A COMPUTER ASSISTED INSTRUCTION SYSTEM FOR TEACHING DECISION-MAKING IN CLINICAL MEDICINE

by

Richard Stephen Dick

A dissertation submitted to the faculty of The University of Utah in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Medical Biophysics and Computing

The University of Utah

August 1978

THE UNIVERSITY OF UTAH GRADUATE SCHOOL

SUPERVISORY COMMITTEE APPROVAL

of a dissertation submitted by

Richard Stephen Dick

I have read this dissertation and have found it to be of satisfactory quality for a doctoral degree.

Date

Homer R. Warner, M. D.
Chairman, Supervisory Committee

I have read this dissertation and have found it to be of satisfactory quality for a

Date

T. Allan [redacted], Ph. D.
Member, Supervisory Committee

I have read this dissertation and have found it to be of satisfactory quality for a doctoral degree.

8/1/78
Date

THE UNIVERSITY OF UTAH GRADUATE SCHOOL

FINAL READING APPROVAL

To the Graduate Council of The University of Utah:

I have read the dissertation of [redacted] in its final form and have found that (1) its format, citations, and bibliographic style are consistent and acceptable; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the Supervisory Committee and is ready for submission to the Graduate School.

Date

Homer R. Warner, M. D.
Member, Supervisory Committee

Approved for the Major Department

Homer R. Warner, M. D.
Chairman/ Dean

Approved for the Graduate Council

ABSTRACT

Clinically-oriented computer assisted instruction (CAI) has been receiving greater emphasis in the last few years. Similarly progress has been made in the development of (a) generative CAI techniques and (b) the refinement of a formalized medical decision-making criterion. These developments have led to the implementation of a nonauthored CAI system for training clinicians in medical decision-making using actual patient data.

The nonauthored CAI system is composed of several subsystems. The principal components include (a) a statistical analysis system providing many options including the generation of histograms, (b) an automated medical decision-making system, and (c) a system for finding patients with desired characteristics and generating populations interactively. The functions and interrelationships of each component are described in detail. The nonauthored CAI system's primary advantages include: (a) it provides the capability of studying medical decision-making procedures in a formal setting; (b) it is a subject-independent system and therefore is relevant to a broad spectrum of clinicians (it is limited by the extent of the clinical data available); (c) because of its generative qualities no courseware "authoring" is required; and (d) because the system requires little training to use, investigations may be conducted in either a formalized group setting or in an independent study mode.

Due to the system's relevance to clinical medicine, several persons affiliated with academic medicine have seen the system demonstrated. In addition to these observers, a demonstration of the system in a more structured setting was explored. A broad spectrum of clinicians repeatedly used the CAI system to investigate subjects of interest during these structured sessions. Details of their investigations and comments are included. Finally, a few future possibilities and enhancements for the system are explored.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

Statement of the Problem and a Proposed Solution

The major emphasis in teaching medical students may be called the deterministic model of disease and its manifestations. That is, students are taught cause and effect relationships through basic science courses and this forms the basis for clinical reasoning during their clerkships. However, the experienced clinician does not rely solely on such a deterministic model but makes many decisions on the basis of a stochastic or empirical model as well. The clinician develops self-conceived statistical models that have evolved from years of clinical experience with many patients. These two models are not conflicting but instead tend to strengthen and support one another. The educational problem is that the student does not experience both modalities, but because of limited time and clinical exposure he is exposed almost exclusively to the deterministic model.

This thesis presents a computerized system for statistically analyzing available parameters on large numbers of actual patients in an interactive setting. In a short period of time the student, aided by the instructor, may investigate data on patients with similar abnormalities. He may form hypotheses from his observations then test them against the actual patient data. The system is useful for the medical student and clinician alike. By examining patient data the user may

identify the criteria that are keys to making diagnostic and therapeutic decisions. It is the purpose of this report to describe a computer assisted instruction (CAI) system designed to provide a student with experience in the application of a statistical approach to decision-making using this model with actual patient data; and (2) he may acquire some skill in derivation of decision logic from empirical data on patients.

An Overview of Computer Assisted Instruction (CAI)

More and more applications are being devised for the use of the computer as an instructional tool to either describe or simulate activities in a wide variety of disciplines. Projections show that this growth is likely to be sustained (Hickey, 1974). Some factors contributing to CAI's growth include evaluation studies which reveal that it is as effective as traditional instruction when used as a replacement, and that students learn more material in less time (Abrahmason et al., 1969). Economic factors contributing to the growth of CAI will in time become increasingly important. One authority (Hickey, 1974) suggests that CAI is already less expensive than traditional instruction in some subject areas. He predicts that within a decade CAI will become cheaper in virtually all disciplines.

Uttal et al. (1970, pp. 1-3) have used the following three broad categories for classifying CAI: (1) "degenerate computer teaching machines;" (2) "selective computer teaching machines;" and (3) "generative computer teaching machines." The nature of these three categories of CAI will be discussed in subsequent paragraphs.

Degenerate Computer Teaching Machines

The degenerate computer teaching machine is characterized by a computer acting primarily as an electronic page-turner. Very little interaction and feedback occurs. Due to their ineffectiveness such systems have not endured (Uttal et al., 1970).

Selective Computer Teaching Machines

Selective computer teaching machines are characterized by author-produced dialogue consisting of "canned" questions and "appropriate" answers. These systems are also known as frame-oriented (or frame-by-frame) systems. Considering the full spectrum of CAI developments, the great majority may be considered frame-oriented systems (Molnar, 1976; Roid, 1974). The question-answer dialogue results from what is called the CAI courseware. Complete descriptions of how courseware is authored appear in the literature (Reed et al., 1974). The following is a brief excerpt to convey the notion of a frame.*

Which one of the following is the most useful in screening patients for Cushing's Syndrome (regardless of etiology)?

A. Metyrapone tests
B. ACTH intravenous infusion tests
C. Plasma ACTH concentration
D. 8 a.m. plasma cortisol concentration
E. Overnight single dose dexamethasone suppression test

?C

Incorrect. ACTH concentration measurement is not a good screening procedure because values may be low, normal, or high in Cushing's syndome depending on the etiology. Try again.

---

*Italics represent computer output.

