storage is not neccessarily a database system, and not all the information provided by computer systems is produced from databases.

I.B.1 Components of databases

A database is hence composed both of data, and of programs or software to enter and manipulate the data. Both data and software are stored within the computers which support the database, and the internal organization may not be obvious to the users. We will now describe some of the components that are part of database software. Databases require the availability of certain technological tools, or software subsystems. Some of these tools, that are used to support databases can also be used independently, and hence they are at times confused with the database system itself. Important subsystems are:

a) File Storage Systems : software to allocate and manage space for data kept on large computer storage devices, such as disks or tapes.
b) File Access Methods : software to rapidly access and update data stored on those devices.
c) Data Description Languages : means to describe data so that users and machines can refer to data elements and aggregations of similar data elements conveniently and unambigously.
d) Data Manipulation Languages : programs to allow the user to retrieve and process data conveniently.

In a database these subsystems have to be well integrated, so that the data manipulation can be carried out in response to the vocabulary used in the data descriptions. Storage is allocated and rearranged as new data enter the database, and access to old and new data is provided as needed for manipulation. To provide the neccessary reliability some redundant backup data is stored separatly and appropriatly identified whenever the database is changed. Optional software components of a database may provide on-line, conversational access to the database, help with the formulation of statistical queries, and provide printed reports on a regular schedule.

I.B.2 File Management Systems versus Database Management Systems

Of primary concern to a database effort is the reliable operation of the devices used to store the data over long periods of time. The programming systems which provide such services, typically inclusive of the tools listed in a) and b) above, are called file management systems (FMS).

When data are to be organized so that they can be accessed by a variety of users, system control extending to the individual users, and to the specific data units which these users will be referencing, may be needed. Control over the data and its use can only be achieved if all users access the database always via programs that will protect the reliability, privacy, and integrity of the database. We achieve reliability when data are not lost due to hardware and software errors. We protect privacy when we guarantee that only authorized access will occur. We define integrity as freedom from errors that could be introduced by simultaneous use of the database by users that may update its contents. A database management system (DBMS) should provide all the required database support programs, including management of files, scheduling of user programs, database manipulation, and recovery from errors. All these should form a well integrated package.

Not every database is managed by a commercial DBMS. Database support can also be provided by programs that use one of the available file management systems. The contents of the database can be identical for a system using a generalized DBMS product or one using programs written specifically for the task. A locally developed collection of programs rarely has the all of the protective features that are desirable when multiple users interact with the database from terminals. The manner in which users gain access will always depend on the choice of the DBMS or the file management system. For instance, a file system does not provide automatic scheduling of user requested activities. Without a DBMS the users will have to schedule their own activities in such a way that simultaneous data entry is avoided. Some file systems will simply disallow such access, in other systems such usage could lead to inconsistent data. If data entry activities are organized so that such conflicts are avoided then there is less need for the complexity of a DBMS. A very popular file management system in medicine is MUMPS, developed at Massachusetts General Hospital to support clinical use of relatively small computer systems [Bowie77].

Both file management systems (FMS) and database management systems (DBMS) are available commercially for most computers. Some DBMS's will make use of an existing FMS, others will perform all but the most primitive file access functions themselves. Since a DBMS interacts closely with the user of the database, we find that distinct types of DBMS's have been developed. DBMS's also differ in terms of the comprehensiveness of software services. Most manufacturers provide an FMS at no additional cost, but acquisition of a DBMS is rarely free.

The choice of a particular type of database management system will influence the structure of the future database. Not every type of DBMS will be available on a given computer, but for most medium to large computers there is some choice. Simplicity versus generality and cost are often a trade-off. Even so-called generalized database management systems impose, to a great extent, the view of the designer or sponsor of such a DBMS. Many of the major systems now being marketed were designed to solve the complexities of specific applications. We hence find DBMS's that excel in inventory management, some do excellent retrieval of bibliographic citations, others

have a strong bias towards statistical processing. Even within the medical area different DBMS's will emphasize one of the many objectives that are found within the range from patient care to medical research. The following table will list some database systems found in medicine with an indication of their objective. We distinguish in this table: general ambulatory patient care, clinical or speciality outpatient care, hospital inpatient care, or patient management and record keeping in these areas. Clinical studies refers to research data collection on defined populations. Guidance refers to the giving of medical advice during the inquiry process. Details of these tyes of application are given in chapter 2 of this review. The types of database organizations can be categorized as tabular, relational, hierarchial, or network. These terms will be defined in section B.4 of this chapter.

Name	Application	Type	FMS used	Computers	Reference
CCSS	Clin.Studies	Tab. DBMS	Seq. files	various	Kronma78
CIS	Med.Guidance	Tab. DBMS	DEC seq.files	DEC11	McDona77
CLINFO	Clin.Studies	Tab. DBMS	DG ISAM	DG Eclipse	Groner78
COSTAR	Amb.Pat.Rec.	Hier.DBMS	MUMPS	DEC 11	Barnet79
GEMISCH	Clin.Recs.	Hier.DBMS	DEC-11 DOS	DEC 11	Hammon78
GMDB	Clin.Studies	Tab. DBMS	IBM VSAM	IBM 370	Wirtsc78
FAME	Clin.Studies	Hier.DBMS	CDC SIS	CDC Cyber	Brown78
IDMS	Pat.Mgmt.	Netw.DBMS	Basic Access	IBM 370	Penick75
IMS	Hosp.Recs.	Multi-hierarch.DBMS	IBM VSAM	IBM 370	Sauter76
LIM	Regional Rec	Hier.DBMS	IBM DL/1(IMS)	IBM 360	Jainz76
MIDAS	Regional Rec	Hier. FMS	Direct files	Univac494	Fenna78
MISAR	Res.Data	Tab. DBMS	MUMPS	DEC 15	Karpin71
MUMPS	Med.Records	Hier. FMS	self.cont.	DEC 15,11 DG and more	Barnet76 Bowie77
OCIS	Clin.Recs.	Mult.Hier.	MUMPS-11	DEC 11-70	Blum79
TOD	Clin.Recs.	Tab. DBMS	PL/1 ind.seq.	IBM 370	Wieder75
PROMIS	Hosp.Recs.& Med.Guidance	Tab. FMS	direct files	Univac V77-600	Schult79
RISS	Hosp.Recs.	Relat.DBMS	RTS11 ind.seq	DEC 11	Meldma78

Database and File Management Systems Found in Health Care

I.B.3 Related systems

Data are collected and stored into a database with the expectation that at a later time the data can be analyzed, conclusions can be drawn, and that the information obtained can be used to influence future actions. Information is generated from data through processing, and should increase the knowledge of the receiver of this information. This person then should have the means to act upon the information, perhaps to the benefit of a larger community.

|| The production of information is the central objective of a database. ||

There are other automated information processing systems which are not considerd databases, although they may share some of the technology. In the remainder of this section two categories of such related systems will be presented.

INFORMATION SYSTEMS store information - often the output of earlier data analyses - for rapid selective retrieval [Beckle77]. A well known example is the MEDLARS system [Katter75, Leiter77], a service of the National Library of Medicine, which provides access to papers published in the medical literature. The task of such an information system is the selection and retrieval of information, but not the generation of information [Lucas78]. Index Medicus for instance only provides the references, and depends on the user's own library [Kunz79]. Even maintenance of personal reference files can be effectively automated [Reiche68]. The benefits are due to the speed and improved coverage with which the documents can be found.

The boundary between information systems and database systems is not at all absolute. One can perhaps even speak of a spectrum of system types. When the queries are simple the two system types are in fact indistinguishable. Retrieval of the age of a patient, for instance, can be carried out with equal facility on either type of system. But when another observation, say cholesterol level, has to be compared with the average cholesterol level for all other patients of the same age, then a computation to generate this information is needed, and a system which is able to do this is placed more on the database side of the spectrum.

DECISION SUPPORT SYSTEMS assist with the manipulation of data supplied by the user [Davis78]. The help may be principally algorithmic - perhaps assuring that Bayes' rule is properly applied. More specialized systems embody medical knowledge [Johnso79], for instance in acid-base balance assessment [Bleich72] and anti-microbial therapy [Yu79]. While these systems could be coupled to databases, so that they become also knowledgeable about a specific patient, today they are typically separate [Gabrie78]. Work in in decision making for health care cost control has indicated a need for database facilities in these applications [BrookW76].

The HELP system, at the LDS hospital in Salt Lake City, does keep a separate file of clinical decision criteria and applies them to the patient database as it is updated. The system then advises the physician to consider certain actions or further diagnostic tests [Warner78]. As medical databases become more reliable and comprehensive we can envisage increased exploitation of the information contained in them by systems which embody medical knowledge.

I.C The Scientific Basis for Database Technology

The emergence of databases is not so much due to particular inventions, but is a logical step in the natural development of computing technology. The evolution of computational power began with the achievement of adequate reliability of complex electronic devices. The mean-time-to-failure reached several hours for powerful computers about 1955. At that point the concerns moved to the development of programming languages, so that programs of reasonable power could be written. These programs had the capacity to process large quantities of data, and in the early sixties magnetic tape and disk devices were developed to make the data available. Operating systems to allocate storage and processing power to the programs became the next challenge. By the late sixties these systems had matured so that multi-user operation became the norm. As these foundations were laid it became feasible to keep data available on-line, i.e., directly accessible by the computer system without manual intervention, like fetching and mounting computer tapes. Now a variety of application programs can use those data as needed. In current systems valuable data can be kept on-line over long periods without fear of loss or damage to the database.

I.C.1 The Schema

The one technical concept which is central to database management systems is the schema. A schema is a formalized description of the data that are contained in the database, available to the programs that wish to use the data. All data kept in such a database is identified with a name, say DOB for date-of-birth. With a schema it is sufficient for application programs to specify the name of the data they wish to retrieve. A command may state:

```
date-of-birth = GET ( current-patient, DOB ) ;
```

The database system will use the schema to match the name of the requested data. When a corresponding entry in the schema is found, the database system can use information associated with the entry to determine where the requested data have been stored, locate the data values, and retrieve them into the application program area (date-of-birth) for analysis or display. During this process it is possible to check that the requestor is authorized to access the data. The DBMS may also have to change the data into a representation that the program can handle [Feins78A]. Similar processes are carried out by the DBMS when old data are to be updated and when new data are to be added to the database.

The schema is established before any data can be placed into the database and embodies all the decisions that have been made about the contents and the structure of the database. Each individual type of data element will receive a reference name. The data to be kept under this name may be further defined. The most important specification is whether the data are numeric, a character string, or a code. Codes then need tables or programs for their definition. Other schema entries give the format and length of the data element, and perhaps the range of acceptable values. For observations of body temperature the five descriptors might be:

TEMP, temperature in degrees C, numeric, XX.X, 36.0 to 44.0.

The data elements so described will have to fit into a structure; a value by itself, say TEMP = 41.9, is of course meaningless. This data element belongs

in an observation record, and the observation record must contain other data elements, namely a patient identification (ID), a date, and a time. These data elements, which are used to identify the entity described in the record, constitute the ruling part; without these there is insufficient information present to make the TEMPerature observation useful. The ruling part data types (ID, DATE, TIME) will also appear in the schema.

The observation record may contain, in addition to TEMP, other dependent data elements as: the pulse rate, the blood pressure nd .the name of the observer. The entire observation record can then be described as a list of seven attributes, as follows:

Observations: ID, DATE, TIME > TEMP, PULSE, BP, OBSERVER;

The first three attributes form the ruling part, the other four are the dependent part; we seperate the two parts with a > symbol. Each attribute has associated with it a schema entry with the five descriptors shown for the TEMP entry above. There will be other kinds of records in the database: a patient demographic data record will exist in most databases we consider. Here the only data element in the ruling part will be the ID field. This record may be in part as below:

Patients: ID > PATIENT-NAME, ADDRESS, DOB, SEX, ... ;

Matching of the ID fields establishes the relationship between patient demographic data records and the observation records. The known relationships between record types should also be described in the schema, so that the use of the schema is simplified [Manach75, Chang076].

We use three types of connections to describe relationships between records [Wieder79], their use is also sketched in the figure below.

a) The Identity Connection - used where the ruling parts are similar, but different groupings are described;
for instance both hospital patients and diabetes clinic patients are patients with patient ID's, but may have different dependent data stored in their files.

b) The Reference Connection - used where there is a common descriptive record referred to by multiple data records;
for instance the physician seen, is a record type referred to from the patients clinic visit records.

c) The Nest Connection - used where there are many subsidiary records of some type which depend on a higher level record; multiple nest connections define an association;
for instance the multiple clinic visits of a specific patient, each with data on his temperature, blood pressure, etc. form a nest of the patient record.
An association occurs in the figure where a physician has admitting privileges at one or more hospitals, and each hospital grants admitting privileges to a number of physicians. The admitting-privileges file has as ruling part both the physician's ID and the hospital name, a dependent data element might be the date the privilege was granted.

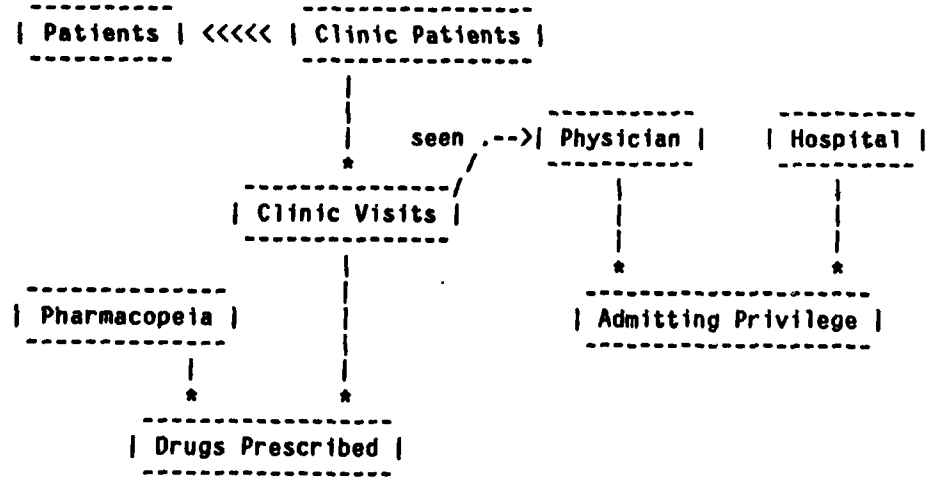
Associated with the connection types may be rules for the maintenance of

database integrity. Such rules can inform the database system that certain update operations are not permissible, since they would make the database inconsistent. For example we would not want to add a clinic patient without adding a corresponding record to the general patient file, if the patient did not yet exist there. Similarily deletion of a physicians record from the database implies deletion of the associated admitting privileges.

I.C.2 The Data Model

In order to provide guidance for the creator of the schema it is important to have design tools. A large database can contain many types of records, and even more relationships between the record types. These have to be understood and used by a variety of people: the programmers who devise data entry and analysis programs, the researchers who wish to explore the database in order to formulate or verify new hypotheses, and the planners who wish to use the data as basis for modelling so that they can predict the response to future actions. A variety of models exist [ACM76]; some models are abstractions of the facilities that certain types of database management systems can provide, other use more generalized, mathematical abstractions to represent the data and their relationships. Recent work in database research is directed towards improving the representation of the semantics of the data [Hammer78, ElMasr79, Codd79] so that the constraints of the relationships that exist in the real world can be used to verify the appropriateness of data that are entered into the database.

Any reasonable model of the database can provide a common ground for communication between users and implementors, without a model there is apt to be an excess of detail [Wieder78]. An example of a data model for a clinical database is shown below.



The nest (--*) connections indicate that there may be multiple inferior instances for each superior instance.
The reference (-->) connections indicate that there may be multiple references to each instance.
The identity (>>>) connection defines a subgroup.

I.C.3 Types of Database Models

A popular approach to database analysis distinguishes several categories of databases. Database system implementations can be associated with each category. These categories are represented by database model types, the best known types are the

Relational model- derived from the mathematical theory of relations and sets.
Hierarchical model- related to tree-shaped database implementations, similar to corporate organization diagrams.
Network model - permits interconnections that are more complex than hierarchies, based on a definition developed by a committee of specialists in commercial system languages.

The structural model can describe the structures of any of these three models, as well as of other database implementations. If only a single record-type - a box in the above diagram - is implemented then we are dealing with a 'universal relation' [Ullman79]. A single box for a complex database would have many columns and rows, and contain many null entries. If the data are organized into several record-types, each corresponding to some meaningful entity, then we are dealing with a 'tabular database'; if a completely general query and processing capability exists in such a system, we have implemented the 'relational model' [Codd70].

At this point the entities stand alone, and some analysis is needed to relate them. If any of the indicated connections have been implemented then we may have a network or a hierarchical database. In the hierarchical model a record-type may have only one nest connection (--*) pointing to it. The implementation of multiple nest connections, which creates a network with associations, is considerably more complex [Stoneb75, Wieder77]. Several of the larger commercial DBMS's are based on work by the Data Base Task Group of CODASYL, and do support such network structures [Olle78]. These systems often do not support the general inquiry capability of the relational model implementations.

It is important to note that there is a distinction between a model and its implementation. A model is an abstraction and provides a level of insight which can cut through masses of confusing detail. In the implementation this detail has to be considered. It is likely that the implementation will differ considerably from the model used to describe it. As more powerful models are developed this distinction may become greater. An implementation may then be best described in terms of transformations that are applied to the model which defines the database at a high conceptual level. Most transformations are done for reasons of operational performance and reliability.

I.D Database Operation

In the section above we have discussed the scientific basis of databases. In order to use and benefit from that science a database operation has to be established, and that involves many decisions of practical, but critical importance. This section will consider such topics.

When the database design has been established, and a suitable software system has been obtained, then data collection can commence. Data is often obtained partially from sources that were in existence before a database was considered. To complete the database, so it can serve the intended broad scope, new data collection points may have to be defined. The value of adding data to the database has to be considerd, since Data collection and entry is costly and susceptible to errors. We will begin with a discussion of issues in entering of data, and then proceed to data storage and organization concerns, discuss data presentation issues, and finish with some remarks about database administration.

I.D.1 Entering Data into the Database

The relatively high cost of data entry is a major concern. It is obvious that data that cost more to collect than they are worth should be avoided. When a certain data element is entered its utility is hard to predict: its usefulness may depend on its value, on the completeness of this patient's record, and on the patient's returning to the clinic, so that follow-up is possible. These factors are not easy to control. The actual problem of data acquisition can, however, be addressed. Much less formal attention has been given in the literature to this subject than to the topic of data retrieval [Greenf76].

When data are to be collected there are the costs of the actual collection, of the transcription to some processable form, and of the actual entry into a computer. The data collection is to a great extent the physician's task. While automated clinical instruments can collect objective values, and the patients themselves can enter their own history [Slack66], many subjective and important findings emanate from the physician.

It may be considered desirable to minimize changes to the traditional manner of medical data recording, so that the physicians continue to collect their findings as notes in free text or by dictation. These reports are then transcribed by clerical personnel into the computer. This format presents the medical information in a way that is least affected by mechanical restrictions. To enable retrieval of such observations the specific statements or paragraphs may be categorized into functional groups as findings, treatment, plans, etc. as proposed by [Korein71]. A system, based on these concepts has served well in a city hospital pediatric clinic setting. Of particular importance was that patient data retrieval for emergency and unscheduled visits became possible [Lyman76].

When textual data are to be used for analysis, we find that they are nearly impossible to process in the form they were entered. An immediate problem is that the natural language text has to be parsed so that its meaning can be extracted. Both the parsers and the associated dictionaries are substantial pieces of software. But even when language understanding is achieved, consistent data for entry may not have been obtained since medical terminology varies over time and among health care providers. In general some encoding is needed. It may then be of benefit both to the

physician and to the system to choose a method of data collection which encodes data immediately into a more rigorous form. Various choices exist to encode data :

1. The encoding can be carried out by clerical personnel [Valbon75].

2. Natural language, i.e. English text, may be analyzed and converted by a program that processes the text within the medical context [Pratt73, Okubu75]

3. A constrained set of keywords for data values, for example the list:
   (no, light, moderate, serious),
   can be attached to the schema entry for a specific data type. These data values will be converted on data entry to an internal code [Wieder75].

4. Where the number of possible data elements, for which data are to be collected, is large, the name of the data element, i.e. 'facial rash', may be encoded in addition to the data value itself [Hammon73, Wong78].

5. Keywords may be checked on a form or selected from a menu presented on a display screen [Schult76]. Selection can de accomplished using touch-sensitive screens, lightpens, cursors operated by joysticks or key-pads, or by entering on a keyboard a digit which refers to a line of he presented menu.

6. Where the list of keywords is too long for screen presentation a hierarchical menu selection can be provided or a subset of the keywords corresponding to a few initial letters can be displayed [Morgan77].

7. The forms or menus to be used for data collection may be generated using the schema of the database managment system [Hanley78].

With the continuing development of fast display technology the latter choices seem to have the most promise. The response for screen selection and presentation of the next menu has to be extremely rapid (0.4 sec. per screen is cited) to encourage direct physician use of the devices [Watson77]. Such speeds are very difficult to achieve today, since the display frames reside on remote disk storage devices and have to be fetched, formatted, and transmitted by file, user, and communication programs for presentation on terminals. When those terminals are connected via telephone lines to the computer another bottleneck appears. To transmit a display frame of 24 lines of 50 characters each, at the fastest available rate, 9600 bits/second, still requires one second. To cope with this problem either special communication lines or storage devices local to the terminal are needed.

Numeric values are not as easily entered on a touch-screen as are choices among discrete elements. Keyboard entry may continue to dominate this part of data entry, unless the values can be obtained directly from medical instrumentation. Typed data requires much editing. Comprehensive commands for specification of input editing are part of the MUMPS language, and have contributed greatly to its acceptance. Modern computer languages, as PASCAL, also provide within the variable declarations a capability to limit the range or the set of choices of values to be entered.

I.D.2 Data Storage

The cost of data storage is now much lower than cost of data entry. This means that if data entry is worthwhile, the entered data can be stored for a reasonably long time. The characteristics of medical record structures can, however, easily lead to a waste of computer storage space which is an order of magnitude greater than the actual data storage space needed. This occurs when data are stored in simple rectangular tables, since the variety of medical data requires many columns, but at one encounter only a few values wil be collected. Hierarchical file organizations allow linkage to a variable number of subsidiary data elements, and in this manner provide efficient storage utilization, whereas the older tabular files dealt poorly with medical data [Greene69]. The encoding techniques used for data entry can also provide compaction of stored data since short codes are used to denote long keywords.

Data structures can often be compressed by suitable data encoding techniques applied to the files. Especially unobserved data elements do not need actual storage space. Data compression can reduce both the storage requirements and the access times greatly [Wieder77]. When space considerations are no longer an important issue in data organization, an apparent tabular format can again be used, and this can simplify data analysis programs. In the clinical databank described in section IV.C, TOD, the compressed data, after encoding to account for missing, zero, or repeating data, occupied only 15% of the original storage space.

Older data often become less interesting, and can be moved to archival storage. Storage on magnetic tape is quite inexpensive and the data can be recovered, if needed for analysis, with a moderate delay. In a well-run operation data can be recovered from tapes for on-line access in about an hour [Soffee76]. The major problems are the development of effective criteria for selection of data for archival storage and the cataloging of archival data, so that they can be retrieved when needed. Candidates for archiving are detailed records of past hospitalizations and episodes of acute illnesses.

I.D.3 Data Organization for Retrieval

The important point in research usage of databases is that information is not produced by the retrieval and inspection of a few values, but rather from the relating of many findings in accordance with hypothesized cause and effect relationships. When the data files grow very large, repetitive scans for data selection may become prohibitively slow, especially during the data exploration phase. We distinguish the following phases in the research use of clinical databases:

1. Initial definition of the data to be collected, with consideration for clinical needs. The expected usefulness is often based on vague or ill-defined initial hypotheses.

2. Exploratory analysis, using tabulations and simple graphics in order to compare subsets of the population.

3. Hypothesis generation based on perceived patterns, definition of independent and dependent variables according to some clinical model.

4. Data validation and sometimes expansion of data collection in the areas in which patterns appear interesting.

5. Subset definition and generation so that differences due to the independent variables can be made explicit.

6. Exhaustive statistical analysis of the subsets to verify or refute the hypotheses.

It is important to have good subsetting facilities and efficient access to defined subsets. Such services are provided in many clinical systems, but the techniques vary widely. Often the subsets are extracted and manipulated as distinct databases [Mabry77]. In other systems a subset is kept as a collection of references to records in the main database [German75], and in yet another system the subset is recreated from the definition of the subset [Todd75].

Since access to data in research is primarily by attribute field rather than by patient record, it can be profitable to transpose the database [Wieder77]. Transposition generates one, possibly very long, record for each attribute of the database. Such a record now contains a sequence of values of this particular attribute for all patients or all visits of all patients. Many current computer systems cannot manage such long records easily but the benefits should be clear: to relate blood pressure results to dosage of an anti-hypertensive drug only two records have to be retrieved from the transposed file. In a conventional file organized by patient visit every visit record is accessed to retrieve the two fields needed to accomplish this comparision.

To avoid scanning an entire conventional file, access structures can be created which speed up the record selection process. Attributes which are expected to be used in record selection are entered into an auxiliary index file, which is then maintained in sorted order. If the attribute is "bloodpressure" all hypertensives will appear at the beginning of the corresponding index file. With every blood pressure value a reference pointer to the corresponding visit record will be kept. Now only the data records for patient visits where the blood pressure was high have to be retrieved.

Bitmaps provide a simplified form of indexing. Whereas an index is based on the actual data values, a bitmap uses simple categorizations of these values. In a list with entries which correspond to the records in the datafile a bit is set to one if the data values in the record meet a certain condition. This condition could be a blood pressure greater than 160/100 [Ragan78]. Both indexing and bitmaps can be viewed as providing the capability of preselection of relevant records. If the selection of indexes or bit map definitions matches the retrieval requests well, access to conventional files can become much faster. The maintenance of such access structures will of course require additional effort at the time of data entry.

There are many cases where more computation at the time of data entry can reduce the effort that required at data retrieval time. In some applications it may be known that certain computable results of the collected data will be needed at a later time. Then such results may actually be already computed and stored within the database when the source data are entered.

Typical of a precomputed or actual result is the maximal value of a clinical observation on a given patient, say blood pressure, which could be kept available so that no search through multiple visits is needed to identify a patient with evidence of hypertension [Melski78]. Other candidates for precomputation are totals, averages, or the range of values of a variable [Wieder75]. The total amount outstanding on a bill and the range of a diabetics blood-sugar level are other examples.

I.D.4 Data Presentation

Data from databases can be presented in the form of extensive reports for manual scanning, as summary tabulations, or as graphs to provide rapid visual comprehension of trends. An extensive data analysis may lead to a printout of statistical findings and their significance, or may provide clinical advice in terms of diagnosis or treatment. When simple facts are to be retrieved the results are apt to be compact and easy to display or print. If much computation is used to generate the information then presentation of the end-results alone is rarely acceptable. Most medical researchers will want an explantion of the data sources and algorithms that led to the output results, as well as information about the expected reliability of the final values.

These requirements increase the volume of the output for research studies, so that printed reports dominate. In most clinical situations less output is used, so that other methods may be practical. We have seen the following alternatives:

1. Detailed listings or rapid video presentations for quick scanning of data.

2. Cross tabulations or graphics to aid human pattern detection.

3. Well-structured summaries, with automatic data selection and advice for patient care.

4. Summaries with explanatory backup available on a terminal when needed.

5. Structured report presentation for outside distribution, as with billing or result publication.

During a routine patient encounter a paper summary is probably least distracting, but in emergency situations video terminal access can be much more rapid. Terminal access helps the researcher in the formulation of queries, and graphics provide insight to clinicians uncomfortable with long columns of numbers. As systems mature and become more accepted the user should be able to move smoothly from one form of output presentation to another, but most systems now in use do not provide many options for data presentation, and even fewer offer a smooth transition between interaction modes.

I.D.5 Database Administration

Even when all the right decisions have been made and a database exists, there has to be an ongoing concern with reliability, adaptation to changing institutional needs, planning for growth, and technical updating of the facilities. In many institutions a new function, that of database administrator, is defined to deal with these operational issues. The database administrator needs strong support from management and high quality technical assistance. Since the function is responsible for day-to-day operations it is not reasonable to expect a high level of innovation from the database administrator, but responsiveness to the institutional goals is essential.

II. USE OF DATABASES IN HEALTH CARE

Now that the concepts and operational issues of databases in health care have been summarized the application of databases in health care settings can be brought forward. We will first relate database uses with the categories of health care settings, and later in the chapter discuss specific types of applications in greater detail.

II.A Health Care Settings and the Relevancy of Database Technology

The obvious area of application of database technology in health care is the maintenance of patient records. These medical records exist in a wide range of health care settings, and effectiveness of databases for their management depends greatly on the environment [Wied&K78].

II.A.1 The Solo Practice

The private solo general practitioner finds that most of his needs for automation are satisfied by relatively simple systems oriented towards billing and schedule keeping. The cost of entering and keeping medical data in a computer does not now provide corresponding benefits, since a paper medical record can be kept close at hand [Castle74, Rodnic77]. The low cost of small processors does create great interest in computer applications by physicians, but current micro-processor systems do not yet provide a convenient basis for the development of programs with complex files [Zimmer79].

II.A.2. The Group Practice

The operation of a group practice, where several physicians and paramedical personnel cooperate in giving care, creates some problems in access to medical records. Data for the record is generated at multiple sites, but the entire record should be complete and legible whenever and wherever it is retrieved. Continuity of care can be greatly aided by a computer-based system [Bresla76]. Here entry and storage of basic medical data, diagnoses, procedures, prescriptions, and follow-up becomes worthwhile.

For multi-user operation a shared database, accessed from the individual health care sites, provides benefits in access to the record and to the data contained therein [Zimmer78]. A group may also wish to integrate its billing service, and the shared medical record can provide the required linkage [Worth78]. In a large group practice or health maintenance organization (HMO) the management benefits of an accessible clinical database are also considerable [Gaus73, Barnet79].

II.A.3 Specialty Practice

A specialty practice or clinic can further exploit the benefits of shared access to data. Since most specialty clinics deal with long-term or chronic diseases a longitudinal record can be collected on the patients, and such a record can reflect the individual's response to tests and treatments [Starme77]. In long term care the ratio of effort devoted towards diagnosis versus treatment decreases, so that more care can be delivered by paraprofessional personnel, and here well organized data presentations can be especially effective [McDona77].

The data management of a specialty practice will also be easier to structure. Whereas in a general practice a bewildering variety of data has to be accommodated, a specialty practice can often organize their data into standardized flowsheets [Fries74]. Such tabular data representations are not only simple to use, but are also significantly easier to process.

The depth of specialized experience, acquired in specialty practice, motivates the physicians to analyze the disease and treatment processes. The physicians are then willing to deal with the procedures that aid data quality maintenance, which are otherwise viewed as a distraction from patient care.

II.A.4 The Hospital

A hospital, dealing with in-patients, presents an entirely different set a problems. In the ambulatory settings discussed above, a patient is seen at most once per day and his record will be kept active in the clinic for several years. A stay in the hospital may oly last a few days, and during that period data entries and retrieval requests can occur within minutes of each other. The active time frame for inpatient services is hence much smaller than the time frame for outpatient services. Data input, processing, and output has to be rapid, but data are not retained in an active state over long periods. The benefits of a Hospital Information System hence are mainly due to the communication provided through a shared database [Watson77]. Rapid communication can lead to reduction in length of stay and minimization of redundant diagnostic orders. When the database is used like a blackboard, as a communication medium by by the treating physicians, it will reduce conflict in patient treatment procedures.

Hospital-based clinics for ambulatory patients have of course the features of general or specialty clinics, and impose their own requirements on a computer system that is to be shared for both hospital functions. Systems that serve both functions well are rare [Collen74, Austin78]. At the NIH Clinical Center, where the patients are referred for inclusion in particular research protocols, work is in progress to create a database system for long-term research and administrative purposes as a byproduct of the on-line hospital system [Lewis77].

II.A.5 Clinical Research

An important use of databases is the support of clinical studies [Palley75]. Both in a controlled clinical trial [Peto77, Silver78] and in open population studies [Fries72], careful management of data is essential. When the studies become even moderately large in terms of population and observation period, a database approach becomes essential. We find that even where no actual database management system is in use that well-defined, programmed or manual procedures are used to complement traditional computing facilities with database tools, such as data dictionaries, generalized query programs, periodic backup, and data verification processes. Without such care computerized medical records will not be trustworthy and not be the resource for research they appear to be [Feinst78].

II.A.6 Non-Patient Databases

There are of course many data collections relevant to medicine which do not contain patient data. Such non-patient databases are not discussed in depth in this report, but they constitute an important facet of database usage in health care. One type of non-patient databases are the data collections that are used to record, monitor, and assess the effect of new drugs on animals prior to their release for clinical trials on humans. Substantial databases exist that support toxicology studies [Oxman76]. The recent rules on 'Good Laboratory Practice' by the Bureau of Drugs of the FDA specifically address data handling requirements for toxicology studies [FDA78]. A databank, the Laratory Animal Data Bank (LADB), has been developed to keep track of laboratory animals used as controls for such studies. This work is done at Batelle Laboratories under sponsorship of a number of agencies concerned with such toxicological testing. The National Library of Medicine manages the program as one of its Specialized Information Services [BCL78].

Databases are also used to improve the management of health care education. Such databases contain student and laboratory data, so that students can receive appropriate assignments, are matched to the patient population, and show adequate progress [Duncan78, Go78, Kreitz78].

Another set of non-patient databases used in health care are collections of reference information for physicians or researchers. Well known are the poison-control centers; these are databases that relate accidents involving dangerous household and industrial substances to appropriate treatment [Yoke178, ODe1178]. The organization of these systems approaches that of information retrieval systems, rather than of databases, since mainly retrieval, rather than data-processing and data analysis, takes place in response to an inquiry.

II.B. Current Health Care Applications of Databases

The use of databases implies the availability of large and relevant quantities of data. A sufficiently large, complete, and accurate collection of data is the basis for achieving believable results from database processing. Since entry of large data volumes is costly, the initial uses of databases have been in areas where the data were simple to obtain and of relatively great value. Databases have been used in the financial areas of health care delivery before we they were used in the clinical areas, and databases are used more often to support specific studies than in general health care delivery situations [OTA77, Stei&B78]. To provide information for public health policy decisions data may be entered initially at high levels of abstraction before the systems are able to compute summarized data from detailed data collections. The problem of data entry has been discussed earlier.

The effort needed to bring database projects into operation, and the inherent delays before results can be produced means that many current database oriented projects are difficult to evaluate. We find broad acceptance today of database technology in the areas where databases were introduced early, but also that some of the pioneering projects suffer today from being bound to outdated database technologies [Brian79]. There appears to be no fundamental barrier to acceptance of databases in medicine, but there is a well founded "show-me" attitude which can be overcome by demonstration of adequate operation and reasonable effectiveness.

II.B.1 Databases Used for Service Reimbursement

The largest databases in use that are related to health care are no doubt the databases associated with the reimbursement mechanisms for health services. In the United States these are the federal Medicare and the state managed Medicaid programs. The latter are generally served by outside private contractors; since the contracts are awarded on the basis of lowest cost the contractors have kept the medical content of the databases as limited as possible. Several other countries maintain databases associated with government sponsored health care delivery [Anders77, Hall77, Nakaya77, Reiche77, Sheple77]. Requirements for inquiry and audit are generating requirements for more complete medical encounter information. This leads the processing organizations which handle reimbursement accounting to consider database technology, although much of the work today is based on periodic processing of large files [KatzjR77]. That these databases can provide useful information for health care delivery policy has been demonstrated by the on-line Medicaid prescription collection system in Alabama [MeselW76] where inappropriate use of several drugs was demonstrated. This was one of the findings which lead eventually to the recent implementation of restraints of prescribing of propoxyphene hydrochloride, Darvon.

II.B.2 Disease-Specific Shared Databases

There is also a broad interest in the capture of population data for those diseases that are so prevalent and costly that national concern, sometimes in the form of disease-specific legislation, is focused on them. In the use of these databases issues of policy and health care delivery are intertwined.

Disease specific databases with important health policy implications have been supported by the National Cancer Institute (NCI) [Haensz66]. One program (SEER) [NCI74] provides surveillance of cancer incidence and survival in nine geographic areas around the country [Young78]. A more recent program, CCPDS, integrates the data collected at the comprehensive cancer centers to evaluate the impact of treatment on disease patient groups [Feigl79]. Similar broad-based efforts exist in psychiatry (MSIS) [Logema73], kidney disease (RENTRAN) [Mishe176], and rheumatic disease (ARAMIS) [Hess74]. The CCPDS database is here described as an example of this type in the Appendix, section IV.A.

Collecting and sharing of data from multiple institutions requires a major effort to standardize data collection. In particular, the consistent encoding of observations, already difficult within one institution, becomes a major problem when the data are collected from multiple institutions. When the data sources are separated by a day's travel the difficulties are even greater. Broad disease classifications, such as ICDA-9 [WHO78], have become accepted, but are inadequate within any specialty. A comprehensive schema can provide a common definition for the data elements that are to be shared. The procedures to encode data within the schema may be made particular to each institution if they have differing conventions for their source data collection..

The solution provided through a schema, as described above, cannot overcome all problems of inter-institutional data comparability. Because different institutions will have differences in patient access, the subpopulations from the cooperating institutions will have different demographic distributions. This means that even when data are coded in a consistent manner, comparability of findings from different institutions is questionable. In general, uncontrolled pooling of multi-institutional data is to be avoided.

To support such shared databases a good communication system is required. Early systems had to build their own communication networks [Logema73, Mesel75]. Now commercial companies like TYMNET (used by MEDLARS, LADB and SUMEX) and TELENET (used by ARAMIS [McShan78] as well as by MEDLARS) can provide such services on a nationwide basis. The linkage to a network also provides a potential advantage for investigators who wish to establish a new database. They can shop around and determine which database system and service is best suited to their needs. Once a large data collection is established within any particular system, switching service suppliers is nearly impossible.

When remote use is made of a central computer the apparent reliability has to be as high that achieved by local systems. That means that the real reliability has to be very high. The apparent reliability is the product of many factors: reliable electric power, reliable central hardware, reliable and easy to understand software at the central site, good communication services with a minimum of noise and inconsistencies, and reliable terminal operation at the users site [Fries78]. The effort to provide all of these in a well integrated form mandates a major management effort, which adds substantially to the cost of a shared operation.

II.B.3 Databases in Health Maintenance Organizations

The larger health care delivery organizations have needs for management and policy setting as well as requirements due to medical services. A prepaid health plan or health maintenance organization (HMO) has to set rates that allow it to be competitive while generating sufficient income to cover the various services. The total cost of services will depend to a large extent on the population being served. Workers, families with children, the indigent, and the elderly all present different service patterns. Good medical accounting practices and reliable data are important to determine equitable rates. The outstanding example of such an operation is seen at the Harvard Community Health Plan [Justic74], where a MUMPS-based system, called COSTAR, provides services to two sites, one with about 20,000 patients. A generalization of this system, using schema-like parameters [Zimmer78], will be described in section IV.D.

The data in the COSTAR database are used in the clinic to provide a printed abstract of the medical record prior to every encounter and are available for on-line inquiry. An associated schedule-keeping program provides the information to allow most of these abstracts to be printed at night, so that the scheduled patient encounters no delay for record delivery. Encounter data are collected by the physicians on forms that are preprinted with problem-specific check lists, and entered subsequently by clerical personnel into the computer files. A small amount of free text can be recorded to cover situations where the check list is inadequate. The hierarchical structure of the MUMPS file system is well matched to the patient oriented view of such a clinical system:

> A patient is seen as having problems,
> for which the patient is seen by specialists in the clinic,
> an encounter results in a number of observations
> and treatment specifications,
> which may in turn have a number of data elements,
> such as drug name,
> with dosage, frequency, and duration.

This hierarchical view was already used in the earliest systems for HMO support [Davis70]. In data model terms we have a nest of drugs within a nest of encounters within a nest of problems for each patient entity.

For management functions, clinical as well as administrative, the access to COSTAR data is indirect. One reported instance in clinical management was the need to prepare a list of patients which had received a certain IUD - which was reported to be potentially harmful. For this task the file has to be searched patient by patient. Such processes are typically done overnight, in effect in batch mode. The overnight delay is certainly tolerable when reports that identify or aggregate many individuals have to be produced. Administrative management is in fact accustomed to much longer delays in conventional data processing operations, where special reports can only be produced as a byproduct of regular periodic processing. Guidelines for the financial management content of these databases have been issued by HEW [Densen72, Gaus73].

Issues to be resolved in the economist's arena are the trade-offs in an HMO between preventive care and restorative care, task assignment to physicians versus paraprofessionals, and the effectiveness of incentives for proper utilization. The analyses must give proper consideration to the patient

population that is receiving services [Hershe79, Luft78]. The data gathered in the database used within an HMO type operation can do much to resolve questions of economic concern. Several broad-based systems exist and have provided valuable data. For instance, the health services provided to a largely indigent population, comprised of the Indians on several reservations, have been recorded in a computer system by the Indian Health Service in Tucson, Arizone [McArth78]. The data have provide information for health care resource allocation.