?D

Incorrect. There is too much overlap with normals for 8 a.m. cortisol concentration to be good as a screening procedure. Try again.

?E

Good. Suppression of 8 a.m. plasma cortisol to a low level occurs in normals after administration of dexamethasone. Patients with Cushing's Syndrome fail to suppress.

Perhaps the most notable attribute characterizing frame-oriented systems is that all questions must be specified by an "author" as well as a set of anticipated student responses to each question. Any branching that may occur must be determined in advance by the "author." Since all questions, answers, and branching to other remedial questions must be specified (an activity called "authoring" of courseware) in advance of any student-computer interaction, time commitments to courseware development make it very expensive. In this mode it is actually prohibitive for teaching some subjects.

One author notes:

> The most critical problem for computer-based instruction is the lack of high quality courseware and the high costs associated with courseware development. Costs can range from $3,000-$20,000 per contact hour and total costs can easily exceed equipment and operating costs. (Molnar, 1976, p. 21)

A contact hour is defined as the material needed to occupy a student for one hour at the terminal actively using the system, receiving questions and responding to them. The difficulties and frustrations encountered when working with frame-oriented systems must be experienced to fully appreciate why the experts say that alternative methods of CAI must be developed.

The following interesting analogy illustrates the limitations

of selective computer teaching machines:

> Selective computer teaching machines thus suffer from the same limitation as the pictographic languages of the Orient. There is little generality to the representational scheme and there must be as many characters (items) as there are concepts (student interactions). The coded languages of the West, however, use a very small number of symbols and a limited number of combination rules to form words representing a wide variety of concepts. This coded, or algorithmic, approach to language is far more efficient, economical and elegant; in addition, it possesses psychological advantages of considerable consequence. The mind-blasting result of memorizing 5,000 to 10,000 weakly discriminable characters has to be experienced to be appreciated. (Uttal et al., 1970, p. 3)

Nonauthored CAI Teaching Machines

The notion of nonauthored CAI or generative computer teaching machines has received limited attention in the literature. Such systems generally employ the computer for generating problems, answers, diagnostic and remedial feedback. Nonauthored CAI departs from frame-oriented systems by operating on the data base such that the output generated produces a dynamic learning environment. What determines if a system is generative? One expert states:

> It is important to recognize that generative CAI is still being defined, . . . the term is defined more by instance than by any class description. Thus, the attempt to demand precision from those who use the term may be compared with attempts to demand a definition of constitutional powers that wise men have left ambiguous. To offset a tendency to want premature closure, researchers need to maintain a set of diversity that welcomes contrasting paradigms of generative CAI. . . . (McMullen, 1974, p. 30)

Some instances will be discussed later. This mode of CAI may potentially alter the future of CAI as seen by the experts. The leading CAI experts were recently interviewed by Hickey (1974). Some 59 topics for CAI research were recommended by the experts. Among these, nonauthored (generative) CAI and ten other topics were given the

highest priority for research.

The appeal of generative CAI, according to Rigney (see Hickey, 1974, p. 49) is that it is one way to relieve the awesome burden of authoring and anticipating every possible response. Licklider (see Hickey, 1974, p. 50) is quite critical of the frame-by-frame approach to CAI. He points out that almost all CAI authors conceive of CAI as being something like a book, except that it is interactive. As an alternative, he suggests that CAI programs should be more model-driven, that is, deriving questions and answers based on some model. Furthermore, he states that we certainly should not stay with the idea of "canned" questions. Suppes (in Hickey, 1974, p. 50) feels that "highly generative courses for which the authors have to write rather sophisticated algorithmic descriptions of classes of problems" as opposed to the conventional selective CAI approach, is the way CAI development in the future will be implemented.

Nonauthored (or generative) CAI as described by these experts is not a trivial capability to develop. For example, some models incorporate sophisticated techniques rooted in artificial intelligence (Molnar, 1976). Molnar states: "The fields of artificial intelligence, algorithmic . . . [approaches], and cognitive learning are all converging. . ." (p. 19) to produce a new and viable approach to CAI. This view is supported by other researchers of CAI systems (Koffman et al., 1975; Peele et al., 1975).

It is readily apparent that some subject areas may not lend themselves to the generative approach. Subject matters suited to generative or nonauthored CAI are limited as noted in the following:

> The idea of generation hinges upon the fact that there are some subject matters which are amenable to algorithmic manipulation. The list of subject matters so amenable is small: currently, it includes the mathematical course materials, other mathematically oriented subjects like the physical sciences, and some special subject matters such as chemistry or logic in which there is a formal system describing the relations between and among parts. . . . (Uttal et al., 1970, p. 3)

Similarly, another CAI expert supports this view by stating that "the properties of Generative CAI make it particularly useful for problem oriented courses." (Koffman, 1973, p. 182)

Although medicine is not listed among the subject areas appropriate for generative CAI development, it possesses some of the required attributes. The procedures in diagnostic and prescriptive medicine may be ordered into a "formal system describing the relations between and among parts," as stated in the above quote (Uttal et al., 1970, p. 3), and therefore are "amenable to algorithmic manipulations." This is a most significant attribute because it means that nonauthored or generative CAI theoretically may be applied to medical education. A few examples may aid in conveying the notion of nonauthored CAI, especially when contrasted with the previous examples given for frame-oriented CAI.

One of the leading CAI experts has created a "generative" (nonauthored) CAI system (Koffman, 1973) which is designed as an introductory course in computer science at the University of Connecticut. The system considers the students' "level" which is dynamic and is derived from past performance. The current "level" is used by the question-generation algorithm to assess the difficulty of the next question. The system generates drill and practice types of questions for the

conversion of numbers between base 2, 8, 10, or 16, and the logical AND of such numbers. The question-generation algorithm uses the current "level" to determine how many digits and the base of the numbers to be presented in the next question. For example, the probability of asking a student to give the logical AND of two hexadecimal numbers increases proportionally with student "level." This brief example illustrates that the computer can generate appropriate problems and evaluate student responses without predetermined, human, one-for-one question-answering production or manual interventions.

Principles from artificial intelligence are increasingly important in the development of some of the more exotic attempts to develop nonauthored CAI systems. For example, SCHOLAR (Carbonell, 1970) is a nonauthored CAI system which utilizes a data base of facts concerning South America; it is not generalized for other subjects. The terms in the data base form a semantic network with associated pointers to other related terms. For example, Brazil is a unit in the semantic network with attributes such as longitude, capital city, etc. The values of the attributes appear as other terms in the network. The network may be used for both question-generation and answering. To illustrate, the system may ask the user the question, "The capital of Argentina is ---- (fill in)?" When generating such a question, the system also retrieves the correct answer for a subsequent comparison with the user's response. Similarly the system permits free form input for soliciting factual information. For example, a student may type, "What is the capital of Argentina?" The semantic network is traversed based on the relationships and pointers and the correct answer is generated and presented

to the user.

Another nonauthored CAI project is designed for:

> . . . assisting a student in constructing a proof by applying the rule of inference designated by him to earlier lines in the proof sequence. If the student's rule cannot be applied, he is given remedial feedback. Efforts to verify that a line typed in by a student is a valid inference from his previous work have been moderately successful. A more ambitious goal now under study is to have the theorem prover suggest ways of completing a proof taking into account the student's incomplete or erroneous work. (Koffman et al., 1975, p. 218)

The authors present the complexity of the problems associated with natural language systems and lament that "none of these systems currently have an effective way of adapting the tutorial dialogue to exploit the information present in an incorrect student response." (Koffman et al., 1975, p. 218)

With such nonauthored systems, one may ask, "are there any generalized, subject-independent systems, or more especially any medically-oriented nonauthored systems? Manual and automated searches of the literature have revealed no such reports concerning these systems. These searches have been conducted by the author and experts at the National Library of Medicine.

Medical CAI in Perspective

Medical education is expensive. Costs ranging from $16,000 to $26,000 per student per year were recently reported by the Association of American Medical Colleges. This and other factors have motivated the investigation of alternative methods of instruction. Most of the medical CAI developments have been aimed at undergraduate medical education. These may be classified into one of two categories:

(1) preclinical (basic sciences) or (2) clinical. A recent survey (Kamp, 1975) of the medical institutions active in CAI developments found that 46 institutions reported 300 units (distinct courses) of medical CAI representing 743 contact hours of student-computer interaction. More than half of all the CAI developed was accomplished at four institutions. This led the Lister Hill National Center for Biomedical Communications to sponsor an experiment in medical CAI between 1973 and 1975. The experiment was conceived to answer several questions, some of which include:

1. If CAI services are made available over a time-sharing communications network, will people be willing to sample them?

2. Can the material offered over the network be integrated in a meaningful way to existing curricula?

3. Can it produce measurable benefits?

4. Will it produce changes in ways medical schools teach?

5. What is the mechanism whereby new CAI material may be generated most efficiently and cost-effectively?

Four contracts were negotiated, one with a commercial time-sharing service, and three with institutions possessing CAI developmental expertise. The latter included (1) Ohio State University College of Medicine; (2) The University of Illinois Medical Center; and (3) Massachusetts General Hospital. Over 100 user institutions participated in the experiment, accessing CAI courseware at one or more of the three host centers.

These studies report that most of the above questions were never answered by the experiment, but question one was answered very

affirmatively. A monthly average usage of more than 2,000 hours shows that not only will users sample the materials, but they repeatedly access them. The available CAI materials included many hours of clinical propgrams. The experiment showed that medical students in their second and third years and house staff used the clinical programs extensively. A discussion of the network as it relates to the above questions is presented by Rubin (1976).

The main reason the other questions were not answered is lack of effective means of measurement and not enough commitment on the part of participating institutions to supply the manpower required for a more critical analysis. The participants were expected to take an active role, but the incentives for such involvement simply were not there. A network evaluation contract with Human Resources Research Organization ("HumRRO") was negotiated and their report has been published (Rubin et al., 1975).

A recent study of 131 U.S. and Canadian medical schools reported that 81 schools made use of CAI, and of these, 53 or nearly 66% were using clinical applications of CAI (Wenzel, 1976). Earlier studies (Brigham et al., 1972; Kamp, 1975) have also attempted to categorize CAI usage into either preclinical or clinical, but they report a lower percentage of clinical applications. These studies reveal an increasing demand for clinically-oriented CAI.

The Deterministic and Stochastic Approaches for Medical Education

The Deterministic Model

The conventional approach to teaching physicians is centered on

teaching basic sciences and two years of clinical clerkships where cause and effect relationships are emphasized. This approach conforms to what may be called the determinstic model. According to this model the student examines a few patients and is expected to evaluate and help manage them, demonstrating an understanding of the cause and effect relationships that were taught in the basic science courses. Perhaps the knowledge most critical to the clinician in understanding these cause and effect relationships is physiology. For this reason the student is encouraged to develop a thorough knowledge of physiology so that he may recognize the mechanisms responsible for both normal and abnormal functions. For example, a recent two-year study (Geyman, 1976) revealed that congestive heart failure ranked as the 19th most common problem seen by family practitioners. With such a common disease it is imperative that the clinician have a thorough understanding of cardiovascular physiology, both for the normal patient and the patient in congestive heart failure so that he may quickly recognize and treat the problem.

The following illustrates how the deterministic model is presented in a physician's training. In the physiology courses the student is taught the physiological mechanisms regulating cardiac output. When the student encounters a patient who may be experiencing congestive heart failure, he can draw upon his knowledge of the circulatory system and the mechanisms regulating cardiac output and properly confirm such a diagnosis and begin treatment. This requires that the student ask the right questions during the history phase such as "How many pillows do you sleep on at night?" and "Does mild exertion leave you

breathless?" Therefore, asking the right questions and performing the proper tests during the history and physical exam is predicated upon a thorough knowledge of the physiology of cardiac output regulation. The appropriate results of the questions and tests may lead the physician to the diagnosis of congestive heart failure.

After making such a diagnosis the student may then be called upon to specify recommendations for treatment. Once again, a thorough knowledge of physiology is important. For example, the student may recommend that the patient be placed on a diuretic, knowing that the symptoms of congestive heart failure are partly due to an increase in blood volume. Therefore, the administration of diuretics is one treatment to achieve the desired result. This brief example, though far from complete, illustrates how the determinstic model operates. First, the student is taught the normal and abnormal physiology and from this knowledge the student can deduce symptoms that result from some of these underlying mechanisms. This is an important and crucial model, but it is not the only model for medical instruction.