Services to poor populations are provided by the federal and state governments through a bewildering array of categorical grants for health services. In South Carolina a system is in operation which was largely motivated by the need to coordinate these various grants [Penick76]. By matching a patients needs to the available services some degree of comprehensive care may be achieved. Here a multi-dimensional view of the database was implemented using a network type database system, IDMS.

Many of these systems define their initial database content using the result of a comprehensive study [Murnag73] of data elements for ambulatory care. The results were issued as a guideline by HEW [NCHS74]. A goal of the study was to achieve a consistent database useful for health-care policy decisions. While this study did not address health plans in particular, it is only in this arena that we find consistent population-based data collection [Linber70]. In some European countries [Gremy77, Hall77, Mase77] where health services are largely provided with government assistance, regional databases have been implemented, although their medical content is often quite shallow.

II.B.4 Surveillance Databases

Surveying the status of a patient population and the generation of appropriate advice, either to health care personnel or to the patient, may be an important function using a database. A simple database which is oriented towards that function, CIS, is in operation at Regenstrief Institute in Indianapolis, serving a number of chronic disease clinics [Bharga74, McDona75].

In the diabetics clinic of Regenstrief Institute, where many of the services consist of of laboratory tests and the advice is often provided by paraprofessionals, automated surveillance to track the patients status is especially effective. A special language has been developed so that the algorithms or rules are easily written. The vocabulary consists of the clinical terms embodied in the schema. The output of the system includes printed advice statements, as well as fields for data elements to be recorded during the current visit. The sequentially organized file is passed rapidly against the rule base. Maintenance of the clinic visit schedule is also provided. Upon leaving the patient receives the copy of the advice report, which is based on the findings up to the current visit [McDona76]. Each visit becomes a new entry in the database. This clinic provides an excellent environment for computer-based assistance. Since the objectives of such a database are initially defined, relatively complete data collection can be attempted. Most patients will cooperate with the goals of the clinic and the requirements of its system. Encounters are scheduled, and data processing can be performed either immediately after or prior to the encounter. The files can be organized for efficient batched processing.

An important application of such a system is also seen in the management of patients on chemotherapy [Wirtsc78]. Here the use of protocols for physicians at a number of remote sites permits an identical therapy regimen to be provided. The consistency is exploited here for in controlled trials, where variations can make the results of the trial unusable. An evalution has shown that difficult protocols are followed much better when the system is used. Eventually it may be feasible to interact with community physicians, sparing the patients the disruption of travel and integrating the primary and tertiary care system.

The aggregate database also serves an important function as a collection of experience in treatment effectiveness as well as patient behavior [Simbor76]. Surveillance is of course not just limited to patients, but can also be employed to track the behavior of the physicians or the clinic operation. Quality assurance surveillance can, for instance, determine if hypertensives are monitored and that patients with positive tests, say for strep-throat, are being treated appropriately [Barnet78]. Such surveillance will improve data recording as well as patient care. Data can only be relied upon if regularly used, and conclusions should not be drawn from data which have not beer verified by use or by audit. An investigation of the operation of a clinical surveillance showed that incomplete record keeping was responsible for many of the findings of lack of patient follow-up.

II.B.5 Specialty Clinical Databases

It is in the speciality clinics that databases have had the most medical impact. An extensive survey [Henley75], on which many of the observations made here are based, found an effective database operation in a private cardiology clinic in Oklahoma City [Wilson78]. The system was written largely by one medical student using MUMPS and maintains patient status data for routine, monitoring, and emergency visits. The system is also used to help in scheduling appointments and the combination of both files provides data for billing purposes. A second example of a speciality database is in an academic rheumatology clinic at Stanford University, using TOD [Weyl75], where signs and symptoms of a new patient can be compared with those of treated patients. The comparision generates a prognosis of treatment effectiveness in the new case. The database structure of these two systems differs dramatically: the former reduced the patient's past history to a concise snapshot for easy review, whereas the latter maintains a detailed time-oriented history for analysis. The cardiology clinic at Duke also relies on an integrated universal patient record [Starme74, Rosati75], whereas a MUMPS-based diabetes system at Washington University records visits over time [Achten75]. We conclude from these examples that the database model of the users is determined by their medical view, rather than by the facilities provided through the database system.

A strong emphasis on follow-up and time markers was noted in a specialized minicomputer database which is oriented towards the tracking of major workman's compensation cases [Leavit72]. An important economic incentive exists here: since the insurance company reimburses the wages lost during the recovery period it can invest in aggressive follow-up and medical treatment in order to reduce its total cost per case. Such motivation is lacking in insurance systems which reimburse for treatment only.

II.B.6 General Clinical Databases

Many of the services to outpatients that are provided by specialty clinics are equally appropriate to outpatients seen in hospital associated general clinics or in major group practices. In large institutions there is also a need for coordination of services carried out by various units, which may be physically or intellectually remote from each other. These larger institutions have had a need for billing their costs to the patients and to the patient's insurance companies, so that basic records of patient visits are already maintained on computers. The reason for the visit also should be recorded to assure that the encounter qualifies for reimbursement. The upgrading of a system from those beginnings to include a medically relevant database is not at all easy, and the motivation for the inclusion of medical experience is less in a general outpatient clinic than in a specialty clinic. In a general outpatient clinic there is typically less shared responsibility for care, and the patient population is more diverse and less committed to follow-up. In non-academic institutions we find databases with little structure beyond the billing date. AUTOMED in Cleveland, which serves private practices, permits addition of free form textual messages under several categories in each visit and provides a keyword oriented search capability. The structuring of data, such as the selection of keywords, the definition of appropriate values, or the creation of access capability to the message text for analysis, is left to the physician-user. A few have used the system well, but most restrict their use to pre-programmed billing services [Wied&M75].

Billing and usage-oriented data can contain information of interest. Records of insurance charges, when processed regionally, have provided a handle for the selection of cases for peer review [Buck74]. Data on usage patterns from the Kaiser-Permanente system has provided information important to health care system planners [Enthov78]. On a smaller scale, in a family practice clinic a system using MUMPS provides feedback in regard to patient visit management to physicians in training [Given77]. The medical record itself is collected by transcribing dictation into the database. Medical data in this form are not easy to analyze for trends across the population. Since there is a pharmacy on the premises, which does collect numeric data, analysis about drug prescribing and drug use, or at least about drug purchase patterns, is possible.

The complexity of data collection in general practice is demonstrated by the multi-page coding form which was used at CHCP, the HMO sponsored by Yale University [Brunje71]. Few data items were actually recorded per visit. Some structured organization of the medical record to simplify data collection appears neccessary. The problem-oriented record structure [Weed71] interacts well with well defined medical data management [Hall76], even if the rigor preferred by Dr. Weed [Weed78] is not maintained.

A comprehensive guide to computing in medical practice settings is provided in "Computers for the Physician Office" [Zimm&R78]. This volume also provides addresses of organizations and journals which are active in this area.

II.B.7 Databases in Research

Clinical databases are also important resources for research. Searching for cases through paper medical records and then abstracting data from them is tedious and costly. When a database is available the collection effort has been carried out. Feedback from the use of the data in patient care should have verified the correctness. Selection of cases and data is now carried out easily, making such a clinical database an especially useful resource. Since the patients, and the data collected on these patients, were not selected with the research objective in mind, some care has to be taken in research-oriented analyses. The conditions under which patients were entered into the database have to be understood, so that problems of selection bias can be dealt with.

For purposes of statistical analysis a rigid, tabular format is preferred, and encoding of observations is essential if problems of differing terminology are to be avoided [Sibley77]. Data to be encoded include diagnoses, subjective findings, stages of disease, and patient demographic characteristics. Many encoding schemes have been used; they are often locally developed to serve particular anticipated analysis needs [Stei&M78]. In general usage for disease and problem classification is the International Classification of Disease, Adapted [WHO78] and the coding scheme for ambulatory care of the Royal College of General Practioners [RCGP72]. Multi-dimensional coding schemes have been developed such as the Systematic Nomenclature Of Pathology (SNOP), with code definitions for each of four dimensions : (topography, morphology, etiology, function) [ACP69]. This coding was later expanded for medical care with the dimensions (diseases, procedures) into SNOMED [Cote77]. Speciality areas are developing their own versions. The Academy of Dermatology has now published SNODERM with codes specific to dermatology in these six dimensions [AAOD78]. The topic of codes is a major field of research in itself, and we will have to forego further exposition here. The actual encoding of clinical observations for computer entry can be carried out in a variety of ways. Common alternatives of encoding data for computing were presented in section I.D.1.

Clinical databases may take many forms, but since in general a cause and effect relationship is to be explored, we will need to collect patient data that represent at least both events. The event which is the cause should precede the event which shows the effect by some interval of time. If the interval is short, then a single patient observation may suffice: studies of effects of different treatments in emergency care have been of that type [Slosbe78]. When data is abstracted from the medical record, then the past history is available from the patient record and can be combined with outcome variables into a single record. When the cause is unknown, but possible correlates of the cause are observed, then a single snapshot may also do: an impressive example is the correlation of US geographical sites with cancer incidence [Breslo75]. The actual cause of the variability of cancer incidence is left to further hypotheses. Verification of such hypotheses may require new, and more extensive databases.

In general the chance of capturing cause and effect with one snapshot is rare. In most situations the appearance of the outcome event is delayed and unpredictable, so that the patient progress has to be observed and kept over a long time period. In the more general databases many visits for diagnosis and follow-up treatment have to be recorded and the time of

observation becomes a critical variable. These longitudinal databases become more complex and often contain missing data [Gree&K77]. Certain events in the patients history are milestones for time measurements, for example disease onset, start of treatment, and recognition of disease stages. Data has to be matched appropriately to these milestones for proper comparision and statistical analysis. An approach to deal with such problems will be described in Section IV.C.

The demands made by researchers on databases are oriented towards rapid access to large quantities of data. The resulting databases are often relatively simple, and data has to be extracted from clinical databases into these research databases prior to analysis. The research database is then not kept up-to-date, but its stability permits comparision of results using multiple analysis techniques. Database systems that are designed specifically to support research are now becoming available commercially or from public institutions and universities. These systems typically provide schemas to simplify referencing of the stored data elements for retrieval of subsets, and convenient linkage for statistical analysis. Most available database systems of this type have the capability for accepting incremental updates, but do not have the full integrity protection and recovery mechanisms associated with the large database systems designed to support on-line operational service requirements. Examples of such database systems are RAMIS (provided by Mathematica of Princeton, NJ), MISAR [Karpin71], WISAR [Friedm77], and RS/1, a derivative of PROPHET (provided by BB&N of Cambridge, MA). The availability and usage of these systems is increasing, since the provide good database capabilities at a lower installation cost than locally developed research support systems.

II.C FUTURE USE OF DATABASES IN HEALTH CARE

There seems little doubt that usage of the various types of databases cited above will continue to grow, even though the number of successful database operations is still quite small. In many current systems the breadth of usage and the degree of medical interaction is much less than was hoped for. As systems develop so that they hold a wider variety of data and become more capable in converting these data to information, they will become more attractive to the the users. Integration of proven services and their implementation within clean and reliable system approaches should remove some of the barriers which now deter acceptance of interesting, but isolated systems. Conceptual advances will depend on how well current research can resolve some important points, which are at a level beyond the simple technical issue of building systems that carry out the defined tasks.

II.C.1 Cost-effectiveness Issues

Economic considerations are a main driving force in the proliferation of databases. Escalating health care costs focus attention on the health care system. Although technology is frequently identified as one of the factors which drive costs higher, the cases cited seem to be instances where excessive equipment has been obtained and is idle. There appear to exist few solutions where a reduction of active technology will reduce health care cost. Medical records and databases are rarely identified as being redundant or inappropriate. A case could actually be made against extensive automated medical records for non-hospitalized patients that do not have any chronic disease. For an acute, self-limiting disease in an otherwise healthy person the physician gains little from the historical medical record in any of its forms. As our population becomes older the fraction of completely healthy people diminishes, so that substantial medical databases remain warranted.

The information from databases is, of course, also vital to the decision-making and planning processes in health care administration [Brande76, Moscov77] and such planning is often directed to health care cost control. Since technology continues to lower the cost of the components of database systems, while most other health care items get more expensive, we can expect further expansion of databases in health care.

II.C.2 Initiatives and Innovation Due to Technology Push

The decreasing cost of microprocessors and computer memories has been well documented. The mass storage devices, disks as well as new solid state technologies, are not far behind. A new and relevant storage medium is provided by magnetic domain or bubble memories. Advances in programming and systems design may be less obvious but are equally important to database technology.

The experience being gathered now in selected projects will provide the filter for new ideas in programming and software. Such a filter is needed since there are already now many technological choices for medical databases. Data entry, for instance, can be accomplished using any of a wide variety of methods, such as: dictation and transcription, free text typing, forms that permit easy encoding of the data, and selection of items from menus presented on a display screen. Only the last two options have been accepted by physicians in a clinical environment [Justic74, Schult76, Watson77].

Rapid interaction is required to present menus on-line, and providing the required file access, processing, and communication speeds have been quite costly. No generalized database system of the current generation can present menus at the rate that physicians can use them. This has, in turn, inhibited broader use of menu selection in medicine. Since the problem appears clear, a search for a technological solution can be made. Prehaps the menus can be distributed in a more direct manner to the terminals, so that the needed performance may be achieved at lower file and communication cost. Hardware that provides improved display distribution is feasible, and when introduced into database systems, menu selection methodology may become dominant.

The example cited above is but one instance where we expect that the availability of new technology will encourage new system development and wider dissemination of medical databases in general. Since data entry is such a bottleneck, improvements in this area are apt to trigger new database application efforts. Automatic scanning of typed documents and computer controlled management of video images are other technologies that are apt to affect medical data management soon.

Developments in software technology can have a similar effect. There are already many ways of information structuring: categorized free text [Lyman76], hierarchically structured data [Greene70], sequential files, relational databases [McLeod75], and network structured databases [Penick76]. No clear picture has yet emerged on which alternative, if any, will dominate [Wieder76]. As our understanding of information structures increases we will apply this understanding to medical databases. Choices can then be made on a rational basis. To apply new software technology to medical databases requires well-trained personnel. Some of the effort can be obtained indirectly: the use of commercial database products in medicine takes advantage of pre-existing tools. Their application still requires medically oriented expertise. Both the complexity and the importance of health care provide the needed incentive to obtain the attention of database oriented scientists.

II.C.3 The Human Element

No advances can be made unless there is a cadre of knowledgeable individuals. Traditionally there has been a lack of competent people who could integrate medical and computing knowledge. Critical decisions affecting the medical interface have been delegated to technicians. We do see a change occurring. Increasing numbers of medically oriented computer science students, microprocessor owners [Orosz78], and physicians with insight into automation will provide the resource for implementation and dissemination of computer applications in medicine. Training programs in Medical Information Science, often sponsored by the National Library of Medicine, are in operation at several institutions. As scientists with the necessary broad scope of expertise become available, the technology and the techniques, many of which exist already now, can be appropriately employed. With better cooperation between medical and computer scientists we can also expect that the impact of database systems will effect medical care in more direct manner, and that new applications for databases will be found.

As the interfaces to the systems become friendlier, and the scope of available services becomes more clinically relevant, direct interaction with the systems will become more rewarding for physicians. There will be less reason to deal with these systems through intermediaries. The physicians demands will in turn provide the impetus for further system and interface improvements.

II.C.4 Sharing of Information

Sharing of knowledge, data, and database technology is essential to progress. Dissemination of knowledge, in the form of research results, occurs traditionally via publication. The communication networks, created originally to connect people to computers, and computers to each other, are beginning to provide a faster and more effective path for knowledge sharing among scientists. This communication occurs through fairly informal messages or eugrams [Lederb78], sent over the network when the sender has an idea or need for some knowledge, and are read by the receiver when he has the time and frame of mind to deal with external inputs. The files used to store these snippers of scientific communication do not use deep database technology, but logs of these eugrams, covering a specific topic, are interesting repositories of knowledge development.

Databases for the sharing of data are more common than computerized sharing of knowledge. Data collected at multiple sites is integrated by some of the disease-oriented database systems, since a single site may not be able to collect sufficient data for analysis. Clinical institutions may also be limited in terms of research interests, so that a shared database can make the collected data accessible to researchers at a variety of sites. Shared access for researchers has been supported by the PROPHET system, a national facility for pharmacological research. The system provides strong hardware and software support for tabular data, computational analysis, and graphical output presentation [Ransi174]. This central computer system allows sharing of the files containing drug data as well as sharing of the computational facilities to process the data [Cast&W74]. Data that are to be shared require shared definitions of the attributes to assure consistent collection and encoding. Lesion sizes and staging in cancer, as needed for the cancer CCPDS database, are examples of difficult data definitions. Comprehensive schemas can help if they make the definitions available on the display screen during data entry. The communality of disease and treatment codes, enforced due to the spread of reimbursement controls, also has benefits here.

Data stored at a central site is typically not used for ongoing clinical activities. Concerns about reliability and local responsibility inhibit such usage, and administrators also fear loss of control over data important for daily operation. Networking and other forms of inter-database communications can increase the number of systems accessible for sharing of data. We can foresee systems organized so that data is kept and maintained locally, but where remote access for research is enabled. Then research projects can aggregate or compare observations from distinct health care service centers. Access to data at remote centers may not be as fast as local access, but should be as easy.

Distribution of systems can make new technology accessible to sites of similar interest. Distribution of COSTAR [Barnet79], CLINFO [Thomps76], the Duke Cardiology system [Starme75], avoids reinvention and reprogramming of complex software. As an important byproduct such system sharing creates human networks of common interest and terminology.

In general the hope is for a symbiosis. Sharing of data will require more adherence to standard definitions. More standardization will make it easier for researchers and analysts to compare their results and to validate each others findings. There have been only a few databanks that have been analyzed outside of the institution which originally collected them, so that the scientific maxim of repeatability of experiments has rarely been demonstrated in the database field. Repeatability is necessary for scientific concensus. Broader agreement of scientific issues in health care should simplify the establishment of effective health care policies.

II.C.5 Privacy in Databases

A continuing concern in medical databases is the guarantee of adequate privacy. An argument often made by technologists is that the methods of keeping the paper record are not very secure, but this not an excuse for inadequate access control in databases, since records that are searchable by computer are much easier misused. Much medical information is distributed by the health insurance system, since reports of physical exams shared among insurers. The technology exists today to make access to computer files very secure [DeMill78]. For instance physicians notes, which would concern only a few database users, can be kept in cryptographically encoded form. Then access is restricted to parties which have been given the decoding key, typically some easily remembered sentence. Such encrypted information is not processable by computer, although text can be filed and moved among computers.

Who should have access to what portion of computerized medical records is an open issue. Since a database program can select individual data elements by name, access can be made much more specific than is possible with paper records or file systems that do not differentiate between data elements. A study sponsored by the Society of Computer Medicine has provided as a tentative guidelines a matrix of access privileges to the basic ambulatory data elements. The accessors considered are health care providers, financial agencies, health care planners, medical researchers, lawyers who represent the patient, and employers [Jelovs79]. The only element to be denied to the provider is the social security number. The extent to which different categories of health care providers should be differentiated in terms of data access is not clear.

Since a major objective of databases is the sharing of data, there is little benefit in the entry and collection of data unsuitable for dissemination. We advocate leaving very sensitive or critical material out of the database altogether. When data is transferred from a clinical to a research environment then identifying characteristics such as name, birthday, id-numbers, etc. are best deleted. Sometimes follow-up requires a reverse linkage capability. Hashing and encrypting techniques exist which permit linkages to be maintained in one direction only [Wieder77].

II.C.6 Missing Data

A specific problem which has to be faced in the effective utilization of health care databases is presented by missing data. Clinical databases, due to lack of relevance at the time of data collection in relation to cost of data collection and entry, will always have incomplete records and missing data. This leads to a number of issues which have been often addressed, but not yet satisfactorily treated in a general sense. The statistical routines used to process such data have to be robust and designed to explicitly process missing data. Many algorithms cannot deal at all with missing data. Then preprocessing is needed to remove records containing missing data, but often the number of valid rows and columns is drastically reduced, so that it becomes difficult to generate significant results. Techniques to replace missing values with appropriate interpolations may have less effect on the significance, but generate discomfort in the mind of the analyst. A prerequisite in either case is that the encoding scheme used recognizes missing data during data entry and that the storage representation never confuses valid data and missing data.

In controlled clinical trials a large amount of effort is expended to make the databases complete so that statistical conclusions can be drawn with confidence [Haensz66]. The ensuing cost limits data acquisition so that the approach of clinical trials is mainly effective when well-defined hypotheses are to be tested. Often the number of candidate patients is already limited due to the prevalence of the given disease and treatment population. Little has been published about the database methodologies which support this important and active area, although the results of studies performed are always made public. A new journal on controlled clinical trials may help to overcome this deficiency [CCT79].