The Stochastic Model

It is generally considered that the experienced clinician's day-to-day practice of medicine is performed in a very different mode than that which is taught in medical school. One of the reasons for internships and residencies is for the student to get experience with the diagnostic, prognostic, and treatment phases of clinical medicine. It is in practice that the clinician begins to build his own stochastic model of the various disease states observed. Up until the time of

the internship or residency the student is more or less practicing textbook medicine. Therefore, under the supervision of the "experts" he may begin to build his own view of the clinical parameters in medical decision-making. It may literally take years to build a model of a disease, but each new case serves to confirm concepts that were generated from previous cases. Each new case has potential for altering the current perspective the clinician has for the disease. Therefore, the clinician is forming an evolving intuitive model of the important relationships between the diagnoses, symptoms, and the treatments of various diseases. The parameters of such a model are frequencies and distributions.

One of the problems in teaching only the deterministic model to medical students is that it encourages medical decision-making based on classical examples with little emphasis on the range of manifestations. Medical texts are replete with generalized statements regarding various parameters for various disease states. For example, it is not infrequent to see medical texts state that parameter X may be elevated in a particular disease, while parameter Y is reduced in the same disease, thus giving the inexprienced clinician no real empirical numbers for decision-making. From this brief discussion of problems encountered in building one's own stochastic or intuitive model of a disease, it is no wonder that it requires years of clinical experience to develop an accurate statistical modelfrom a single physician experience.

The computer system developed as part of the research described in this thesis is designed to provide the clinician with the following advantages for building his own empirical models of disease:

1. The system does not require a faculty member to author a dialogue such as that seen in the examples of the frame-by-frame-oriented CAI systems. This could be called the unstructured approach.

2. It allows a student to improve his problem-solving skills by developing hypotheses which he can test against actual patient data. For example, he may explore relations between one laboratory test value and another. He may also test for independence of laboratory test values from each other within a set of patients. Similarly, the student may explore the relative frequencies of the various disease manifestations.

3. The system permits the student to develop medical decision-making criteria from the study of these relationships and then test these criteria using data from real individual patients or sets of patients. That is, it allows the student to explore relationships among variable data which can then be used in medical decision-making on future patients.

4. Another advantage afforded the user is the ability to explore other data and decisions on patients who exhibit exceptionally low or high values for any specific data item (e.g., outliers on a statistical distribution). This feature permits the user to review the complete set of data collected for a single patient, which hopefully will be enlightening in determining what caused the patient to be exceptional.

5. A student may develop skills in deductive reasoning based on data derived from real patients and improve his ability to develop stochastic models through the experience with deductive reasoning obtained as he uses the system.

6. This system is self-documenting in the following ways: (a) the logic developed by the student to define populations and variables is retained by the system and a hardcopy is available to the user upon request. (b) The sequence of hypotheses developed by the student and the results of testing these hypotheses is retained in the form of an audit trail file. A hardcopy of this sequence may be printed by the user at any time. (c) the data base may be saved in its current state (i.e., at the end of one session in a series), and later it may be restored at the user's request. This feature allows the other users of the teaching mode to pursue their investigations independently without lengthy delays for data base regeneration.

CHAPTER II

The generative teaching system presented in this thesis consists of several component subsystems. This chapter describes the functions performed by the subsystems and how they are integrated to form the teaching system.

HELP As a Subsystem for Decision-Making

HELP is a computerized medical decision-making system that is an integral part of this teaching system. HELP is designed to simulate the decision-making processes employed by the medical expert. Many medical decision-making algorithms are presented in the medical literature, such as those presented by Ledley et al. (1959), Warner et al. (1961), Gorry et al. (1968), Cornfield et al. (1973), Moore (1973) and others. From these efforts it is apparent that computerized medical decision-making is not only possible, but practical in a variety of circumstances. These studies have concluded that the computer can simulate the expert and arrive at conclusions similar to those reached by medical experts, given the same data.

The HELP system was developed by Warner and his associates. It is an advanced decision-making system which is in clinical use as a consulting tool at the LDS Hospital in Salt Lake City.

HELP is used in the teaching system to (a) create medical decision-making logic, and (b) generate populations and variables. The

remainder of this section explains the fundamental concepts pertaining to these two functions.

Medical Decision-Making

In designing HELP, four fundamental criteria were considered: (1) patient data should be acquired insofar as possible by direct input to the computer from automated equipment (examples include ECG, spirometry, and clinical laboratory data); (2) the system should provide consultation to the physician in the form of decision suggestions (Pryor et al., 1975); (3) the medical decision-making logic should be readily transferrable; and (4) the computer response time should be acceptable (under two seconds). Thus, the HELP system is designed (1) to make available the expert physician's logic to nonexperts on a 24-hour basis, and (2) to standardize the decision-making procedure. Articles by Warner (1972), Pryor et al. (1975), and Gardner (1975) provide further details about the HELP systems.

A general requirement for the HELP system to operate effectively is the existence of a computerized data base. The patient data file, the medical decision-making logic file, and the problem file (e.g., a list of problems noted for the patient) all contribute to the HELP decision-making process. The "patient data file" contains data acquired from several sources throughout the hospital (i.e., patient history, heart catheterization laboratory, clinical laboratory, etc.). The "medical decision-making logic file" includes the medical logic criteria upon which an expert would base a medical decision. The "problem file" contains the decisions made on each patient by the

HELP logic. The problems in each patient's record are used in planning, reporting, and reviewing patient care. Each patient's decision list is available to the physician via terminals or as hardcopy.

A typical sequence of events is as follows: A patient is admitted to the hospital and proceeds through the admission screening laboratory where a computerized self-administered history is taken. This is followed by the acquisition of ECG, spirometry, hematology, vital signs, etc. The storage of this data in the patient's file automatically triggers the HELP system's medical decision-making logic to produce diagnostic suggestions and store these in the patient's decision file.

The steps involved in deriving and processing a medical decision associated with an ECG include the following: (1) the electrocardiogram is sampled by the computer; (2) morphology, rhythm, and other parameters (e.g., RR interval, ST elevation, QRS duration, etc.) are extracted; (3) the extracted parameters are stored in the patient's data file; (4) the result of the analysis of the parameters by HELP is an ECG finding; (5) the findings or decisions are stored in the patient's file; (6) a report based on these findings is generated.

The medical decision-making logic is composed of basic building blocks called HELP sectors. A HELP sector for determining a normal ECG is seen in Figure 2.1. This example is a very simple sector which was selected because it omits the more complex or advanced features of HELP which are sufficiently complex to confuse the uninitiated. Figure 2.1 illustrates an easily understood sequence of logic which would come at the end of a chain of previous decisions. The ECG would be determined

```
NORMAL ECG
FVAL    A and B and C and D and not E
A       Normal Rhythm
B       Normal P
C       Normal QRS
D       Normal ST and T
E       Borderline ECG
```

Fig. 2.1. HELP sector logic for determining a high-level decision concerning the ECG.

to be normal if the final evaluation (FVAL) was found to be true, and this decision, of course, would be stored in the patient's decision list. Items A through E would have been previously evaluated based on ECG parameters and other HELP sector logic sequences. Therefore, this HELP sector simply recalls the previous decisions and makes a higher level decision (normal/abnormal ECG) based in previous decisions and logical statements within this HELP sector. First, the decision of whether A (normal rhythm) was true is recalled, followed by the retrieval of the true or false condition decisions of B, C, D, and E. Then the FVAL is determined. In order for the final evaluation (namely, a normal ECG) to be true, then A and B and C and D would have to be true and E would have to be false. That is, in order for the final evaluation to be true (that this person has a normal ECG), the person would have to have a normal rhythm, a normal P wave, a normal QRS complex, a normal ST and T, and not a questionable or borderline ECG. HELP, having

made a decision, stores the decision in the patient's file.

An additional example of HELP will further illustrate the impact of the total system. Assume a patient receiving an ECG test has an abnormal rhythm due to the effects of digitalis. The medical decision-making logic calls for the analysis of such parameters as digitalis, serum potassium, and serum creatinine from the patient's record in order to evaluate the diagnosis of digitalis toxicity with potassium depletion. Thus, some of the ECG medical decision-making logic may require data not only from the ECG test, but also from the clinical laboratory and the pharmacy.

HELP as a Tool for Generating Populations and Variables

The following is a description of the words "population" and "variable." A population may consist of patients who have blood urea nitrogen (BUN) level greater than 30 mg/dl upon admission. A population consists of identification (ID) numbers of all patients who satisfy a criterion (e.g., BUN > 30). Variables consist of values of a particular parameter for patients in a population. For example, a variable may consist of the maximum bilirubin value for each patient in a population. The process of population or variable generation is accomplished by (1) creating the HELP sector logic and (2) using the logic against the individual patient data to either select patients and/or retrieve the values of variables to be included in the data base. Creating populations using HELP is the first, yet not the only, method used in population-generation.

The following illustrates use of the HELP logic to generate

(1) patient populations and (2) variables for patients in the populations. Assume that a population is to be generated consisting of patients who had a reticulocyte count performed within one day before receiving iron and who also had another reticulocyte count ordered within eight days after receiving the iron. Using the resulting logic, patients would be evaluated individually. Only those patients who met the criteria would be included in the population being generated.

Similarly, in generating variables, first the HELP sector logic is created, followed by a comparison of the logic against the data for each patient in the population. For example, assume that a variable, the lowest hematocrit, is to be created for patients in a population. The logic to select the lowest hematocrit would be generated using HELP. This logic would then be used to extract the lowest value among all the hematocrit values for each patient in the population. Therefore, the lowest (or minus zero for patients with no results) hematocrit value would be the new variable being generated.

The process of creating the HELP logic for the generation of populations and variables may require five to 20 minutes or more depending on its complexity. The comparison of the logic against the patient data is a lengthy process requiring ten seconds to several minutes per population. The actual time may vary depending on the complexity of the logic and the size of the population, but more importantly on the current load on the time-shared computer system.

The FIND System

FIND is a subsystem for rapidly finding patients who satisfy

a set of user-defined medical decisions. The FIND subsystem permits the rapid retrieval of patients that satisfied specified HELP sectors or HELP decision logic during their hospital stay without having to reevaluate the medical decision logic and run the HELP logic against the patient data. To perform the latter function while in the teaching mode is totally unacceptable, due to the excessive time required to generate new populations using HELP.

The following is a description of how the FIND subsystem works: All medical decisions are grouped into broad categories called data classes. For example, cardiology decisions fall into data class 03 (see Figure 2.2). These decisions may be further subcategorized by what are called field codes. Decisions based on the ECG morphological analysis are assigned the field code 00 (see Figure 2.3). Specific decisions then appear under the data class and field code (see Figure 2.4). The FIND subsystem reserves disc space for saving pointers to patients who have had decisions made under certain data classes and field codes (see Figure 2.5). A decision is made on a patient based on the ECG parameters. A numeric code for the decision and a pointer to the patient are stored in the next available word of disc storage reserved for cardiac decisions (see Figure 2.5). As other decisions are made based on the ECG parameters, their respective decision number and patient number are appended to the list. The file structure described for FIND is an inverted file designed to improve retrieval times.

The next section provides further details on retrieving a set of patients who have all had the same medical decision made on them during

Data Class (Octal)	Decision Type
.	
.	
.	
03	Cardiology decisions
.	
.	
.	
06	Blood gas and pulmonary decisions
.	
.	
.	
15	Chemistry decisions
.	
.	
etc.	

Fig. 2.2. Examples of various data classes.

Data Class	Field Code (Octal)	Decision
.	.	
.	.	
.	.	
03	00	Decisions based on ECG morphology data
03	01	ECG decisions requiring pharmacy data
.	.	
.	.	
07	40	Decisions based on history data
.	.	
.	.	
.	.	
etc.	etc.	

Fig. 2.3. Examples of field codes within various data classes.

Data Class	Field Code	Decision Number	Decision Message
.	.	.	
.	.	.	
03	00	17	Left bundle branch heart block
03	00	24	Old myocardial infarction (MI)
.	.	.	
.	.	.	
etc.	etc.	etc.	

Fig. 2.4. Decisions appear under data class and field code.