New ways of dealing with missing data will have to be developed and integrated into clinical database systems so that conflict between collection limitations in a practice and research demands for completeness of data can be mitigated. One approach may be to perform the data analysis at a higher level of abstraction [Blum78] so that multiple observation types can be combined into a single medical concept. Then, if selected observations are missing, the general medical concept can still be supported from related, existing observations.

II.C.7 Problems of Current Interest

To make databases more responsive to medicine a number of specific scientific issues have to be addressed. When reasonable solutions appear possible the results will have to be embodied in experimental systems and presented to a broad audience of medical users to test the validity of the approaches.

We are defining databases too much in computer-oriented terms rather than in medically-oriented terms. Database schemas require that data be identified as numeric, real or integer, or as variables that hold a string of characters. Variables describing disease, type, stage, and location are forced to be explicitly encoded outside of the database system, so that natural terms are not used for input and absent from database output. Even well understood and common data element types, such as the date of an event, are handled ineptly, in a variety of formats that are often not obvious and hence error prone.

We are not yet dealing well with data that represent time or relationships of events that are time-oriented [Bolour79]. Manipulation of such data is essential if we are to build valid cause and effect models. The problem of inference from recorded events over time can be approached with various methods, but the methods now available are either simplistic or quite ad hoc. Statistical techniques to deal with data observed over time are based on models that are overly simple. Time-lagged correlation assumes equal time intervals, but in general we do not observe our patients at fixed intervals, nor can we assume that a disease progresses at a fixed rate, independent of the individual. Heuristic inference methods can capture time dependencies, but are often quantitatively weak.

Adequate higher level models are essential to the effective utilization of the contents of databases, as has been strongly argued in [Schnei75]. The physicians knowledge of the relationships of signs and symptoms to diagnoses and outcome is as important as the observations themselves [Pople75]. The acquisition and management of such knowledge is a central topic of research in applied artificial intelligence [Pople72]. Application of knowledge about the content of the database promises to help greatly in making effective use of future databases [Kuliko77, Schnei78, Slamac77, Baskin78]. The data in turn can be used to add quantification to the rules used to represent medical knowledge.

Many of the problems in medical databases are attacked where and when they become bottlenecks, usually first in a particular application. Rarely is literature consulted at that point, a programmer is pressed to design and code a solution. The fact that Medical Information Science is increasingly recognized as an autonomous field, with its own journals [JMS77], will help to disseminate new solutions to problems to the workers in the area, so that future, better systems can be built on the results of current research.

Scientific advances can increase the utility and the depth of database usage in all health care application fields, although implementations of databases will continue to differ in their emphasis in the according to the health care environment.

II.D Effect of Databases on Health Care Cost, Quality, and Access

The sharing of information, made possible through the use of databases, is expected to have positive effects on the health care system. The mechanisms that lead to such improvements include:

a. Readily available information will reduce the need for duplication of laboratory tests.

b. Databank analyses that advise physicians of possible drug-drug interactions for a patient will reduce the frequency of iatrogenic illnesses [Hulse76, Morre177, Cardon78].

c. Databank analyses that advise physicians of possible drug-laboratory test interactions for a patient will reduce the number of invalid laboratory tests [Young72].

d. Tracking of individual patients who are at risk can prevent inadequate follow-up, reduce morbidity, and the associated long-term care costs [Johns77].

e. Computerized problem lists in the medical record can help assure that all problems of each patient receive attention, rather than just the most obvious ones.

f. The availability of a copy or an abstract of the medical record at all of the candidate encounter sites and at the time of the encounter can prevent misdiagnosis and over-prescription[Lyman76].

g. Records, perhaps with data selected and formatted for the particular site, can improve the effectiveness and scope of community and paraprofessional personnel, and thus support a multi-modal health care delivery system [Mesel76, Zielst77].

h. Data for health care research at various levels can be extracted out of clinical databases, so that costs of otherwise redundant data collection can be avoided.

These factors can all lower health care access costs; others affect the quality of health care [Barn&W78]. The use of paraprofessionals should improve access to health services.

It has been difficult to demonstrate the benefits of these systems to a high level of statistical significance. The measures collected by the study on the implementation of a hospital information system at El Camino Hospital have that problem, even though the trends favoring the automated system appear clearly [NCHSR77]. The effects to be demonstrated occur over a large patient population and a long span of time. The before and after measures of outcome show only small differences, but due to the high cost of health care makes even small differences are impressive in absolute terms. Since health care does not stand still for measurement, many confounding events take place during a study. It is nearly impossible to find two similar

health care institutions which can serve the roles of test and control sites for an experiment. But informal assessment of operational database systems takes place continously. It is actually doubtful that any inappropriate system will remain in operation very long. Eventual acceptance of a medical database systems can provide another, quite stringent, evaluation of effectiveness.

Databases do provide the information needed to measure cost, quality, and access, and can be viewed from this point alone to be a tool in the improvement of health care [Donabe78]. In our study we found that, contrary to our hypothesis, that the presence of federal funding did not make the eventual system less vital than systems that were privately developed [Henley75]. Overall the failure rate was quite high, much higher than the success rate; but failures can also be ascribed to many other factors than the database technology. It is obviously important to have a system that is well matched to the setting. The computers and their software typically make up less than half of the system cost during operation, the time and effort spend by medical personnel with the system is very valuable, so that medical relevance weighs more than direct system cost considerations.

The effect that research results, obtained through use of medical databases has on health care delivery is even more difficult to measure. Since most large clinical studies rely on databases the assessment has to shift to the efficacy of such research. There is little doubt, that if reasonable models of cause and effect are used, that such research increases our understanding of many disease processes. Such medical evaluations, will improve health care quality and reduce cost. If understanding of disease processes also leads to the provision of appropriate entry points into the system, by say systematic screening of populations likely to be affected, then health care access has also improved.

III. STATE OF THE ART OF DATABASE TECHNOLOGY IN HEALTH CARE

There is of course a strong interplay between database development in general and its application in health care. Most basic research takes place outside of the health care field. But applications of databases are neccessarily application area related, and current work spans the entire spectrum of database use, since the ideal database does not yet exist anywhere.

III.A Systems in Research or Development Status

Many new data base systems and experiments are motivated by specific problems that are common in health care and in the systems that attempt to serve the area. In this section we will touch upon two such problem areas and indicate methods that have been tested or are under test in current systems. No single system can attempt to advance the state of the art in all problem areas at the same time. As specific problems reach reasonable levels of resolution the most successful methods should be integrated into systems which have concentrated on resolving problems in other areas. This integration of solutions has not always happened, partially because of difficulties in the tranfer of new and incompatible technology, sometimes because of lack of awareness of approaches that were demonstrated in similar, but not directly related areas of application. Some types of problems may affect one category of database more than another, since system requirements depend on the institutional setting or on the type of user served by the database. In the Appendix we will describe in detail several systems which are at the forefront of their area of application and which embody many of the aspects needed for dissemination of database technology. We have already considered the critical problems of data entry and management of missing data because of their particular importance to medical databases.

The support for controlled clinical trials, presented in section II.B.4 above, is now being extended with the use of distributed small computers (Data General Micro-Novas), which are to be placed at the health care delivery sites. Data entry and protocol advice can be supported locally; at night the systems communicate with the central database (GMDB), where the long term data for the southwest cooperative Study Group are being maintained.

The work on the PROMIS system has recently concentrated on the data entry and transmission problems [Schult75]. Special terminals with touch-sensitive screens are used, and data is transmitted in packets over a shared coaxial cable, a method which provides a very high performance at a reasonable cost [Wanner78]. Languages to simplify the definition and use of the display frames on the terminals are part of several system efforts. Other techniques to deal with data entry are not associated with particular system developments. Avoidance of redundant entry, perhaps combined with collection close to the source of data, using mini- or micro-computers interacting with medical personnel, appears to be a fruitful direction. Voice data entry will become available as a means to collect data using limited vocabularies, similar in style to menu selection schemes.

Artificial Intelligence techniques to deal with missing data are being explored in the RX project [Blum78]. Here the detailed data observations from the ARAMIS database are being integrated into higher level concept, which may then be used to define a patients progress at a more clinically relevant level. Proper management of time-oriented causality is an important aspect of this work.

III.B Industrial Status

There are two directions in current commercial efforts. First there are companies which address specically the medical market, and then there is general database system development. Some of these general systems may serve certain medical application areas quite well.

III.B.1 Medical Database Systems

We have defined throughout that databases consist of data, and systems to manipulate the data. While hardware and software vendors will make file and database systems available, the data collection and information applications remain in medical hands. Since it is difficult to develop medically relevant systems without having the required medical expertise we find that commercial firms often develop their prototypes within a specific medical environment. The chosen healthcare institution may benefit by obtaining a system which satisfies its particular needs, without the expense of paying for a handcrafted system, and the vendor benefits from the access to healthcare expertise, which would be nearly impossible to duplicate in the vendor's domain. But there are also major costs for the institution in such a cooperative venture. The system development, since it needs to have general applicability, will take more time than would be required for an institutions specific system and that the vendor will find some aspects of the institution are so particular that they will not served well by the system. It is also easy to underestimate the cost in physician, nurses, and management time involved in participating in a development project. Excessive expectations may lead to disappointments by either party, the greatest disappointment occurs if the vendor fails, either in terms of producing an acceptable system or totally as a business.

Most systems now on the market had their beginnings in cooperative efforts. The Technicon Medical Information System for hospitals was developed at El Camino Hospital in Sunnyvale, California, and even though there were traumatic moments, the hospital now obtains services at a favorable price, and Technicon markets the system to other hospitals. Another example of a hospital system is given in the Appendix in section IV.E. Here a tripartite partnership was in operation: a software company, Dynamic Controls developed the programs, the hospital obtained the IBM hardware, and the result is marketed with IBM assistance.

The MUMPS and COSTAR systems were developed at Massachusetts General Hopital, by its Laboratory of Computer Science, with major support from the National Center of Health Services Research. The programs are hence in the public domain, and a number of companies have taken them, improved them, and started to provide services based on them. A problem due to the multiplicity of MUMPS vendors has been that a number of language dialects have developed. In 1978 a standard definition of MUMPS was accepted by the American Standards Institute, and it appears that all new work will be based on this standard.

In Germany a project of some health insurance agencies is developing a system, based on PASCAL-oriented microcomputers, for distribution to their client physicians. The intent is to combine the keeping of simple medical records and claims processing in the office, and transmit the information to central computers via maiable disks or by direct linkage.

III.B.2 General Database Systems That Are Applicable to Health Care

Whereas in the past few commercial database systems have had the required flexibility, reasonable scale, and pleasant human interface to be useful in a medical setting, we see today that this picture is changing. Simple database systems are becoming available on many midi- and mini-computers. Terminal interaction, using formatted displays, is becoming better understood. Requirements for reliable long-term storage are aided by improved software technology and improved hardware. Hierarchical commercial systems, such as the MRI/INTEL System 2000, are being used to store patient-oriented records in clinical settings. CODASYL based network databases are seen in large institutional settings [Penick76], and a FACOM system serves an eye clinic in Osaka [Watana78]. The Laboratory Animal Data Bank sponsored by NLM uses the Batelle BASIS Information Retrieval system, as well as TYMNET for access, and this combination has provided comprehensive and reliable services at relatively low cost.

Data General's DBMS system, INFOS, is used with specialized hardware and software to manage documents which track the recovery process of workmen's compensation cases [Bucho178], and to assure follow-up of care services. The National Spinal Injury Research Data Center at the Good Samaritan Hospital in Phoenix, Arizona, collects regional information on spinal injury, its treatment, and the costs of treatmeent and rehabilitation. It uses a a simple network system, IMAGE3000, available for the Hewlett-Packard HP-3000 computers [Jewel79]. These two systems are distinguished in that they do not only consider the medical costs, but also the cost of having a person disabled during the recovery period.

There is also a flow from medicine to the commercial world. The MUMPS system, developed for and in a medical environment is now being marketed by DEC for general minicomputer data management as DSM-11 [Bowie76]. A standardization effort has focused attention on the language and system [Wasser76], and recent implementations make use of modern data structures, specifically B-trees which provide the hierarchical access without excessive dependence on specific hardware parameters [Wieder77].

The demands made in medicine for statistical analysis are a model for interactive statistics in other planning environments. Many statistics packages have had their origin in medical research, and have now entered a more general market. A particular example is RS/1, a research support database system sold by BBN, shows its heritage to PROPHET, the pharmacological research support system developed by BBN under NIH sponsorship, and employs for the data analysis tasks the BMDP statistical package, which was developed at the Health Sciences Computing Facility at UCLA, also under NIH sponsorship.

III.C Current Directions of Development

Improving the interaction and high level usability of medical systems is a major ongoing effort. The addition of decision criteria [Warner78], knowledge about the data [Levy78], and heuristic rules to relate medical events [Short179] are basic to this line of development. Permitting queries formulated in natural language can remove barriers to accessability [Epstei78]. Improved coupling of these techniques to databases can make the collected experience more valuable. Alternatives to knowledge extraction from databases can change the manner in which research results reach the physician [Blum78]. A review of computer projects supported by the National Center for Health Services Research, many of which address issues discussed here, csn be found in [NCHSR79].

While the use of a DBMS is traditionally associated with fairly large computers, a number of DBMS's are now becoming available for small machines. Large machines have not always been able to provide the reliability, priority of access, and low cost desired in the medical area so that DBMS technology was not easily available in health care. The majority of medical databases in use today do not use a DBMS, but it is likely that the usage of DBMS's will increase. The continued reduction of computer hardware prices favors this tendency. The cost of writing software is not changing as fast, so that inefficiency of hardware utilization, caused by the use of a more powerful, but standard product is probably less costly than the effort of installing and maintaining a specialized and optimal system. There will remain situations where specially tailored database support is needed to bring performance within the constraints of critical time limits. For high volume operation some operational costs reductions might be gained from such tailoring. Problem specific adaptations are often more easily applied to a file system than to a database system.

We have already discussed issues associated with access to databases on a distributed basis. The availability of public networks will accelerate this development and eventually permit accessing multiple related databases within one analysis task. Communities of medical scientist will be bound together by shared interests, exemplified in the databases, rather than by the boundaries of the institutions which employ them. Centers of excellence will maintain specialized databases, while their associates can be remote and interact with collegues, patients, and students in their institutions.

Databases will then be the repositories for expertise and knowledge will be tested for validity against the collected data, and quantitative parameters will be based on database analyses.

Acknowledgement

I wish to thank the providers of the sample material presented in the Appendix for their willingness to share their work. Much intellectual stimulation has come from my colleagues on many evaluation committees. Marty Epstein of NIH DCRT, Dr. Bob Blum of Stanford, and Dr. Alan Rector of Nottingham have provided important remarks and insights into the development of databases. Dr. Blum was also of major assistance in making this paper more readable and Marty Epstein provided many corrections and references.

IV. Appendix: Examples of Current Databases in Health Care

IV.A Public Health: THE CENTRALIZED CANCER PATIENT DATA SYSTEM

[This material is based on the introductory description by Polly Feigl and other personnel of the SAQC, Fred Hutchinson Cancer Center, Seattle, WA.]

INTRODUCTION

The Centralized Cancer Patient Data System (CCPDS) is a standard system for registering persons with reportable malignant neoplasms, who are patients of comprehensive cancer centers in the United States. Eligible patients were first admitted to a center on or after July 1, 1977, and are reported to the Statistical Analysis and Quality Control (SAQC) Center in Seattle, Washington.

OVERVIEW OF SAQC

The SAQC Center, located at the Fred Hutchinson Cancer Research Center, consists of three units to carry on the technical activities: a Field Liaison Unit, a Data Processing Unit, an Epidemiology Unit, plus an Administrative section. Most data is received at SAQC on tape and is immediately subjected to computerized analysis. Resultant reports inform the submitting center of cases accepted into and rejected from the database. Rejected data will be corrected at the source for resubmission. Communication regarding technical matters is generally carried out by field representatives assigned to centers. Centers' data coordinators collectively provide advice to SAQC via the Technical Advisory Committee. Within this group, special subcommitteess are involved with Quality Control and Training, Data Utilization, and Research Planning. A Policy Advisory Committee, also composed of center representatives, affords further advice on a policy level.

Thirty-eight items of information are collected on each patient, including demographic characteristics, diagnosis, therapy, and survival. Standardized definitions of data items have been documented in the "CCPDS Data Acquisition Manual" (DAM). This manual also includes recommended procedures for abstracting, coding, submitting data to SAQC, and quality control.

OBJECTIVES OF CCPDS

The database being collected has a number of purposes. Selection of cases over a broad population will allow clinical researchers to locate similar patients for detailed comparative analyses. Treatment conventions may differ among the cancer centers, and such differences and their effects may be identified through the CCPDS database. Knowledge about the prevalence of various types of cancers and the changing effectiveness of treatments can warrant increased efforts in specific areas.

Initially, standard definitions and codes had to be established for reportable patients and tumors, as well as for each of the thirty-eight data items. Then, criteria for quality control were set up to assess accuracy, completeness and timeliness of reporting. There is a continuing effort to maintain intercenter comparability and compatibility with other national and international cancer reporting systems. CCPDS data is disseminated according to policies and procedures developed by a Policy Advisory Committee for that purpose.

On a long term basis, modular expansion is seen as the way for CCPDS to function for collaborative analytic studies. Special areas of interest identified include staging for certain sites, therapies, etiologies, health services, and rare neoplasms. The central registry can serve as a means of identifying patients for studies or for identifying institutions with the potential to contribute such patients to such studies.

PATIENT DEFINITION

A patient who is reportable to CCPDS is any individual with a frankly malignant tumor who is seen as an in- or out-patient at the center and is assigned a hospital or clinic number. Included are patients whose diseases have been clinically diagnosed but not microscopically confirmed; patients not diagnosed at the center but referred for therapy of of recurrent or late metastatic disease, and patients who are clinically free of disease, but are admitted to the center for adjuvant or prophylactic anti-cancer therapy, if that admission occurs within two months of the initial treatment.

Patients who are excluded due to this definition are consult-only cases, cases diagnosed at autopsy and former cancer patients with no evidence of residual disease who are admitted for rehabilitation or for treatment of some other condition. Also excluded are cases of basal and squamous cell carcinoma of the skin.

CCPDS DATABASE AND COMPREHENSIVE CENTERS

The data processing system at SAQC has been set up to register about 47,000 new patients cases each year. Follow-up data is added annually on all registered patients. Patient data is entered into the database only after successfully passing all of the SAQC edit checks.

As of March 1979, twenty-two cancer centers have been designate as comprehensive centers, based on criteria which include having an adequate statistical base. Some centers are actually consortiums of hospitals which collectively submit data to SAQC. The centers are: University of Alabama, Colorado, Duke University, Florida, Fox Chase & University of Pennsylvania, Fred Hutchinson, Georgetown & Howard Universities, Illinois Cancer Council, Johns Hopkins, Los Angeles County & USC, Mayo Clinic, M.D. Anderson, Ohio State, Roswell Park, Sidney Farber, Sloan-Kettering, UCLA, Wisconsin, Yale University, Detroit, and Columbia University.

QUALITY CONTROL

A major on-going effort at SAQC is directed toward assessing the quality of CCPDS data. A Data Monitoring Plan has been written for monitoring accuracy, timeliness and completeness of data, as well as compliance with established SAQC data acquisition rules and procedures. Coding practices at contributing centers are measured in several ways, one of which is by applying computerized edit checks to data submitted to SAQC. Also, SAQC field representatives visit each center annually to independently reabstract and recode a random sample of previously reported cases. A coding reliability study was conducted during 1978 for which a standard set of test cases was sent to each center for abstracting and coding. Item error rates up to 16% were found, and up to 32% in staging codes.

The various procedures for assessing quality of data allow for comparing coded data both between centers as well as between centers and SAQC. These early efforts toward enhancing data quality will pay off when the data is utilized for research studies.

LIST OF CCPDS DATA ITEMS

INITIAL REGISTRATION

Item No. data type

Identification
1 Institution Code
2 Patient Identification Number and Check Digit
3 File Number
4 Birthdate

Demographic Information
5 Birthplace
6 Race/Ethnicity
7 Sex
8 Residence at Time of Admission

Diagnosis
9 Date of First Admission to Center for This Tumor
10 Sequence
11 Date of Initial Diagnosis
12 Primary Site
13 Laterality
14 Histology
15 Diagnostic Confirmation
16 Date of Best Diagnostic Confirmation
17 Stage of Disease at Time of First Therapy at Center

Therapy Cancer Therapy Prior to Admission to Center
18 Surgery
19 Radiation Therapy
20 Chemotherapy
21 Endocrine Therapy
22 Immunotherapy
23 Other Cancer Therapy
24 Date of Initial Therapy at Center

Initial Therapy After Admission to Center
25 Surgery
26 Radiation Therapy
27 Chemotherapy
28 Endocrine Therapy
29 Immunotheraphy
30 Other Cancer Therapy

Patient Status
32 Date of Last Contact/Death
33 Autopsy
34 Cancer/Treatment Related to Death

FOLLOW-UP ITEMS

Identification as before

Follow-Up Information
35 Date Report Prepared
36 Method of Follow-Up
13 Laterality
32 Date Last Contact/Death
33 Autopsy
34 Cancer/Treatment Related to Death

PROTOCOL ITEMS

Identification as before

Protocol Information
37 Date Entered on NCI Protocol
38 NCI Protocol Identification
37 Corrected Date Entered on NCI Protocol
38 Corrected NCI Protocol Identification

IV.B Randomized Controlled Clinical Trials: HARVARD PUBLIC HEALTH

[This information was provided by Dr. William G. Costello of the Harvard School of Public Health.]

The Division of Biostatistics and Epidemiology in the Sidney Farber Cancer Institute of the Harvard School of Public Health provides data processing and analysis services to groups engaged in clinical trials. The primary groups being served are consolidated as the Eastern Cooperative Oncology Group (ECOG) and the Radiation Therapy Oncology Group (RTOG).

The data processing system for the ECOG and RTOG consists of software, hardware, and procedures which have been developed over the past 3 years and now operate smoothly for all studies of these groups. It encompasses every aspect of the processing, from data collection to data updating and error checking, and retrieval and analysis of data. A modular design was adopted which allows for flexibility of operation and routine incorporation of new Group studies or data formats. Although the software is largely general-purpose, the system is specifically oriented toward clinical cooperative group activities.

Overall System Design - Input

The overall design of the data processing for ECOG and RTOG may be seen by tracing the flow of data through the system. Data forms are received from member institutions via the Group's Operations Office, and are logged and checked for completeness. Data Managers conduct extensive manual checking of the data to ensure high quality. This includes verifying that:

1. The patient was eligible for the particular study.
2. The treatment was given according to protocol.
3. The toxicities were reported correctly.
4. There was adequate documentation for tumor response evaluation.
5. The required data items have been answered.

Any queries about the data or requests for more information are sent back to the contributing institution. During 1977, more than 1100 query letters were sent to ECOG investigators. When complete case records are available, the Data Manager prepares a Case Evaluation Form. This provides a vehicle for noting eligibitity or protocol compliance problems or any problems in evaluating what happened to the patient. This is sent to the Study Chairman to aid in reviewing the records. Ultimately, the contributing institution receives a copy of this evaluation, so that a formal feedback mechanism is available to them.

Most forms used in these groups have codes printed with the boxes to be checked so that they are completely ready for data entry personnel once checked by the Managers. All self coding forms and data entry documents are sent to data entry where they are keyed and verified.

Data items are checked for syntax and range as they are keyed. The records are then transmitted to the DEC-20 computer operated by the division and distributed to the responsible Data Managers' directory along with electronic mail notifying the Data Manager that he or she may now update the study files. The update is initiated by submitting a control file. Updates are run in batch mode at night.

During the update run, new data items may be calculated, for instance survival time. Both these calculated items and the primary items are checked with automated editing procedures which include:

1. Format of input data
2. Presence of "must fill" fields
3. Proper values - range checks and checks for special "allow" values
4. Logical checks - comparison of data items in the patient record for consistency and plausibility.

Any exceptions in the data which are detected by these automatic procedures are reported to the Data Managers. They are responsible for correcting or resolving the discrepancies; query letters to institutions may be generated at this stage, as well. Once past the editing stage, the input data are merged into the master study data files.

In the event of hardware or software failures, backup copies of the data files are maintained, so that recovery of data files is possible. Multiple backups are kept in physically separate and remote locations.

Supporting the automatic editing and updating functions is a system of data description files, called the Data Dictinary. These files allow the editing and updating programs to proceed automatically, because they contain the formats, field widths, allowable values and cross-checking procedures which are used for these functions. Central maintenance of the Data Dictionary is the responsibility of the Data Base Administrator. The Data Base Adminstrator also coordinates all data file maintenance and initialization operations.

Overall System Design - Output

All outputs from the study data files are based on two principal features: The QUIRE retrieval system and the Data Dictionary facilities. These capabilities allow retrieval of any data, even by non-programmers, with minimal training and work. The system has data independence features which make it possible for users to be insulated from any changes in data formats or additions of new data types. This simplifies the use of the data and provides for dynamic growth in the structure of the data base.

The QUIRE retrieval system is a collection of programs which allow the user to retrieve any data in the study data files, without requiring detailed knowledge about data formats or any computer programming on the part of the user. In order to specify the desired output, the user may merely indicate which study or studies he is interested in, and give a list of identifiers of the data items he wishes to retrieve. The QUIRE software refers to the Data Dictionary to find out how the data is stored and other necessary details. The system automatically generates the output data file. In addition, the system generates a description file which contains the necessary formats and labels to describe the retrieved datafile. This is an important aspect, because it allows building of report generation and analysis programs which can automaticall access that data file without further programming effort on the part of the user.

The output subsystem includes software for data transformation (creating new data items, re-scaling or re-grouping data values, etc.), listing of data items by patient, and generation of cross-tabulations. Analysis programs include the SPSS statistical package, the IMSL subroutine library, and many programs developed in the Statistical Laboratory which are important to the analysis of clinical trials data. These include sophisticated procedures for evaluating treatment and other effects based on multidimensional contingency tables and censored survival data. Programs may produce computer listing output, special data files for later analysis, or graphical output. The latter is particularly useful both for analysis of complex data relationships and for representation of results in compact form to clinical investigators in the Cooperative Groups.

The study data files are organized as separate files for each study. Each file has a sequential data layout. This allows good performance for the analysis of data from a given study by statisticians.

The Data Dictionary is composed of several fields describing various aspects of the data. Generally, its organization reflects the input data to the system. It is based on segments of data, each segment consisting of a number of related data items. For example, a self-coding form may be described as a segment. A file exists for each segment, and records contain:

1. A short identifier,
2. A short alpha-numeric descriptor,
3. Details about where and how the information is stored, and
4. Data range checking information to be used for automatic editing when the item is updated.

There are general segment description records in another file. Finally, different views of the data are supported through a system of directory files, which catalogue the description files needed for a given application. The user need only identify the application (for example, all data for a given study) and the QUIRE system takes care of finding the appropriate descriptions through the directory files.

Report Generating Capability

For instance, an excerpt from the System Data Dictionary (SDD) file for ECOG data could be a directory file describing the view of the data corresponding to the protocol ECOG 4274. This would contain the names of all segments comprising the description of ECOG 4274. Each segment description contains variable description information for each data element and includes:

1. Data element key
2. Data element FORTRAN format type and length (for instance: I5)
3. Starting position in data line (16)
4. Logical segment type corresponding to a form (A)
5. Data line type (C)
6. Some descriptive information for each data element
7. Minimal legal value for each variable
8. Maximum legal value for each variable
9. Four "allow" values for each variable even if they do not fall into the range given by 7 and 8.

In updating extensive cross variable checking procedures are also used.

IV.C Clinical Research: TOD, THE TIME-ORIENTED DATABASE SYSTEM

[This material is based on an overview paper by Drs. Dennis McShane and James Fries of the Stanford University Medical School.]

The Time-Oriented Database System supports data banks primarily dealing with chronic diseases. A major user is the ARAMIS project with six operating databanks in rheumatic disease from six institutions. Other current users include the Northern California Cancer Center and the National Stroke Data Base.

The individual data banks utilize the same file structure, format, and common entry and retrieval programs. Programs and software are schema-driven and content-independent.

Two components underlying are ARAMIS are a database definition for rheumatic diseases and the TOD Software System. The Uniform Database for Rheumatic Diseases, promulgated by the American Rheumatism Association Computer Committee, consists of a vocabulary of 422 variables, describing attributes of the various rheumatic disease processes and ranges from descriptors of demographic information to symptoms, signs, laboratory values, diagnoses, and therapies. This database definition is the standard vocabulary followed by rheumatologists seeking to clinically describe patients with rheumatic disease and has widespread acceptance in the rheumatology community.

The TOD System was developed within the ACME project at Stanford University for ease of data retrieval in useful clinical research formats and is now maintained by the Stanford University Computing Facility. It is being further developed by ARAMIS programmers, and currently operates on an IBM 370/168 under a locally maintained timesharing system, Orvyl, with an adaptation of IBM PL/1. Programming objectives are beyond the scope of this review, but include heavy use of macros common to many programs, internal self-documentation, and optimization for retrieval, even at the expense of entry or storage considerations. Every effort has been made to keep system design simple for the physician-user and to keep search strategies intuitively reasonable.

Each TOD databank contains two distinct structures. The main file maintains patient records, each containing an array of data from a single patient visit. Visits are entered or updated interactively The transposed file contains corrected and validated data and is used for retrieval. Each of these files can be considered separately in more detail.

TOD Main File

The main file consists of all information, organized by patient visit. In the TOD System a patient course is conceptually considered to be a two-dimensional array of numbers in flowchart form. Each column represents a series of observations (or elements) for a single patient made at the same time variable in the same patient over time.

By adding additional patient courses, a third dimension is created, whereby any value in the data bank may be accessed by three coordinates within a conceptual cube: the name of the variable, the name of the patient, and the time-point of the observation.

Thus, a clinical data bank in ARAMIS can be defined as serial observations for a given set of variables in a defined population. While this three-dimensional structure may seem obvious, hierarchical system designs in medicine have not formalized the critical time dimension.

TOD Transposed File

The transposed or retrieval file is a rearranged main file in which all values for each attribute of the Uniform Database are placed into as a separate record. For instance, in the TOD/ARAMIS System there are 422 such records.

This file makes data readily available for individual study, for detecting correlations and interactions between variables, for comparison of therapeutic interventions with changes in clinical course or laboratory values, for life-table outcome, analysis, for charting distribution of values within a given patient population, and for a host of additional and potential search programs.

TOD Schema File

The schema is the first file in the TOD System. It is a map which defines each database and differentiates it from other TOD databases. The schema is the direct reflection of the user's research concerns, and therefore must be designed carefully in order for the resulting database to meet the user's needs.

Underlying all TOD data handling is the descriptor file. The descriptor file is a machine-readable file, derived from a schema defined by the data bank user, which serves as a template for the stratifying, intervention, or outcome variables, for which values are to be collected in the physician-patient encounter.

All TOD programs reference the schema to get information for the meaning of the different elements. Therefore, the schema must exist in a machine-readable form. The process of developing a schema involves creating a human-readable text file in the computer which syntatically describes each element in the schema completely. Then a TOD program called TRANSLAT creates from the schema source file a machine-readable file, called the descriptor file. Another program lists the descriptor file, producing a document which reflects the contents of the operating schema.

Storage of data in a TOD database is cost-effective because the data files contain only data values. The meaning of those values is provided through the schema. For example, if the value of Parameter 27 for Patient 123 at Visit 7 is 42.6, the schema could be consulted to discover that Parameter 27 is defined as Hematocrit.

Header and Parameter Elements

Information is then entered interactively from the patient chart which is patterned after the descriptor file by means of the entry program which references the descriptor file. Patient information is collected into one of two files, termed "header" or "parameter", in accordance with whether the data are demographic and, hence has only the patient number as its ruling part, or subject to change over time, so that the ruling part consists of both the patient number and the visit date.

Header elements are those defined once for each patient; such as name, birth date, sex, etc. Note that although some variables, such as the patient's address may occasionally change, it is normally not necessary to preserve these changes through time; thus, the street address is another typical header element. Often a database owner will also want to keep certain dates as header elements, such as the date of the first symptom of the disease, the date of the first diagnosis, and so forth.

Parameter elements are those items recorded at each visit, so that multiple values of the element exist for each patient. Parameter elements must be numeric and should be amenable to statistical analysis.

Data Types

There are seven ways of representing clinical information in TOD, as listed. These types were developed to best describe elements of the disease process. The data types in TOD are: VALUE, CHARACTER, +RANGE, DATE, DISCRETE, CODED, and OCTAL. A CONFIDENTIAL datatype remains unused.

VALUE specifies a continuous variable (such as a laboratory test) which may have any degree of precision: serum creatine = 1.4 .

+RANGE elements are assumed to be semi-quantitative disease descriptors as are commonly utilized to describe degrees of severity or abnormality on a scale of 0 to 4+ : wrist pain = 3+ .

DISCRETE is employed for integer values: number-of-pregnancies = 2 . This definition limits certain computational procedures.

CHARACTER type provides for the representation of textual data in the computer: name = Susan Jones . Computation capability does not exist for such data, since the encoding is not controlled. The data type CODED provides for controlled strings. Only references to CHARACTER strings are kept in visit files, so that computation is not delayed by large character entries. The character strings themselves are kept on a remote file. The current ARAMIS implementation limits character type attributes to header items. in TOD as "1" and "0" for male and female, respectively.

CODED variables are ways of keeping attributes with controlled definitions in the system. The terms are keot and internally assigned to numeric values for compactness and limited computation. For example, "sex" wil be internally represented in TOD as "1" and "0" for male and female, respectively. By having a coded data type for this element, the entry clerk would not need to remember that "0" stands for female, but could type female or F and have the system understand that this should be represented by "0".

DATE allows the capture and computation of times: birthdate = 30JUL67 and age = TODAY - birthdate are permitted entry forms, but computation of differences and intervals is possible on the internal form.

OCTAL is a means whereby eight integers may be stored under one variable. Through this representation on the Stanford Orvyl file system, the use of storage and, hence, cost may be decreased.

Programs

The Chart Dump Program is the only retrieval program which uses the main

Programs

The Chart Dump Program is the only retrieval program which uses the main files, and is used to reconstitute a patient record for use within the clinic setting. The Transpose Program creates, from the main header and parameter files, the files used for retrieval. These are organized by individual variable. Rapid and efficient retrieval in the traditional scientific formats occurs from these files with reference to the descriptor file. The Subset Program creates a library of patients meeting user specified criteria, and this library is accessed by the other retrieval programs in studying defined subsets of the accumulated data.

The following table lists some of the currently available retrieval programs in TOD. When asking a research question, the individual TOD investigator will typically choose several programs.

RETRIEVAL PROGRAMS

PROFILE	-	Histographic Distribution
CRITTER	-	Diagnostic Criteria Counts
SCATTER	-	X-Y Graph of Variables
MULTREVU	-	Mean and SE of Variables in Subsets
OUTCOME	-	Life Table Analysis of Variables
AUTOSET	-	Computer Consultation
LIST	-	List of Variables in a Subset
TIMESCAT	-	X-Y Graph of Variable Over Time
SUBSET	-	Group Meeting Specified Criteria
MVV	-	Ranking Variables by Logistic Regression

Thus, one might execute a Profile analysis on a given population to ascertain the presence of sub-populations for further subsetting by the Subset Program. If this pertains, a Multrevu, which examines the elements of the Uniform Database between sub-populations and looks for differences in mean values between groups, might be selected. Or an Outcome, utilizing life-table analysis, might detect differences in prognosis between groups over time. The dynamic nature of the system allows for interactive reformulation of the evolving research question. A typical retrieval program requires less than a second of interactive computer time and less than three minutes of investigator time, including problem specification and output printing.

IV.D HMO Support: COSTAR 5

[This material is based on the Functional Description of the COSTAR 5 System. The Laboratory of Computer Science, Massachusetts General Hospital, Cambridge, Mass. gave permission to use this copy-righted material here.]

INTRODUCTION AND OBJECTIVES

COSTAR (COmputer-STored Ambulatory Record) is a computer-based ambulatory information system which improves and expands upon the capabilities of a traditional medical record. Although use of the term "record" has historical precedence, COSTAR is more appropriately considered an information and communication SYSTEM designed to meet both the medical care and financial/administrative needs of either a fee-for-service or a prepaid group practice health maintenance organization (HMO).

The central objectives of COSTAR are to:

1) Facilitate patient care by improving the availability of medical information in terms of accessibility, timeliness of retrieval, legibility, and organization.

2) Enhance the financial viability of the medical practice by providing a comprehensive billing system with accompanying accounting reports.

3) Facilitate medical practice administration by providing the data retrieval and analysis capability required by management for day to day operation, budgeting, and planning.

4) Provide data processing support for administrative and ancillary services e.g., scheduling, laboratories and planning.

5) Provide the capability to generate standardized management reports and support manager-specified inquiry and report-generation on any elements of the database.

6) Support programs of quality assurance by monitoring the content of the database according to physician-specified rules and to report automatically any deviations from these standards of care.

COSTAR is designed to have minimal impact on the physician's habit pattern of recording information. It is also designed to be the SINGLE integrated information system for the practice and thus to eliminate duplicate data recording and duplicate information processing for medical care, billing, and administration. All data are collected by the medical staff using specifically designed forms and are entered into the system by clerical staff through simple interactions with the computer using video terminals directly connected to the computer. The collected data are stored on magnetic disks so that the information is always available and simultaneously accessible at multiple locations.

SYSTEM MODULES

The basic design of COSTAR presents a modular system with a large variety of available options, allowing the system to be customized to the specific needs of each group practice. The modules which are basic to the system are:

SECURITY AND INTEGRITY MODULE. These routines, which are an integral part of all modules, provide for identifying and logging in/out all terminals and users for the purpose of preventing unauthorized access to medical and administrative information. The module will also provide the support routines to monitor the functioning of the system, provide transaction logging, and prevent data loss in case of machine failure.

REGISTRATION MODULE. These interactive routines are used for the entry and review of all identification date, demographic, insurance, and administrative, for each patient and family. It is possible for the practice to select the items to be collected in the registration sequence from a large menu of pre-coded fields. If necessary, the practice may also define additional registration items.

SCHEDULE KEEPING MODULE. This set of routines allows on-line booking and cancellation of appointments, review of current appointments, and production of legible, accurate schedules and day sheets. For the scheduling of non-members or new patients a minimal registration sequence is available.

MEDICAL RECORD MODULE. This series of routines provides for data input from encounter forms, and accessibility to the total medical and administrative database. Direct inquiry into this database can be done through computer terminals. The computer-generated medical record is made available for each scheduled visit. This module represents the core of the information system and provides a large variety of options for recording, manipulating, organizing and displaying the data.

BILLING AND ACCOUNTS RECEIVABLE MODULE. This subsystem uses the patient identification data captured on the encounter form to prepare monthly statements for each account and, at practice option, to produce superbills and third party claim forms. Complete accounts receivable audit trails are maintained and a wide selection of accounting reports are available.

MANAGEMENT REPORTING MODULE. These routines provide pre-programmed, standard reports (Densen'tables [Densen72], utilization and membership reports, and revenue analysis reports.) This module also allows the practice to specify the parameters for search routines which operate on the database to produce patient listings, and standardized tabulations and cross-tabulations.

CARDINAL COSTAR FEATURES

We will present the Medical Records Module in more detail below. In the classical, hand-written medical record, a provider has almost unlimited freedom of expression, since the record begins with a blank sheet of paper. In contrast, there are certain procedural rules in COSTAR which inherently constrain medical recording practices. On the other hand, because of these restrictions, and because COSTAR is a data-based system rather than a document-based system, there are significant advantages in terms of accessibility of recorded information. The procedures and advantages which most clearly differentiate COSTAR from a manual medical record system are:

The practice must enter at least a minimal set of registration data on each patient. This provides a single data file which is always available for patient identification, insurance and billing information, family linkage, and demographic information. This file can be accessed by authorized users at remote terminals by either patient number or alphabetic look-up.

Data are collected at each patient visit by recording both administrative and medical information on a form which is specifically designed for the needs of the particular medical group and/or speciality. This ENCOUNTER FORM is the single source document which is designed to capture all data which providers find necessary, and routinely collect, in clinical practice. This recording technique facilitates practice efficiency, cost-effectiveness, and data integrity in that the data from this single document supplies the multiple needs of medical records, accounts receivable, management reporting, quality assurance, medical audit, and research. The ENCOUNTER FORM provides for the recording of information in a structured format so that each particular type of datum (e.g., telephone number, medication) is uniquely identified. The encounter form is a self-encoding check-list; the important data elements at each encounter (e.g., names of diagnoses, medications, procedures, and laboratory tests) are recorded by the provider, who checks the appropriate item on the form. Next to the box for the checkmark a 5-character code has been printed. Within the computer-stored database all information is organized and accessed by the designated code. Detailed information concerning the particulars of the diagnoses, therapies, test results, etc., are recorded in narrative text (using either hand-written notes on the encounter form or associated dictation). However, this narrative information is linked to the encoded information and is always accessed and displayed with this code.