Data Class	Field Code		
03	00		
Decision Number	**Disc Number**	**Patient's Relative ID Sector Address**	
017	6	7462	NOTE: A patient with a left bundle branch heart block
024	6	7360	NOTE: Patient has old MI
.	.	.	
.	.	.	
.	.	.	
		24	Next available word of disc storage

Fig. 2.5. Example of the FIND subsystem pointers to patients with similar problems.

their hospital stay. The user may desire to retrieve the set of patients from a population who exhibited a left bundle branch heart block. To retrieve such a set, the user would enter the keywords associated with the set of patients desired. In this example the user could enter the keyword "block." The computer would then present the messages of all HELP decisions containing that keyword. Among the list of messages presented would appear the message: "left bundle branch heart block" and the user would select this from the list.

All medical decisions in the HELP system are coded with a number. The left bundle branch heart block is decision 17 within the broad category of ECG decisions (see Figures 2.4 and 2.5) The FIND subsystem has an inverted file containing the decision number and the patient number (patient relative identification sector address) under each of the broad categories referred to previously (see Figure 2.5). In this example, as FIND begins the search for patients who exhibited a "left bundle branch heart block," it accesses data in the inverted file. It begins searching the decisions made under data class 3, field code zero, looking for decision 17 (see Figures 2.4 and 2.5). As it searches the inverted file patients who qualify are included in a separate list consisting of only patients who had a "left bundle branch heart block." After searching the entire list of decisions for patients with cardiac problems, all patients who have had a left bundle heart block and are in the specified population are stored on disc and the list of these patients is presented at the user's terminal. The user then may select one patient at a time from the list to review the individual patient's data, analyzing other patient problems and other

parameters that may be of interest, or he may use the group as a population and perform further statistical analysis.

The Original STRATO System

STRATO is a statistical subsystem for the analysis of clinical data. It is a system used for statistically analyzing patient populations and the variables associated with the patients. The STRATO system was developed by William Miller, a Ph.D. candidate in the University of Utah Department of Medical Biophysics and Computing. The STRATO system has not yet been described in detail in the literature. STRATO gets its name from the fact that the populations and variables may be stratified; that is, HELP may be used to create subpopulations and variables from the data base. For example, using a population of patients who had received more than 50 units of blood, one could use HELP logic to divide this population into two subpopulations, males and females. After these two subpopulations are created, variables for them may be created.

In the original STRATO system, the only method available for population and variable generation was to create HELP sector logic and use it against the data base. This process was described in the previous section explaining the HELP subsystem.

There were several restrictions which made it impossible to use the original STRATO system in a teaching mode. The following are a few of these restrictions: (1) all programs were of such a restricted size that only limited processing (small number of patients) could be done; (2) excessive time was required for the generation of new populations

and variables which meant that all populations and variables had to be generated prior to the interactive session; (3) there was some ambiguity in the user messages, so that the user was frequently confused by the messages; and (4) many options were offered that were not pertinent in the teaching mode. The following section presents the approaches taken to overcome these and other restrictions.

Significant modifications to STRATO were made in ten areas: (1) alterations to user messages; (2) population-generation procedures; (3) variable-generation procedures; (4) FIND pointer generation; (5) a linking of the HELP system to STRATO; (6) patient data review linkage; (7) ranking of patients by a variable; (8) the teaching mode password; (9) variable and population hardcopies; and (10) the audit trail file. Each of these ten new features will now be described in further detail.

Alterations to User Messages

Most of the messages in the STRATO system have been expanded, sacrificing brevity for increased clarity. Messages now present a "prompt," indicating when the user is to respond. The types of messages presented to the user are password-dependent. All passwords other than the teaching mode password present the full spectrum of messages for manipulating, generating, and processing populations and their variables. When using the teaching mode password, messages are greatly expanded for increased clarity and yet actually fewer options are available to

the user, so that there is less chance of destroying the data base or getting into complex processing which may utterly confuse the uninitiated. For example, in the regular STRATO system the user is offered the option of clearing the patient data disc and generating a new patient data disc. This option and its subsequent messages are not appropriate during a teaching session. Therefore it is not seen by users while using the teaching mode password.

Population-Generation Procedures

The user may generate populations using the HELP system, as has always been the case with the STRATO system, or he may now use a new option in generating new populations by specifying the range of a variable associated with a parent population. This assumes the desired variable has already been created for the parent population. Having specified the range on the chosen variable, ID numbers of patients who fall within the specified range are stored in a list as a new population. Thus new populations of patients may be created from existing or parent populations and variables to any level desired. This option is displayed to the user as "generation of subpopulations from a parent population and a parent variable." Simultaneously, the patient numbers may be saved for patients (outliers) whose variable value was not in the specified range. To illustrate, the variable containing serum iron values for the parent population might be searched to generate one population of patients with a serum iron between 0 and 50, and another group who did not have a serum iron in that range. Appropriate text is provided by the user to identify each population for use in

subsequent stages of analysis.

Another population-generation method is available which permits a population to be created consisting of only patients that are common to two other populations. For example, one subpopulation may contain patient numbers for all males and a second for all patients who received one or more blood transfusions. It may be desirable to create a new subpopulation consisting of all males who received a transfusion. This option permits the generation of such a subpopulation by creating a new population from all patients who are common to the two aforementioned subpopulations.

Variable Generation Procedures

Two new methods are available to the user to generate new variables for subpopulations. In the above example under "population-generation" the user specified a range on the variable that a patient must satisfy to be included in the new population. At the time the patient number is saved in the new population, the value for the variable used as the discriminator is also saved for the patients who had satisfied the range that has been specified by the user. So we may generate a new population and a new variable simultaneously. The variable so generated may be discarded or saved at the user's option. The same is true for generating variables for the outliers (those who did not fall within the range specified).

The second method available for generation of variables is accomplished by specifying a parent variable, a parent population, and a subpopulation. This procedure generates a variable for the subpopulation using the parent variable and parent population. For example,

a parent population might consist of patients who had a hematocrit less than 30, and a parent variable of maximum bilirubin values for each of the patients in the parent population. Having created a subpopulation of males who had a hematocrit less than 30, we could create a new variable for the males, from the parent population, and parent variable. The new variable generated would consist of the maximum bilirubin values for each of the males with a hematocrit less than 30. The main idea is that if an existing variable is already saved for the parent population, the user should be able to easily generate the same variable for a subpopulation without having to go through the more time-consuming approach of using HELP again.

FIND Pointer Generation

The FIND system has been discussed previously. It is a system for rapidly retrieving patients who have had a certain set of medical decisions made during their hospital stay. The FIND pointers themselves consist of a decision number and a patient number, grouped into broad classes and stored on random access storage for rapid retrieval. The FIND pointers are generated in the following manner: The patient data is read from magnetic tape and the decision list for the patient is searched to retrieve the decision numbers. The patient's relative ID sector address on the patient data disc (which was created from tape) and the decision number (or HELP sector number) are saved in the inverted file that has been previously discussed. Every decision in the decision list for every patient on the tape is stored in the file. This permits rapid access to patients who satisfy the FIND

criterion (any logical combination of HELP decisions) specified by a user. For example, if the user desires to find all patients in a population who had angina pectoris and normocytic hypochromic anemia, he can specify these criteria as a search strategy and all patients who satisfy that criteria would be arranged in a list as a potential new population available for review and analysis.

Linking of HELP System with STRATO

The HELP compiler is used to create the medical decision-making logic. It is used with the STRATO system to create the logic for population and variable generation. In the original version of STRATO, it was impossible to generate HELP sectors or HELP decision logic while using the STRATO system. One had to terminate the current processing in the STRATO system and enter the HELP compiler. HELP sectors or decision logic needed for future processing of variables and populations under STRATO control are created exclusively by using the HELP compiler. Therefore, a direct link between the HELP compiler and the STRATO system has been developed. An option available in the first option list presented in STRATO permits the user to transfer control directly to the HELP compiler and then return to STRATO without terminating the STRATO system. This is beneficial because it makes possible the use of two separate systems without interruption or any perception of an annoying starting/stopping sequence that is otherwise observed. On entering the HELP compiler, all processing proceeds in the normal manner except when in the teaching mode. This exception will be discussed in a subsequent section. Alterations have been made to the HELP

compiler such that it recognizes when it was initiated from STRATO, so that it presents the option of returning to STRATO at the appropriate time.

The Patient Data Review Linkage

The standard programs in the system for reviewing patient data such as (1) admit diagnosis; (2) discharge diagnosis; (3) complications and surgical procedures; (4) laboratory data; (5) medical decisions in the decision list; (6) blood gas data, etc., have all been modified so that the users of STRATO may use each of these programs to review the selected patient data for the patients in the study. The linkage permits the user to return to the STRATO system after reviewing patient data. All linkages are transparent to the user, so that he does not have to terminate the STRATO system to review patient data. After reviewing the data on a set of patients, the user may select the option of saving the patients that he has reviewed in a new population. The user may select one of three options in doing so: (1) save all of the patients that he has reviewed or that are in the list available to be reviewed; (2) he may save all of the patients in the list except those that he indicates not to save; or (3) he may specify certain patients from the list to be included in a new population in the STRATO system. To illustrate: the FIND system may be used to retrieve a set of patients who had an acute myocardial infarction and a heart block. Ten such patients may be retrieved from the population of patients with a hematocrit less than 30. Their patient numbers, length of stay, etc., would be displayed at the terminal and the user could then save those

ten patients as a new population in the STRATO system. After reviewing data on each of the ten patients retrieved with FIND, the user may have noted that two of them died while hospitalized. The user may, for some reason, want to reject any patient who died from the new population being generated. Therefore he could eliminate the two who died and save the other eight patients in a new population in the STRATO system.

Ranking of Patients by a Variable

It may be desirable to arrange all patients in a population such that their patient numbers appear in order of their highest maximum bilirubin value. For example, a population of males may have been previously generated but now we would like to see the male patient numbers reordered in descending order on their maximum bilirubin values. A user may specify any population or variable that he wants sorted in ascending or descending order. In the above example the descending order would be selected and a population would be generated with the patient numbers rearranged, such that the patient with the highest maximum bilirubin would appear as the first patient number in this population, and the last patient would have the lowest maximum bilirubin value. This permits the user to review patient records for outliers with high or low values, depending upon whether the population is sorted in ascending or descending order on a variable, in an attempt to investigate why certain patients had high or low values for the variable under consideration.

The Teaching Mode Password

The teaching mode password causes an expanded option list to appear at the user's terminal. The messages are more lengthy than similar messages presented in STRATO under the nonteaching mode passwords. An effort has been made to make it a cordial system for the uninitiated. Therefore in the teaching mode, the users may proceed independently with little tutoring from an experienced user. As one enters the HELP compiler from STRATO under the teaching password, the HELP compiler recognizes it and does not request the data class and field code, or any password of the user. STRATO in the teaching mode passes a preassigned data class and field code to the HELP compiler automatically. The underlying philosophy for developing a teaching mode has been to provide rapid interactive processing at all levels in the system, so that long delays (more than 30 seconds) are eliminated. The teaching mode has been planned and designed so that a faculty member with one hour of demonstration and practice may make use of all of the powerful options that are available to him for instructing students in the teaching mode.

Variable and Population Hardcopies

When using the STRATO system the user had to list the variables and populations and record the appropriate indexes manually or frequently display the lists interactively to refresh his memory concerning which index to use. Therefore it was apparent that a printout of both the populations and the variables was needed. Appropriate software was created to produce these lists. Examples of variable and

population printouts appear in the tables shown in Chapter III.

The Audit Trail File

Another problem for users of the STRATO system was that all recordings of statistical outcomes had to be done manually as they occurred. This added burden naturally extended the time limit for user investigation. An automated data collection scheme was implemented which could be activated by the user. This permitted the user to perform all investigations, confident that the paths being pursued would be automatically tracked by the system. This is known as the audit trail file. The audit trail file contains such information as the date and time of the investigation, the course pursued, statistical outcomes of each test used, and appropriate messages entered by the user. The messages could act as a further explanation of special conditions.

Nonauthored CAI System for Teaching Decision-Making

This section is a discussion of a nonauthored CAI system for teaching the stochastic model from laboratory data. It will give an overview of the sequential steps in the procedures followed during a typical teacher/student learning session. A typical teaching session is conveniently divided into three phases: (1) preparation of the data base; (2) conversational interaction of the teacher, the student, and the computer; and (3) hypothesis testing and evaluation. The characteristics and activities involved in each of these phases follows.