Medical record data are provided by a computer-generated printed output for routine (i.e., scheduled) patient-care. In COSTAR, the computer always prints the most up-to-date information. Several printed copies of the patient's record may be simultaneously available in different locations. In contrast to a hand-written record, COSTAR is not restricted to displaying medical information in the temporal sequence or form in which it was entered; instead, the computer is programmed to select the pertinent subset of the data and present this information in different formats according to the needs of different specialties. The organization of the computer-generated output emphasizes medical data. The objective of the organization is to present the information in a form that facilitates scanning of the relevant data in a minimal period of time. Since this information is always stored in the computer's files, the computer output can be discarded after use.

COSTAR enables immediate access to all patient and administrative information through use of one of the computer terminals. Direct inquiry into the database is simple and rapid; the user responds to a series of questions posed by the computer, entering on the keyboard the patient's name or identification number and the type of information desired. Although all the data are directly accessible, the user may choose to examine only certain information such as telephone number, most recent visit note, latest laboratory test values, etc., or may direct the computer to present the information as a flowchart of particular types of data (e.g., all blood pressures displayed together with all cardiovascular medications and serum potassiums). The user interaction at the computer terminal may be an iterative series of requests resulting in a series of different displays of patient data. This process is a greatly extended analog of leafing through a written medical record.

COSTAR is an ACTIVE or RESPONSIVE system in that the processing and display of information is a function of the content of the data. Because much of the record is coded, it is possible for the computer to tailor the output according to the characteristics of the individual patient and of the care which has been given. This contrasts with a manual system which is a completely passive archival system and therefore insensitive to the content, meaning, or significance of the information. The ability of COSTAR to "understand" the encoded data makes it possible for the physicians to develop automated programs for quality assurance. Computer programs can be written to monitor the recorded care of every patient according to standards of care defined by the particular group practice. Whenever a deviation from the standard occurs, COSTAR can automatically notify the appropriate physician or nurse, allowing corrective action to be taken for that particular patient care situation. Active surveillance and automatic feedback are two features of COSTAR which cannot be easily duplicated in a manual system and which represent unique additional capabilities for facilitating patient care.

COSTAR provides a capability for easy analysis of the database, either through standardized management reporting programs or via programs which allow user specification of search strategies and reports. (A major weakness of manual medical record systems is that it is costly in personnel time to perform aggregate data analysis of groups of patients; a similar weakness of most automated accounting systems is that the data are unavailable except through standard pre-specified reports. COSTAR provides an interactive language to allow the non-programmer to generate a variety of analysis routines or reports by a simple specification of the search parameters needed to select the desired groups of patients, and a report generator program to permit the user specification of the listings, tabulations or cross-tabulations desired.

COSTAR has been designed to be adaptable to a variety of practice settings. The system can be tailored to the needs of a specific practice because of its modular construction, and because of its extensive use of directories as the method of defining the structure and content of the record. Suggested content for these directories (diagnostic codes and modifiers, medication terms, laboratory test normal values) are supplied with COSTAR, but the content may be easily modified or extended by each practice. This permits each practice to take advantage of the COSTAR system, and yet individualize actual operation to meet local needs.

PROVIDER EDUCATION

One of the dominant limiting factors in the application of computer technology to medical practice is the necessity of achieving physician acceptance. COSTAR provides a buffer between the physician and the computer technology by having the physician record on paper forms which are then input into the computer system by clerical personnel. The main difference in recording practice introduced by COSTAR is that information must be recorded in specific patterns associated with a single medical entity, e.g., all the clinical findings associated with a specific disease must be associated with the code for that disease. This philosophy of record-keeping is a modified form of "problem-orientation" and seems to be acceptable to a wide variety of practices. COSTAR is designed so that the medical staff can learn to use the system after only a few minutes to an hour of explanation. However for a practice to take full advantage of all the features of COSTAR, it is necessary that there be a more extensive period of provider education, since some of the concepts such as the "status" of a diagnosis, or the recording

of modifiers for a diagnosis are unique in comparison to standard medical record practices. In general, the more the medical staff understands the functional capability of COSTAR, the more the system can be used to improve record keeping and patient care.

Interaction at the display terminal

Interaction at the terminal is designed to be simple and consistent. For instance the user is frequently offered the option to edit a previous response. In all such cases, the old information is presented, enclosed in carets, e.g., <OLD INFORMATION>, and the new information may be entered immediately following the ">" symbol. If the current telephone number is incorrect, the editing would be accomplished as follows:

TELEPHONE <821-3114>821-3141

In this case, the user has typed the correct number followed by pressing the ENTER key. Retaining the old information is accomplished simply by pressing the ENTER key without entering any new information.

In some cases, editing consists of removing existing information from the patient file. For example, if a patient is no longer employed, the office telephone number should be deleted from the file. This is done by entering a minus sign "-".

OFFICE TELEPHONE <965-0811>-

By using the minus sign, the data for the field OFFICE TELEPHONE has been deleted from this patient's record.

The Medical Record

The COSTAR Medical Records module is designed to provide the medical practice with timely and legible medical records. COSTAR improves the accessibility of medical data by optimizing: a) the availability of the information; b) the appropriateness of organization; and c) the style of presentation.

The encounter form is the primary document used to record medical data. The information on the form is structured by data type; and each element is associated with a code, the date of collection, and the name of the provider involved. Because of this structure and this coding, it is possible for COSTAR to generate output which highlights the important components of the medical information and which is tailored to the needs of the particular specialty for which the record is being generated.

The computer generated output is used for the routine care of scheduled patients, for walk-in patients, for telephone calls, for follow-up of selected patients, for patients selected as being of particular concern by the automated quality assurance studies, for consultations between different providers, and for transmission to other physicians, hospitals, or insurance companies.

Because of the COSTAR structure and coding, it is also possible to use the database forquality assurance, for medical audit, for descriptive studies of the patient population (in terms of patterns of disease, treatments given, and outcomes) and for medical research. Although the codes are unique to COSTAR, the system contains translation tables which can convert COSTAR codes to the coding system required by the particular third party carrier.

Output Documents

There are three different types of output provided by COSTAR: ENCOUNTER REPORT, FLOWCHART, and STATUS REPORT. For scheduled visits a combination of these three different types is routinely generated, depending on the particular needs of the specialty and the practice.

Encounter Report. This computer-generated report is equivalent to a medical "note" reflecting the activity at a single visit or encounter with a patient. This ENCOUNTER REPORT displays in a standard format both those data collected and entered via single encounter form, and the data reflecting laboratory test results associated with that encounter, which may be entered separately. The data from each such encounter form may be retrieved as an encounter report, which is identified by patient, date and provider. The encounter report displays all codes entered on that visit, with the status flag(s), modifier(s), associated text and/or results(s). All data are presented in the following sequence:

A. Patient Identification
B. Encounter Identification
C. Provider Name(s)
D. Disposition(s)
E. Diagnoses and Problems
F. Physical Examination Data
G. Medications and Therapies
H. Procedures
I. Laboratory Tests
J. Administrative Data

Flowcharts. This form of computer-generated report emphasizes the temporal course of the disease process or the variation in clinical findings over time. The display is a chronological listing, by date, of all occurrences of particular coded items with associated text and or results. The medical practice may create any number of flowchart format "templates" that specify which COSTAR codes are to be displayed, and the output format of the report. The template is organized by columns. Each column may include one or more codes. For example, the report generated from a sample template intended for follow-up of hypertensive patients contains colums labeled WGT, BLOOD PRESSURE, CREA, URIC ACID, and K+; and includes all statuses, results, and textual information associated with the COSTAR codes for weight, blood pressure, serum creatinine, serum uric acid and serum potassium respectively. The column marked MEDICATIONS includes many anti-hypertensive drugs. When mulltiple codes are specified for one column the name of each code is given in the flowchart.

A template may have any number of associated "trigger" codes. The presence of any one or more of these codes in a patient record will cause the corresponding flowchart to be generated whenever a STATUS REPORT is printed. For example, the diagnosis of hypertension could be given as a trigger for the hypertension flowchart template.

Status Report. The STATUS REPORT serves both as an index to the content of the computer-based medical record and also as a summary of the most recently collected data. The STATUS REPORT consists of seven components: The HEADER information which contains the patient identifying information, demographic and personal data.

The DISPOSITION information which represents the plan (e.g., future appointments) with this patient.

The DIAGNOSIS or Problem information which represents the medical assessment entered by the physicians or nurses who have given care to this patient.

The PHYSICAL EXAM data (e.g., vital signs) which are available on this patient.

The THERAPY section which lists the Medications and Therapies prescribed for this patient.

The PROCEDURE section which lists operations, immunizations, invasive tests, etc.

The TEST RESULTS section which lists the most recent test results on the patient. Abnormal test results are flagged with an asterisk. If the patient has ever had an abnormal result for a particular test, a flowchart of the five most recent results for that test is displayed.

The STATUS REPORT gives the date the particular medical item was first mentioned (e.g., the date a specific diagnostic term was first used), the number of encounters at which the item had been checked, and the last date at which the particular item was mentioned. Detailed information, as free text, is given only for the most recent instance in which such text was associated with that particular code.

System Summary

COSTAR is programmed in Standard MUMPS and can be supported by any computer system configuration that supports Standard MUMPS. COSTAR is designed to take advantagee of the recent advances in computer hardware technology which have resulted in a dramatic reduction in the cost of the computer processor and disk storage. It is anticipated that in most practices, COSTAR will be an in-house system with a variety of peripherals and storage capability based upon practice needs and requirements. Depending upon the size of the system and the configuration chosen the system cost will probably be between $75,000 and $200,000 - with the smaller system being appropriate for a small group practice (e.g., five physicians) and the larger systems being used for practices of 15 or more physicians. This price should make COSTAR a cost-effective alternative for many offices currently using manual or partial service-bureau systems. The computer configuration on which COSTAR is now being implemented is a Digital Equipment Corporation PDP-11.

IV.E Hospital Systems: PATIENT ORDER MANAGEMENT AND COMMUNICATION SYSTEM (Used at the Coral Gables Variety Childrens Hospital)

[This material is based on descriptive material provided by Dynamic Control and IBM.]

The Patient Order Management and Communication System (POMCS) was developed as a hospital-wide computer information system that provides a communication link between the admissions office, nursing stations, ancillaries, and the accounting department. It was developed by Dynamic Control Corporation of Coral Gables, Florida and is installed at Variety Childrens Hospital to facilitate, expedite, and integrate the delivery of health care services and the operation of the hospital. This hospital is a 188 bed facility with an average stay of 9 days. Doctor's orders are entered through video displays at the nursing stations. The orders are automatically transmitted to the appropriate ancillary departments and added to the patients record. These departments have the ability to display work to be done and to enter results. Results may then be transmitted back to the nursing station in hardcopy form to become a part of the chart. The result also becomes a part of the patient's record and is available for inquiry through a video display to authorized personnel. Cumulative summaries may also be produced. Charges for patient services are automatically collected and made available to the accounting system.

The system now provides support for the ancillary departments at Variety Childrens Hospital. The system is designed to be utilized in a modular fashion, so if installed at another hospital it can support those departments deemed necessary there.

OBJECTIVES

POMCS has as objectives to provide revenue increases, to lead to increased personnel productivity, cost savings, and improved patient care quality, through the following:

Revenue Increases

1. Automatic generation of patient charges and control
2. Reduced forms cost
3. Accountability of floor stock
4. Significantly reduced lost and late charges
5. Automatic pricing and control of prices
6. On-line census information for improved bed utilization
7. Charges can be generated immediately for outpatients and receipts can be entered on-line.

Personnel Productivity and Cost Savings

1. Reduction of clerical activity for nursing by reducing the need to transcribe orders to multiple working documents
2. Entry costs due to the use of display menus for data collection
3. Standard ordering procedures
4. Automatic communication of test results to nursing units
5. Improved document legibility
6. Errors resulting from order transciption can be reduced

Patient Care Quality

1. More nursing time for patient care, through reduction of clerical work.
2. Improved control of order status from order time to receipt of result.
3. More expedient order processing by eliminating the transfer of forms between requester and provider.
4. Cumulative result summaries for the medical record.
5. Duplicate orders can be eliminated.
6. Inquiry as required to determine order status and/or results.
7. Reduce order rejects because of legibility.

SYSTEM DESCRIPTION

POMCS uses an IBM System 32 to support a network of interactive display terminals and printers which link nursing stations and the major service areas of the hospital. The Patient Order Management and Communication System is written in RPG II.

FUNCTIONS

POMCS includes four major on-line interactive functions:

1. Admission/Discharge/Transfer
2. Order Communications
3. Patient Billing & Accounts Receivable
4. Outpatient/Emergency Room Registration

The program modules which implement these functions are described below.

ON-LINE PRE-ADMISSION AND ADMISSION, DISCHARGE, TRANSFER, AND CENSUS: Patients are preadmitted, admitted, transferred, or discharged using display terminals via interactive screen data entry and editing. Patient data is entered only once - at pre-admission time. Demand census from a display station provides current information on patient location.

ON-LINE OUTPATIENT REGISTRATION: The system provides outpatient and emergency room registration for prior patients, as well as new patients. On-demand outpatient billing information is also available to facilitate collection from outpatients as they leave the facility.

ON-LINE ORDER ENTRY AND ORDER CONTROL: Orders may be entered through display terminals utilizing successive menus or from a single screen if service codes are known. The status of orders may be tracked from the time of entry until acknowledgement or result entry by the ancillary department occurs.

1. Order Entry: The use of standard assumed options and override capabilities permits fast, accurate, and individually tailored order information.
2. Order Status: Orders may be displayed for a patient or an ancillary department.
3. Order Acknowledgements can be placed in the system by the respective service departments upon receipt or completion of the ordered service. The order is then flagged as complete, charges are posted to the patient account and the transactions are passed to the billing system.

4. Order Result Entry: Certain orders may be coded as requiring results such as laboratory orders. In place of acknowledging the completion of the service, results may be entered using a number of user defined formats. Once entered, the results are printed at the patient's nursing unit.
5. Repeat orders need to be entered only once because the system will regenerate them for the time period required.
6. Order Maintenance: Orders requiring only an acknowledgement may be altered to change the service provided before the charge is entered.

ON-LINE RESULT REPORTING: The posting of results to various orders is provided. Once the results have been entered by the service department, they are printed immediately at the patient's nursing station. The results of that day are combined to produce daily and final cumulative reports. The form of these reports provide a tabular graph of similar results over time. Results may be displayed and corrected for a specific patient by the ancillary department, however the original result reported claims as part of the Patient Master Record for audit purposes.

ORDER ENTRY CHARGE COLLECTION: The Catalog file contains prices for patient services and indicates when to post the charge, i.e., at order time, acknowledgement time, or at result reporting time. For an entered order, the charge is automatically posted at the specified time.

ON-LINE FILE MAINTENANCE: The system contains facilities for on-line maintenance of the Doctor, Insurance, Catalog, Screen, and A/R files. Each ancillary department may update their own section of the Catelog file. Price changes are limited by a security code. Departments may also alter the format of their display screens as necessary. Each ancillary department should be responsible for maintenance of their files.

MESSAGE PROCESSING: During order processing, free form text messages such as "patient needs wheelchair" may be included. A message file is provided from which automatic messages can be generated with certain orders by having appropriate codes in the Catalog File.

PATIENT HISTORY FILE: A file is available in the system to maintain patient records. Should a patient enter the facility as an in- or outpatient, the data on the patient needs merely to be reviewed or updated before the system automatically prepares the new registration or admission forms.

PATIENT BILLING: The system includes the following billing functions:

Demand Bill Inquiry: The business office may display a patient's bill at anytime.

Outpatient Billing: Bills are printed daily to show the itemized charges for that day. Follow-up mailers can be printed as required.

Inpatient Billing: Final bills will usually be requested by the business office. Prorated patient bills for insurance coverage are printed after the final bill is requested.

ACCOUNTS RECEIVABLE: The Accounts Receivable functions in the system include Bad Debts Reporting, Patient Trial Balances, Insurance Company Reports, Cash Receipt Reports, and Patient Accounts Receivable Activity Reports.

PATIENT CENSUS: Patient by room census is available as an on-line function. Daily batch census reports can include ADT logs, alpha & numeric patient room reports.

SYSTEM SECURITY AND CONTROL: The basis of control in the system is a non-displayed operator security code. After signing on the system, the operator may perform only those functions permitted with his security code. Each program has its own allowable security code requirements. Cancellation of orders is controlled by the security code. All order entry transactions entered into the system are logged by date, time of entry, and identified with the security code of the person performing the transaction. All other transactions are logged by date and security code.

REPORTS: Some of the reports which can be generated by the system include:

Patient lists
Doctor lists
Lab Flow Sheets
Order Result Summaries
Orders outstanding by department, nursing station and patient
Patient Bills and Billing reports
Outpatient reports, including patients registered during the day with no charges entered
Patient Census reports
Bad debt reports
Cash receipts reports
Patient Trial Balances
Insurance Company Accounts Receivable
Transaction journals
Department journals
Admission forms

SUMMARY

This system provides an example of the new generation of hospital information systems. It runs on a relatively inexpensive computer, so that it is feasible for the hospital to own the equipment, and hence control its expenditures to a large extent. The hospital has no programming personnel and relies wholly on a software vendor. Once a basic and reliable operation is established system improvements have to be negotiated with the vendor. While this causes some delay, it also makes the hospital administration aware of the costs associated with new and changed software specification, a problem commonly underestimated when software is written and maintained in-house.

IV. REFERENCES

AAOD78 SNODERM: Systematized Nomenclature of Dermatology; American Academy of Dermatology and College of American Pathologists, Crawford & Brown, 1978.

Achten75 Achtenberg, J., J.P. Miller, P.E. Cryer, and V.V. Weldon, "A Diabetes Center Patient Registry"; Proc. 1975 MUMP User's Group Meeting.

ACP69 SNOP, Systemized Nomenclature of Pathology; American College of Pathologists, Skokie, IL, 1969.

ACM76 ACM Surveys: Special Issue; Data Base Management Systems; ACM Vol. 8, no. 1, 1976.

Anders77 Anderson, J.: "Clinical Computer Information Systems in the United Kingdom"; J. of Med. Sys., Vol. 1, no. 4, pp. 375-383, 1977.

Austin78 Austin, C.J. and B.R. Greene: "Hospital Information Systems: a Current Perspective"; Inquiry, Vol 15, no. 2, pp. 95-112, 1978.

Barnet76 Barnett, G.O., D.E. Souder, J.E. Bowie, N. Justice: "MUMPS: A Support for Medical Information Systems"; Medical Informatics, Vol. 1, pp. 183-189, 1976.

Barnet78 Barnett, G.O., et al: "Quality Assurance Through Automated and Concurrent Feedback Using a Computer-Based Medical Information System"; Medical Care, Vol. 16, pp. 962-970.

Barnet79 Barnett, G.O., N.S. Justice, et al: "COSTAR: A Computer-based Corporation, Medical Information System for Ambulatory Care"; Proc. of the IEEE, Vol. 67, no. 9, September 1979, pp. 1226-1237.

Barn&J78 Barnett, G.O., N. Justice, et al: "COSTAR-A Computer-Based Medical Information System for Ambulatory Care"; Proceedings of the second Annual Symposium on Computer Applications in Medical Care, IEEE, Washington, D.C. pp. 486-487, 1978.

Barn&W78 Barnett, G.O., R.N. Winickoff, J.L.Dorsey, et al: "Quality Assurance through Automated Monitoring and Concurrent Feedback using a Computer-based Medical Information System"; Medical Care, Vol. 16, pp. 962-969, 1978.

Barsam79 Barsamian, H.: "Technology and Architecture for Data Management: An Overview and Perspective"; IEEE Computer Vol. 15, no. 5, pp. 8-10, May 1979.

Baskin78 Baskin, A.B. and A.H. Levy: "MEDIKAS--An Interactive Knowledge Acquisition System"; Proceedings of Second Annual Symposium on Computer Applications in Medical Care, IEEE, New York, pp. 344-350, 1978.

Beckle77 Beckley, R.F. and H.L. Bleich: "Paper Chase: A Computer-Based Reprint Retrieval System"; Comp. Biomed. Res., Vol. 10, pp. 423-430, 1977.

BCL78 Batelle Columbus Laboratories: "Technical report (Phases I-III) on Development and Implementation of the Laboratory Animal Data Bank"; submitted to the National Library of Medicine, Bethesda Md, Contract NO1-LM-5-4747, 1978.

Bharga74 Bhargava, B.K. and C.J. McDonald: "Tree Systems for Medical Information Processing "; Biomedical Sci. Instrum., Vol. 10, pp. 93:99, 1974.

Bleich72 Bleich, H.L.: "Computer-based Consultation - Electrolyte and Acid-base Disorders"; Am. Jour. Med., Vol. 53, pp. 285-291, 1972.

Blum78 Blum, R.L. and G. Wiederhold: "Inferring Knowledge from Clinical Data Banks Utilizing Techniques from Artificial Intelligence"; Proceedings of the Second Annual Symposium on Computer Applications in Medical Care, Washington, D.C., IEEE, pp. 303-307, 1978.

Blum&L78 Blum, B.I. and R.E. Lenhard, jr.: "Privacy and Security in an Oncology Information System"; Proceedings of the Second Annual Symposium on Computer Application in Medical Care, Washington, D.C., IEEE, pp. 508-508, 1978.

Blum79 Blum, B.I. and R.E. Lenhard, jr.: "An Oncology Clinical Information System"; Computer (IEEE), Vol. 12, no. 11, November 1979, pp. 42-58.

Bolour79 Bolour, A.: "The Process Model of Data"; Tech. Rep. 38, Medical Information Science, Univ. of Calif., San Francisco, March 1979.

Bowie76 Bowie, J.E. and G.O. Barnett: "MUMPS - An Economical and Efficient Time-Sharing System for Information Management"; Computer Programs in Biomedicine, Vol. 6, pp. 11-22, 1976.

Bowie77 Bowie, J.E.: "Methods of Implementation of the MUMPS Global Database"; Proc. of the MUG of Japan Meeting, Nagoya, Japan, 1977.

Brande76 Brandejs, J.F.: Health Informatics - Canadian Experience; North Holland Publishing, Company, New York, 1976.

Breslau76 Breslau, N. and M.R. Haug: "Service Delivery Structure and Continuity of Care: A Case Study of Pediatric Practice in Process of Reorganization"; Journal of Health and Social Behavior, Vol. 17, pp. 339-352, 1976.

Breslo75 Breslow, N. and J. Eustrom: "Geographical Correlations between Cancer Mortality Rates and Alcohol-Tobacco Consumption in the United States"; Journal of the Nat. Canc. Inst., Vol. 53, pp. 631-639, 1975.

BrookW76 Brook, R.H., K.H. Williams, and A.V. Avery: "Quality Assurance Today and Tomorrow: Forecast for the Future"; Annals of Internal Medicine, Vol. 85, pp. 809-817, Dec. 1976.

Brown78 Brown, B.W., J. McConathy, M. Meigs, B. Munder, and M. Schimek: "File Arrangement and Maintenance Executive (FAME), A General Purpose Schema Driven File Maintenance System"; Comput. Biol. Med., Vol. 8, pp. 159-172, 1978.

Brunje77 Brunjes, S.: "An Anamnestic Matrix Toward a Medical Language"; Comput. Biomed. Res., Vol. 4, pp. 571-584, 1971.

Bruce71 Bruce, R.A., et al: "Non-Invasive Predictors of Sudden Cardiac Failure"; Cardiology, Vol. 39, pp. 833-840, 1977.

Buck74 Buck, C.R. and K.L. White: "Impact of a System Based on Billing Claims"; New England Journal of Med., Vol. 291, pp. 877-883, 1974.

Buckho78 Buckholtz, T.J.: "Large-scale Real-time Document Distribution"; Proceedings 3rd USA-Japan Conf., AFIPS-IPSJ, pp. 501-504, 1978.

Bush45 Bush, Vannevar: "As We May Think"; Atlantic Monthly, Vol. 17 no. 1, pp. 101-108, 1945.

Cardon78 Cardoni, A.A. and T.J. Thompson: "Impact of Drug Information Services on Patient Care"; Amer. Jour. of Hosp. Pharm., Vol. 35, pp. 1233-1237, 1978.

Castle74 Castleman, P.A.: "The CAPO Project Phase I - Evaluation of an Automated Medical History in Office Practice"; HRA 110-71-2244, PB 294 665, Bolt, Beranek and Newman, Inc., 1974.

Cast&W74 Castleman, P.A., F.N. Webb, C.A. Hollister, J.R. Siegel, S.R. Zdonik, and D.M. Fran: "The Implementation of the Prophet System"; Proceedings of the 1974 NCC, AFIPS Vol. 43, pp. 457-468, 1974.

CCT79 Controlled Clinical Trials, Design and Methods; North Holland Publishing, journal starting 1979.

Chang76 Chang, S.K. and B.H. McCormick: "Intelligent Coupling of the User to Medical Information Systems"; Proceedings of the Third Illinois Conference on Med. Inf. Systems, pp. 29-46, Univ. of Illinois at Chicago Circle, 1976.

Chan&O76 Chang, S., M. O'Brien, J. Read, R. Borovedc, W.H. Cheng, and J.S. Ke: "Design Considerations of a Database System in a Clinical Network Environment"; Proc. 1976 National Computer Conference, AFIPS, Vol. 46, pp. 277-286, 1976.

Codd70 Codd, E.F.: "A Relational Model for Large Shared Data Banks"; Comm. of the ACM, Vol. 13 no. 6, pp. 377-387, 1970.

Codd79 Codd, E.F.: "Extending the Database Relational Model to Capture More Meaning"; ACM Transactions on Databases, Vol. 4, no. 4, Dec. 1979, pp. 397:434.

Collen74 Collen, M.F.: Hospital Computer Systems; John Wiley & Sons, New York, 1974.

Cote77 Cote, R.D. (Ed.): Systematized Nomenclature of Medicine; College of Am. Pathologists, Skokie, Ill., 1977.

Davis70 Davis, L.S.: "Prototype for Future Computer Medical Records Comput. Biomed. Res., Vol. 3, no. 5, pp. 539-554, 1970.

Davis78 Davis, R.: "A DSS for Diagnosis and Therapy"; ACM SIGBDP, Vol. 8, no. 3, 1978.

Densen72 Densen, P.M.: Guidelines for Producing Uniform Data for Health Care Plans; US DHEW, Washington, D.C., 1972.

DeMil178 DeMillo, R.A., D.P. Dobkin, A.K. Jones, and R.J. Lipton: Foundations of Secure Computation; Academic Press, Inc., 1978.

Donabe78 Donabedian, A: "Needed Research in the Assessment and Monitoring of the Quality of Medical Care"; NCHSR Research Digest Series, DHEW, Pub. no. (PHS) 78-3219, 1978.

Duncan78 Duncan, K.A.: "A Total System for Dentistry"; in Information Technology in Health Science Education, E. Deland, Ed., Plenum Publishing Co., 1978.

ElMasr79 El-Masri, R. and G. Wiederhold: "Data Model Integration Using the Structural Model"; Proceedings of the ACM-SIGMOD Conference, pp. 191-202, 1979.

Enthov78 Enthoven, A.D.: "The Consumer Choice Health Plan"; New England Journal of Medicine, Vol. 298, March 23 & 30, 1978, pp. 650-658 & 709-720.

Epstei78 Epstein, M.N.: "Natural Language Access to a Melanoma Data Base"; Proceedings of Second Annual Symposium on Computer Applications in Medical Care, Washington, D.C., IEEE, New York, pp. 320-327, 1978.

FDA78 Federal Drug Administration, Dept. of HEW, Bureau of Veterinary Medicine: "Laboratory Practice for Nonclinical Laboratory Studies"; Federal Register Vol. 43, no. 247, pp. 59986-60020, Dec. 22, 1978.

Feigl79 Feigl, P.: "The Collection of Standardized Patient Data in US Comprehensive Cancer Centers"; Cancer Campaign, Vol. 3, Gustav Fisher Verlag, Stuttgart, p. 9, 1979.

Feins70A Feinstein, A.R.: "Taxonorics"; Archives of Internal Medicine, Vol. 126, pp. 679-693 (October), pp. 1053-1067 (December), 1970.

Feins70B Feinstein, A.R.: "Quality of Data in the Medical Record"; Comput. Biomed, Res., Vol., 3, pp. 426-435, 1970.

Fenna78 Fenna, D., S. Abrahamson, S.O. Loow, and H. Peterson: The Stockholm County Medical Information System; Springer Verlag 1978.

Friedl73 Friedlander, J.: "Clinical Search for Information"; Journal of the American Society for Information Science, pp. 65-69, Jan-Feb., 1973.

Friedm77 Friedman, R.B. and D.H. Gustafson: "Computers in Clinical Medicine, A Critical Review"; Comput. Biomed. Res., Vol. 10, pp. 199-204, 1977.

Friedm78 Friedman, R.B., S.M. Entine, E.A. Anderson: "Laboratory Data and Clinical Decision Making"; Proceedings Fourth Illinois Conference on Medical Information Systems, pp. 31-38, 1978.

Fries72 Fries, J.F.: "Time Oriented Patient Records and a Computer Data Bank"; JAMA, Vol. 222, no. 12, pp. 1536-1554, 1972.

Fries74 Fries, J.F.: "Alternatives in Medical Record Formats"; Medical Care, Vol. 12, no. 10, 1974.

Gabrie78 Gabrieli, E.R.: "Knowledge Base Structures in a Medical Information System"; Proceedings of the 8th Annual Conference of the Society for Computer Medicine, pp. 1.2.9 to 1.2.11, 1978.

Gaus73 Gaus, C.R.: "Information Requirements for Management of New Health Care Organizations"; Medical Care, Supplement, Vol. 11, p. 51, 1973.

German75 Germano, F. Jr., and S. Weyl: "A Database Organization to Support a Time-Oriented Medical Record"; Proc. ACM Pacific 75 Conference, Boole and Babbage, Palo Alto, CA., 1975.

Giebin75 Giebink, G.A., et al: "Current Status of Ambulatory Health Care Computer Applications"; IEEE Computer, Vol. 18, pp. 28-32, 1975.

Gieb&H75 Giebink, G.A. and L.L. Hurst: Computer Projects in Health Care; Health Administration Press, Ann Arbor, 1975.

Given77 Given, C.W., et al: "The Use of Computer Generated Patient Profiles to Evaluate Resident Performance in Patient Care"; The Journal of Family Practice, Vol. 5, pp. 831-840, 1977.

Go78 Go, H.T.: "Development and Use of an Information System in Primary Care Education"; The Journal of Ambulatory Care Management, Vol. 1, pp. 75-87, 1978.

Greene69 Greenes, R.A., A.N. Pappalardo, C.W. Marble, and G.O. Barnett: "A System for Clinical Data Management"; Proceedings of the 1969 FJCC, AFIPS Vol. 35, pp. 297-305, 1969.

Greene78 Greenes, R.A., G.O. Barnett, S.W. Klein, A. Robbins, and R.E. Prior: "Recording, Retrieval and Review of Medical Data by Physicians/Computer Interaction"; New England Journal of Medicine, Vol. 282, pp. 307-315, 1978.

Greenf76 Greenfield, R.H.: "Clinical Data Base Usage"; Proceedings of the Third Illinois Conf. on Med. Inf. Systems, pp. 47-56, Univ. of Illinois at Chicago Circle, 1976.

Gree&K77 Greenfield, R.H., M.A. Kass, and J.P. Livingston: "A Computerized Glaucoma Data Base"; Arch. Ophthalmol, Vol. 95, no.8, August 1977, pp. 1365-1367.

Gremy77 Gremy, F.: "Medical and Hospital Data Processing in France"; J. of Med. Sys., Vol. 1, no. 4, pp. 385-393, 1977.

Groner78 Groner, C.F., M.D. Hopwood, N.A. Palley, et al: "An Interactive Data Management and Analysis System for Clincial Investigators"; J. Lab. Clin. Med., Vol 92, no. 3, pp. 325-340, 1978.

Haensz66 Haenszel, W. and W.I. Lourie, Jr.: "Quality Control of Data in a Large-Scale Cancer Register Program"; Methods Inf. Med., Vol. 5, no. 2, pp. 67-74, April 1966.

Hall76 Hall, W.D.: "The Defined Data Base"; in H.K. Walker, W.D. Hall, J.W. Hurst (Eds.), Clinical Methods. The History, Physical and Laboratory Examinations, pp. 13-16, Butterworths, 1976.

Hall77 Hall, P.: "Outstanding Computer Applications for Patient Care in Sweden"; J. of Med. Sys., Vol. 1, pp. 395-401, 1977.

Hammer78 Hammer, M. and D. McLeod: "The Semantic Data Model, A Modelling Mechanism for Data Base Applications", Proc. ACM SIGMOD, pp. 22-36, 1978.

Hammon73 Hammond, W.E., et al: "GEMISCH: Mini-Computer Information Support System"; Proc. of IEEE, Vol 61, pp. 1575-1583, 1973.

Hammon78 Hammond, W.E., W.W. Stead, M.J. Straube, and J.D. Collins: "Experience in the Development of an Ambulatory Care Medical Information System"; Proc. 3rd USA-Japan Conference, AFIPS-IPSJ, pp. 452-455, 1978.

Hanley78 Hanley, J.A.: "A Language for the Computer Generation of Medical Data Forms"; Biometrics, Vol. 34, no.2, June 1978,pp. 228-297.

Henley75 Henley, R.R., G. Wiederhold, J.V. Dervin, M.A. Jenkin, I.M. Kuhn, E. Mesel, D. Ramsey-Klee, and J.E. Rodnick: "An Analysis of Automated Ambulatory Medical Record Systems"; Vol. I, Findings and Vol. II, Background, Univ. of California, San Francisco, Office of Medical Information Systems, Technical Report #13, 1975. Available from NTIS (Springfield, VA 22151) as PB254234.

Hershe78 Hershey, J.C., D.H. Kropp, and I.M. Kuhn: "The Productivity Potential of Physicians' Assistants"; Journal of Medical Systems, Vol. 2, no. 2, 1978.

Hess74 Hess, E.V.: "A Standard Database for Rheumatic Disease"; Arthritis & Rheumatism, Vol. 17, pp. 327-336, 1974.

Hulse76 Hulse, R.K., S.J. Clark, J.C. Jackson, et al: "Computerized Medication Monitoring System"; Am. J. Hosp. Pharm., Vol. 33, pp. 1061-1064, 1976.

Jainz76 Jainz, M. and T. Risch (Eds.): "A Data Manager for the Health Information System Berlin"; Comp. Prog. in Biomed, Vol. 6, pp. 166-170, 1976.

Jelovs79 Jelovsek, F.R., R.E. Bolinger, R.E. Davis, J.M. Long, B.B. Oberst, R.A. Reid, and J.Zimmerman: "Guidelines for User Access to Computerized Medical Records"; J. of Med. Syst., vol.2, no.3, 1978.

Jewel79 Jewel, J.: "System Aids Spinal Injury Patients"; Mini-Micro Systems, Vol. 12, no. 11, Nov. 1979, pp.75-78.

JMS77 Journal of Medical Systems; Plenum Press, 1977.

Johns77 Johns, C.J., et al: "Minirecord: An Aid to Continuity of Care"; Johns Hopkins Medical Journal, Vol. 140, pp. 277-284, 1977.

Johnso79 Johnson, P.E., D.G. Severance, and P.J. Feltovich: "Design of Decision Support Systems in Medicine: Rationale and Principles from the Analysis of Physician Expertise"; Proceedings of the 12th Hawaii International Conference on System Sciences, B. Shriver, T.M. Walker, and R.H. Sprague, Jr., Western Periodicals Co., 1979.

Justic74 Justice, N., G.O. Barnett, R. Lurie, and W. Cass: "Development and Implementation of a Medical/Management Information System at the Harvard Community Health Plan"; Proc. 1974 National Computer Conference, AFIPS Vol. 43, pp. 159-165, 1974.

Karpin71 Karpinski, R.H.S. and H.L. Bleisch: "MISAR, A Miniature Information Storage and Retrieval System"; Comp. and Biomed. Res., Vol 4, pp. 655-660, 1971.

Katter75 Katter, R.V. and K.M. Pearson, Jr.: "MEDLARS II: A Third Generation Bibliographic Production System"; J. Libr. Auto., Vol. 8, pp. 87-97, 1975.

Katzje77 Katzjer, M. and N.P. Ross: "Analysis of Profiles in the Monitoring of Health Care Services"; MEDINFO 77, North Holland, pp. 17-21, 1977.

Korein71 Korein, J., M. Lyman, and L.J. Tick: "The Computerized Medical Record"; Bulletin N.Y. Acad. Med., Vol. 47, pp. 824-826, 1971.

Kreitz78 Kreitzberg, C. B, M. L. Stocking, and L. Swanson: "Computerized Adaptive Testing: Principles and Directions"; Comput. Educ., Vol. 2, no. 4, pp. 319-331, 1978.

Kronma78 Kronmal, R.A., K. Davis, D. Lloyd, R.A. Jones, and M.J. Gillespie: "Data Management for a Large Collaborative Clinical Trial. CASS: Coronary Artery Surgery Study"; Comput. Biomed. Res. Vol.11, no.6, pp. 553-566, Dec. 1978.

Kuliko77 Kulikowski, C.A.: "Problems in the Design of Knowledge Bases for Medical Consultation"; in Proceedings of the First Annual Symposium on Computer Application in Medical Care, IEEE, 1977.

Kunz79 Kunz, J.: "Index Medicus: A Century of Medical Citation"; JAMA, Vol. 241, pp. 301-398, 1979.

Lairso74 Lairson, P.D., J.C. Record, and J.C. James: "Physicians Assistants at Kaiser: Distinctive Patterns of Practice; Inquiry, pp. 207-219, September 1974.

Leavit72 Leavitt, S.S., T.L. Johnston, and R.D. Bever: "The Process of Recovery - Patterns in Industrial Back Injury"; Industrial Medicine and Surgery, Vol. 40, no. 8, 1971 and Vol. 41, no. 1, 1972

Lederb78 Lederberg, J.: "Digital Communications and the Conduct of Science: The New Literacy"; Proceedings of the IEEE, Vol. 66, pp. 1314, 1978.

Leiter77 Leiter, J.: "On-Line Systems at the National Library of Medicine"; Proceedings of MEDINFO 77, North-Holland, pp. 349-353, August, 1977.

Levy78 Levy, A.H. and A.B. Baskin: "Clinical Computing - 1977: A Selective Review and Perspective"; J. Med. Systems, Vol. 1, no. 4, pp. 361-374, 1978.

Lewis77 Lewis, T.L. and G.C. Macks: "Adaptation of a General Hospital Computerized Medical Information System to the Research Hospital Environment"; Proceedings of the First Annual Symposium on Computer Applications in Medical Care, IEEE, pp. 111-115, 1977.

Lindbe70 Lindberg, D.A.B.: "A Statewide Medical Information System"; Comp. Biomed. Res., Vol.3, pp. 453-463, 1970.

Lindbe79 Lindberg, D.A.B.: The Growth of Meedical Information Systems in the United States; Lexington Books 1979.

Logema73 Logemann, G.W.: "Multi-State Information System - Problems and Solutions"; Proc. ACM Annual Conf., pp. 271-272, 1973.

Lucas78 Lucas, H.C.: "The Use of an Interactive Storage and Retrieval System in Medical Research"; CACM, Vol. 21, pp. 197-205, 1978.

Luft78 Luft, H.S.: "How Do Health Maintenance Organizations Achieve Their 'Savings', Rhetoric or Evidence"; New England Journal of Medicine, Vol. 298, pp. 1336, 1978.

Lyman76 Lyman, M.: "A Data Support System for Emergency Health Care"; Proceedings of the Third Illinois Conference on Med. Inf. Systems, pp. 88-97, Univ. of Illinois at Chicago Circle, 1976.

Mabry77 Mabry, J.C., H.K. Thompson, M. Hopwood, and W.R. Baker: "A Prototype Data Management and Analysis System (CLINFO)"; Proc. of MEDINFO 77, North Holland, pp. 71-75, 1977.

Manach75 Manacher, G.V.: "On the Feasibility of Implementing a Large Relational Data Base with Optimal Performance on a Mini-Computer"; Proceedings of the First International Conference on Very Large Data Bases, D.S. Kerr, ed. , ACM, pp. 175:201, 1975.

Mase77 Mase, E.: "Medical Computer Applications in Italy"; J. of Med. Sys., Vol. 1, no. 4, pp. 403-407, 1977.

McArth78 McArthur, J.I.: "Patient Care Information System for American Indians and Alaska Natives"; Proc. 3rd USA-Japan Conf., AFIPS-IPSJ, pp. 436-440, 1978.

McDona75 McDonald, C., B. Bhargava, and D. Jeris: "A Clinical Information System (CIS) for Ambulatory Care"; Proc. of the 1975 NCC, AFIPS, Vol. 44, pp. 749-756, 1975.

McDona76 McDonald, C.J.: "Use of a Computer to Detect and Respond to Clinical Events: Its Effect on Clinician Behavior"; Ann. Intern. Med., Vol. 8, no. 2, pp. 162-167, 1976.

McDona77 McDonald, C.J., et al: "A Computer-Based Record and Clinical Monitoring System for Ambulatory Care"; American Journal of Public Health, Vol. 67, no. 3, pp. 240-245, 1977.

McLeod75 McLeod, D. and M. Heldman: "RISS - A Generalized Mini-Computer Relational Data Base Management System"; Proc. 1975 NCC, AFIPS, Vol. 44, pp. 397-402, 1975.

McShan78 McShane, D.J. and J.F. Fries: "ARAMIS: The Pilot National Arthritis Data Resource"; Proc. 3rd USA-Japan Computer Conference, AFIPS-IPSJ, pp. 343-347, 1978.

Meldma78 Meldman, M.J., D.J. McLeod, R.J. Pellicore, and M. Squire: RISS, A Relational Data Base Management System for Minicomputers; Van Norstrand Reinhold, 1978.

Melski78 Melski, J.W., D.E. Geer, and H.L. Bleich: "Medical Information Storage and Retrieval Using Preprocessed Variables"; Comp. Biomed. Res., Vol. 11, pp. 613-621, 1978.

Mesel75 Mesel, E. and D.D. Wirtschafter: "On-Line Medicaid System for Physician Services"; Comp. Biomed. Res., Vol. 8, pp. 479-491, 1975.

Mesel76 Mesel, E., D. Wirtschafter, et al: "Clinical Algorithms for Cancer Chemotherapy - Systems for Community Based Consultant-Extenders and Oncology Centers"; Meth. Inform. Med., Vol. 15, pp. 168-173, 1976.

MeselW76 Mesel, E. and D.D. Wirtschafter: Editorial "Automating Ambulatory Medical Records: A Claims-Based Medical Profile"; Comp. Biomed. Res., Vol. 9, pp 89-91, 1976.

Mishel76 Mishelevich, D.J., P. Stasing, R. Ellis, R. Gail, and S.G. Mize: "RENTRAN, An On-line Computer Based Regional Kidney Transplant Matching System"; Transplantation, Vol. 22, pp. 223-228, 1976.

Morgan77 Morgan J.L.: "Computerized SNOMED Encoding: a Practical Solution"; Medinfo 1977, North-Holland, pp. 277-281, 1977.

Morrel77 Morrell, J., M. Podlone, and S.N. Cohen: "Receptivity of Physicians in a Teaching Hospital to a Computerized Drug Interaction Monitoring and Reporting System"; Medical Care, Vol. 15, pp. 68-78, 1977.

Moscov77 Moscovice, I.: "A Method for Analyzing Resource Use in Ambulatory Care Settings"; Medical Care, Vol. 15, pp. 1024, 1977.

Munnec77 Munnecke, T., R.F. Walters, J. Bowie, C.B. Lazarus, and D.A. Bridges: "MUMPS: Characteristics and Comparison with Other Programming Systems"; Med. Inform., Vol. 2, no. 3, pp. 173-196, 1977.

Murnag73 Murnaghan, J.H. (ed.): "Ambulatory Care Data: Report on the Conference on Ambulatory Care Records"; Medical Care, II (supplement), 1973.

Nakaya77 Nakayama, T.: "A Review of Medical Computing in Japan"; J. Med. Sys., Vol. 1, no. 4, pp. 409-416, 1977.

NCHS74 National Center for Health Statistics: Ambulatory Medical Care Records: Uniform Minimum Basic Data Set: A Report to the United States Committee on Vital and Health Statistics; Vital and Health Statistics, Series 4, Number 16, Washington, D.C., U.S. Government Printing Office, 1974.

NCHSR77 National Center for Health Services Research: Demonstration and Evaluation of a Total Hospital Information System; HRA 77-3186, Hyattsville, Md., 1977.

NCHSR79 National Center for Health Services Research: Computer Applications in Health Care; DHEW (PHS) 79-3251, Hyattsville, Md., 1979.

NCI74 National Cancer Institute, Biometry Branch: Cancer Surveillance, Epidimiology and End Results Reporting, Program Description and Data Format; HEW, NIC, 1974.

ODell78 O'Dell, R.W., M.D. McClung, D.A. Spyker, et al: "Evaluation of an On-Line Poisindex Information System"; Veterinary and Human Toxiocology, Vol. 20, pp. 465, 1978.

Okubo75 Okubo, R.S., W.S. Russell, B. Dimsdale, B.G. Lamson: "Natural Language Storage and Retrieval of Medical Diagnostic Information"; Computer Programs in Biomedicine, 1975, pp. 105-130; also in Computers in Laboratory Medicine, ed. D. Enlander, Academic Press, 1975, pp. 54-60.

Olle78 Olle, W.: The Codasyl Approach to Data Base Management; John Wiley & Sons, November 1978.

Orosz78 Orosz, G.M. et al: Physicians Microcomputer Report, Box 6483, Lawrenceville, N.Y., published since 1978.

OTA77 Office of Technology Assessment: Policy Implications of Medical Information Systems; Report of the Congressional Office of OTA, Washington, D.C., Nov. 1977.

Oxman76 Oxman, M.A., H.M. Kissman, J.M. Burnside, et al: "The Toxicology Data Bank"; J. Chem. Inf. Comput. Sci., Vol. 16, pp. 19-21, 1976.

Palley75 Palley, N. and G.F. Groner: "Information Processing Needs and Practices of Clinical Investigators - Survey Results"; Proc. of the 1975 NCC, AFIPS Vol. 44, pp. 717-723, 1975.

Penick76 Penick, R.W. and F. M. Crowder: "Integrated Data Communications System for Ambulatory Care"; Proceedings of the Third Illinois Conference on Med. Inf. Systems, Univ. of Illinois at Chicago Circle, pp. 98-109, 1976.

Peto77 Peto, R., et al: "Design and Analysis of Randomized Clinical Trials Requiring Prolonged Observation of Each Patient. II. Analysis and Examples"; Br. J. Cancer, Vol. 35, pp. 1, 1977.

Pople72 Pople, H. and G. Werner: "An Information Processing Approach to Theory Formation in Biomedical Research"; Proc. of the 1972 FJCC, AFIPS Vol. 41, pp. 1125-1138, 1972.

Pople75 Pople, H.E., J.D. Myers, and R.E. Miller: "The DIALOG Model of Diagnostic Logic and its Use for Internal Medicine"; Proceedings of the 4th IJCAI, pp. 848-855, 1975.

Pratt73 Pratt, A.W.: "Medicine, Computers, and Linguistics"; Advances in Biomedical Engineering, Vol. 3, Academic Press, pp. 97-168, 1973.

Ragen78 Ragen, D.P. and S.A. Jones: "High-Level Language Implementation of Bit Map Inverted Files"; Comp. Biomed. Res., Vol. 11, pp. 593-612, 1978.

Ransil74 Ransil, B.J.: "Application of the PROPHET System in Human Clinical Investigation"; Proceedings of the 1974 NCC, AFIPS Vol. 43, pp. 477-484, 1974.

RCGP72 Royal College of General Practioners Classification; RCGP, London SW7 1PN, 1972.

Reiche77 Reichertz, P.L.: "Medical Computer Applications in Germany"; J. of Med. Sys., Vol. 1, no. 4, pp. 417-424, 1977.

Reiche68 Reichertz, P.L: "PIRS, Personal Information Retrieval Systems"; Meth. Inf. Med., Vol. 7, no.3, pp. 165-172, 1968.

Rodnic77 Rodnick, J.E.: "The Use of Automated Medical Records"; The Journal of Family Practice, Vol. 5, pp. 253-264,1977.

Rosati75 Rosati, R.A., J.F. McNeer, C.F. Starmer, et al: "A New Information System for Medical Practice"; Arch. Internal Med., Vol. 135, pp. 1017-1024, 1975.

Sauter76 Sauter, K., P.L. Reichertz, et al: "A System to Support High-Level Data Description and Manipulation of an Operational Database System"; Medical Informatics, Vol. 1, no. 1, pp. 15-26, 1976.

Schnei75 Schneider, W. and Bengtsson (Eds.): "The Application of Computer Techniques in Health Care"; Comp. Prog. in Biomed., Vol. 5, pp. 171-249, 1975.

Schnei78 Schneider, W.: "Design Considerations for a Knowledge Base and Language Interface to a Drug Information System"; Proceedings of the Second Annual Symposium on Computer Applications in Medical Care, IEEE, pp. 326-329, 1978.

Schult76 Schultz, J.R.: "PROMIS: Problem-Oriented Medical Information System: A Medical and Technical Solution to Four Problems in Medicine"; Proceedings of the Third Illinois Conference in Med. Inf. Systems, pp. 1-20, Univ. of Illinois at Chicago Circle, 1976.

Schult79 Schultz, J.R. and L. Davis: "The Technology of PROMIS"; Proc. of the IEEE, vol. 67, no. 9, September 1979, pp. 1237-1244.

Sheple77 Shepley, D.J.: "Medical Computer Applications in Canada"; J. of Med. Sys., Vol. 1, no. 4, pp. 425-431, 1977.

Shortl79 Shortliffe, E.H., B. Buchanan, and E.A. Feigenbaum: "Knowledge Engineering for Medical Decision Making: A Review of Computer-Based Clinical Decision Aids"; Proceedings of the IEEE, vol. 67, no. 9, September 1979, pp. 1207-1228.

Sibley77 Sibley, W.L., M.D. Hopwood, G.F. Groner, W.H. Josephs, and N.A. Palley: "Data Management for Clinical Research"; Proc. of the 1977 NCC, AFIPS Vol. 46, pp. 63-68, 1977.

Silver78 Silvers, A., J.R. O'Fallon, and J. Bill: "Data Base Management Structure in Cancer Clinical Trials"; Proc. of the 3rd USA-Japan Computer Conference, AFIPS-IPSJ, pp. 353-354, 1978.

Simbor76 Simborg, D.W., et al: "Information Factors Affecting Problem Follow-Up in Ambulatory Care"; Medical Care, Vol. 14, pp. 848-856, 1976.

Slack66 Slack, W.F., G.P. Hicks, C.E. Reed, et al: "A Computer Based Medical History System"; N. Engl. J. Med., Vol. 274, pp. 194-198, 1966.

Slamac77 Slamecka, V., H. Camp, A.N. Badre, and W.D. Hall: "MARI: A Knowledge System for Internal Medicine"; Info. Proc. Manag., Vol. 13, pp. 273-276, 1977.

Soffee76 Soffeer, S.B., S.M. Entine, T. Wolfe, A. Eggert et al: "A Database System for Archival Storage of Clinical Lab Data"; Proc. of 3rd Illinois Conf. on Med. Inf. Sys., pp. 278-288, Nov. 1976.

Starfi77 Starfield, B.H., et al: "Coordination of Care and Its Relationship to Continuity and Medical Records"; Medical Care, Vol. 15, pp. 929-938, 1977.

Starme74 Starmer, C.F., R.A. Rosati, and J.F. McNeer: "Data Bank Use in Management of Chronic Disease"; Comp. Biomed. Res., Vol. 7, pp. 111-116, 1974.

Starme75 Starmer, C.F. and R.A. Rosati: "Computer Based Aid to Managing Patients with Chronic Illness"; Computer(IEEE), Vol. 8, pp. 46-50, 1975.

Stei&B78 Steinwachs, D.M. and T. Bledsoe: "A Minimal Data Set for Ambulatory Care: Aplications of a Computerized System for Management and Quality Assurance"; Management Information Systems: A Collection of Case Studies, American Hospital Association, Chicago, 1978.

Stei&M78 Steinwachs, D.M. and A.I. Mushlin: "The Johns Hopkins Ambulatory Care Coding Scheme"; Health Services Research, Vol. 13, pp. 36-49, 1978.

Stimso78 Stimson, D.H. and G. Charles: "A Computer-Based Information System in an Ambulatory Care System and Case Study"; Proc. of the 3rd USA-Japan Comp. Conf., pp. 447-451, AFIPS IPSJ, 1978.

Stoneb75 Stonebraker, M. and G. Held: "Networks, Hierarchies, and Relations in Data Base Management Systems"; Proc. ACM Pacific 75 Conference, Boole and Babbage, Palo Alto, CA, 1975.

Szolov78 Szolovits, P. and S.G. Pauker: "Categorical and Probabilistic Reasoning in Medical Diagnosis"; Artificial Intelligence, Vol. 11, no. 1/2, pp. 115-144, August 1978.

Thomps76 Thompson, H.K., J.C. Mabry, M. Hopwood, W.R. Baker: "User Experience with a Prototype Data Management and Analysis System for Clinical Investigators (CLINFO)"; Proceedings of the Third Illinois Conference on Med. Inf. Systems, Univ. of Illinois at Chicago Circle, pp. 21-28, 1976.

Todd75 Todd, S.J.P.: "Peterlee Relational Test Vehicle"; Proceedings of the 1975 VLDB, ACM, pp. 554-556, 1975.

Ullman79 Ullman, J.D.: Principles of Database Systems; Computer Science Press, Potomac MD, 1980.

Valbon75 Valbona, C., et al: "Computer Support of Medical Decisions in Ambulatory Care"; Meth. of Inf. in Med., Vol. 14, no. 2, pp. 55-62, 1975.

Wanner78 Wanner, J.F.: "Wideband Communications System Improves Response Time"; Computer Design, Dec. 1978, pp. 85-92

Warner78 Warner, H.R.: "Knowledge Sectors for Logical Processing of Patient Data in the Help System"; Proceedings of the 2nd Annual Symposium on Computer Applications in Medical Care, IEEE, pp. 401-404, 1978.

Wasser76 Wasserman, A.I. and D.D. Sherertz: "A Balanced View of MUMPS"; Proc. of Conference on Programming System in the Small Processor Environment, ACM SIGPLAN Notices, Vol. 11, no. 4, pp. 16-26, April 1976.

Watson77 Watson, R.J.: "A Large-Scale Professionally Oriented Medical Information System - Five Years Later"; J. of Med. Sys., Vol. 1, no. 1, pp. 3-21, 1977.

Watana78 Watanabe, A., H. Hamano, R. Iwai: "Small Business Computers in Medical Treatment and Management Systems"; Proc. of the 3rd USA-Japan Comp. Conf., pp. 441-446, AFIPS IPSJ, 1978.

Weed71 Weed, L.L.: Medical Records, Medical Education, and Patient Care; The Press of Case Western Reserve University, Cleveland, OH, 1971.

Weed78 Weed, L.L., et al: "Representation of Medical Knowledge and PROMIS"; Proceedings of the Second Annual Symposium on Computer Applications in Medical Care, IEEE Press, pp. 368-400, 1978.

Weyl75 Weyl, S., J.F. Fries, and G. Wiederhold: "A Modular Self-Describing Clinical Databank System"; Computers and Biomedical Research, Vol. 8, pp. 279-293, 1975.

WHO78 International Classification of Diseases, Adapted for Use in the United States, Ninth Revision, U.S. Government Printing Washington, D.C., 1978.

Wieder75 Wiederhold, G., J.F. Fries, and S. Weyl: "Structured Organization of Clinical Databases": Proc. of the 1975 NCC, AFIPS, Vol. 44, pp. 479-485, 1975.

Wieder76 Wiederhold, G.: "Alternatives in Database Design for Ambulatory Records"; Proceedings of the Third Illinois Conference on Med. Inf. Systems, University of Illinois at Chicago Circle, pp. 78-87, 1976.

Wieder77 Wiederhold, G.: Database Design; McGraw-Hill Book Company, New York, (Computer Science Series), May 1977.

Wieder78 Wiederhold, G.: "The Data Model of MUMPS Globals"; Proceedings of the 1978 MUMPS User Group Meeting; P. Zimmerman (Ed.), MUG, Bedford, Mass., pp. 159-183, 1978.

Wieder79 Wiederhold, G. and R. El-Masri: "A Structural Model for Database Systems"; Stanford CS Report 79-722, May 1979.

Wied&K78 Wiederhold, G. and I. Kuhn: "Effective Services of Automated Ambulatory Medical Record Systems"; Policy Analysis and Information Systems, pp. 22-32, 1978.

Wied&M75 Wiederhold, G. and E. Mesel: "Ambulatory Automated Medical Record Systems"; Proc. 1975 MUMPS Users' Group Meeting, J. Zimmerman (Ed.), Biomedical Computer Lab., St. Louis, Missouri, pp. 182-191, 1975.

Wilson78 Wilson, D.H.: "Computer Assisted Medical Record"; Proc. of the 3rd USA-Japan Comp. Conf., AFIPS IPSJ, pp. 383, 1978.

Wirtsc78 Wirtschafter, D.D., M.Scalise, H.Cort, C.Henke, and R.A.Gams: "Clinical Algorithms in A Cooperative Multi-Institutional Clinical Trial"; Proceedings of Fourth Illinois Conference on Medical Information Systems, University of Illinois at Chicago Circle, 1978.

Wirtsc79 Wirtschafter, D.D., J.T.Carpenter jr., and R.A.Gams: Facilitating the Medical Innovation Adoption Process with Clinical Algorithms: Experience with Two Anti-neoplastic Chemotherapy Protocols"; Proceedings of the 3rd Jerusalem Conference on Information Technology, North Holland 1978/1979.

Wong78 Wong, R.L., R. Plah, A. Baldwin, and T. Hain: "Experience on a Surgical Pathology Database and a Lexicon Compiled from Logically Derived Medical English"; Proceedings of the 2nd Annual Symposium on Computer Applications in Medical Care, pp. 326-327, 1978.

Worth78 Worth, R.M.: "An Integrated Computer-Assisted Medical Record and Billing System"; Medical Group Management, Vol. 25, pp. 42-50, 1978.

Yokel78 Yokel, R.A., L.T. Sigell, E.D. Nelson, et al: "A Comparison of Four Toxiocology Resources in Respect to Rates of Retrieval and Time Required; J. Pediatr., Vol. 92, pp. 145-148, 1978.

Young72 Young, D.S., R.B. Friedman, et al: "Effects of Drugs on Clinical Laboratory Tests"; Clinical Chemistry, Vol. 18, pp. 1041-1303, 1972

Young78 Young, J.L., A.J. Asire, and E.S. Pollack (eds.): "SEER Program: Cancer Incidence and Mortality in the United States 1973-1976"; NCI, HEW, Bethesda Md., DHEW pub. (NIH) 78-1837, 1978.

Yu79 Yu, V.L., B.G. Buchanan, E.H. Shortliffe, S.M. Wraith, R. Davis, A.C. Scott, and S.N. Cohen: "Evaluating the Performance of a Computer-Based Consultant", Computer Programs in Biomedicine, Vol. 9, pp. 95-102, 1979.

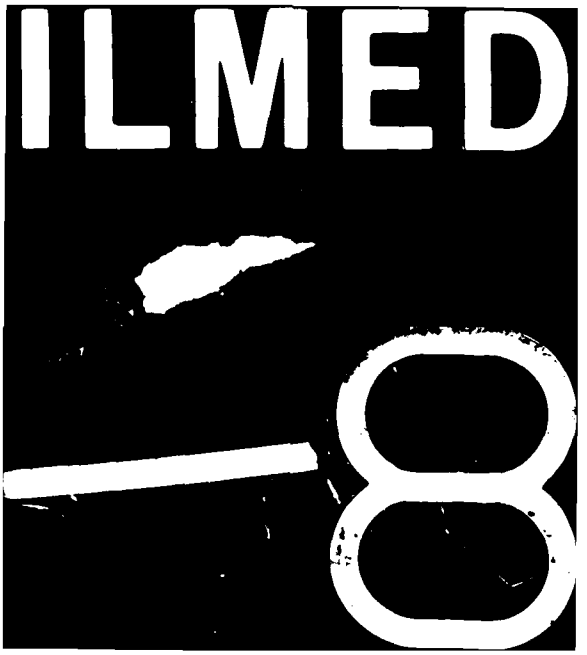
Zielst77 Zielstroff, R.D. et al: "Experience with a Computer-Based Medical Record for Nurse-Practioners in Ambulatory Care"; Comp. and Biomed. Res., Vol. 10, pp. 61-74, 1977.

Zimmer77 Zimmerman, J.: "Towards Generalized Automated Ambulatory-Care Record Systems"; MEDINFO 77, IFIP, North-Holland Publishing Company, 1977.

Zimmer78 Zimmerman, J., R.S. Gordon, D.K. Tao, and S.B. Boxerman: "The Acceptability of Computer Applications to Group Practices": Journal of Medical Systems, Vol. 2, no. 1, pp. 15-24, 1978.

Zimmer79 Zimmerman, J., S.B. Boxerman, and A.L. Rector: "Are Micro-Computers Appropriate for your Practice? "; Journal of the AMA, Vol. 242, no. 17, pp.1887-1890, Oct. 1979.

Zimm&R78 Zimmerman, J. and A. Rector: Computers for the Physician's Office; Research Studies Press, Forest Grove, OR., 1978.


ILMED
8