Preparation of the Data Base

The Medical Biophysics and Computing Department (MBC) at the LDS Hospital maintains archival tapes containing information recorded on each patient admitted to the hospital during his hospital stay.

From these archival tapes, other subtapes may be created, transferring only those patients who meet certain criteria. All of the patients on the resulting subtape form the original population in the data base. The subtape contains all available data collected on each patient during the hospital stay. All of the data or selected portions of it are transferred to the patient data disc.

Using the patient data disc and the original population, the remainder of the data base can be created. The remaining steps include: (1) the creation of HELP logic, defining further populations and variables; and (2) the generation of populations and variables using HELP logic. The variables generated during this phase consist of values for parameters associated with patients in the original population. The archival tapes for a selected time interval are searched and a subtape created. The data for patients on the subtape are transferred to the patient data disc and the original population is thus generated.

All further processing is based on this original population. All subpopulations will consist of a subset of patients from the original population. Hence the original population is frequently referred to as the parent population.

Similarly, variables associated with the original population are called parent variables. The following are some examples of

typical variables that might be generated for investigating an anemic population: serum iron, maximum iron binding capacity, maximum reticulocyte count following iron therapy, and the minimum hematocrit.

All population and variable generation procedures are accomplished via the HELP system prior to the interaction of the student with the system since this requires considerable time to complete as noted previously.

Conversational Interaction of Teacher, Student, and Computer

The system is designed primarily to support faculty-student investigations, where the instructor leads the investigation by giving each student a copy or computer-produced listing of the parent populations and variables. He presents the notions of (1) a population and (2) a variable. For example, he indicates that a population has the following five attributes (see Figures 2.6 and 2.7 for illustration of these five characteristics of populations): (1) a name associated with it; (2) an associated index number; (3) number of patients in a population; (4) a sequence of numbers which are patient identification numbers; and (5) the patient ID numbers positionally identified (e.g., the fifth patient in Fig. 2.7 has a patient ID of 173458).

Similarly a variable has (1) a name; (2) an index number; (3) a number of entries; (4) values for the variable; and (5) the values positionally arranged in the same order as the patient numbers in the population. Figures 2.8 and 2.9 illustrate these attributes.

The instructor illustrates the relationship between a variable and a population. For example, he may elucidate their relationship by

Index	Number of Entries	Name of Population
1	384	Diabetic
2	265	Diabetic, female
3	119	Diabetic, male
.	.	
.	.	
.	.	
36	27	Diabetic, age < 13
37	6	Diabetic, male, age < 13
.	.	
.	.	
.	.	

Fig. 2.6. Illustration of characteristics of populations.

Index	Name			
37	Diabetic, male, age < 13			
Count	Patient Numbers			
1	276534	295437	372195	527426
5	173458	654127		

Fig. 2.7. Example showing populations consisting of patient ID numbers.

Index	Number of Entries	Name of Variables
1	384	Maximum glucose
2	384	Maximum creatinine
.	.	
.	.	
.	.	
48	27	Maximum glucose, age < 13
.	.	
.	.	
.	.	
51	6	Maximum glucose for male diabetic children
.	.	
.	.	
.	.	

Fig. 2.8. Illustration of attributes of a variable.

Index	Name			
51	Maximum glucose for male diabetic children			
Count	Values			
1	350	407	703	386
5	647	506		

Fig. 2.9. Illustration of values for a variable.

showing that the fifth male diabetic child (patient #173458) had a maximum blood glucose level of 745 mg/dl during his hospital stay.

Once the students have mastered these fundamental concepts, the instructor leads them through similar explanations on how to generate (a) subpopulations and variables and (b) histograms. The instructor also demonstrates how to perform statistical analysis using selected patient parameters and how to use FIND to locate sets of patients with similar problems. These demonstrations and explanations (1) familiarize the student with the data available; (2) illustrate the capabilities of the system; and (3) illustrate the logic, straight-forwardness, and ease of using the system.

The above activities introduce the students to the system. The instructor then proceeds to ask questions that can be answered using the data available. The subsequent discussion is a hypothetical case of an investigation by an instructor and a few students studying diabetes. This section will present the case as seen by another observer.

Teacher: What are the two major kinds of diabetes and how can you determine which kind is exhibited in a patient?

Student A: Diabetes insipidus and diabetes mellitus.

Teacher: And how do you distinguish each of them clinically?

One student said he thought the specific gravity of the urine was higher for the diabetes mellitus patient. Another student disagreed, so the teacher used the nonauthored CAI system to generate a histogram to compare the specific gravity of the two groups. There was no significant difference between the two groups, as the teacher expected, and he explained that the physiological mechanisms

responsible for the low specific gravity is the same in both forms of diabetes. This permitted the students to see that questions can generate hypotheses and that the data can then be analyzed to confirm or reject such hypotheses.

Another student recalled that both forms are metabolic disorders but that diabetes insipidus is caused by injury to the neurohypophysial region while diabetes millitus is caused by hypoinsulin production in the pancreas. Further he suggested that this being so, one would expect the ADH levels in the two groups to be different. Since ADH is released by the neurohypophysis to a lesser extent in insipidus patients, a test for blood levels of ADH should reveal which type is indicated. The instructor pointed out that his reasoning was logical, but that the test for the hormone suggested was not frequently done due to economic restraints. Further he stated the types may be generally separated by three common tests: (1) volume of urine output over time; (2) urine ketone levels; and (3) blood glucose levels. The instructor then compared the data on the first two parameters for both subpopulations. The urine output per unit/time and blood glucose level of diabetes mellitus significantly exceeded that of diabetes insipidus patients. Diabetes mellitus patients also excreted more ketone bodies in the urine than did the insipidus subpopulation. The instructor suggested that they reexamine the data, this time looking for trends and thresholds that might be formalized into a decision-making algorithm.

This hypothetical example illustrates the sequence: (1) ask clinically relevant questions; (2) consider alternatives and form hypotheses; (3) test these hypotheses using the system; (4) make

generalizations from the analysis; (5) apply to individual patients for decision-making. Step #5 may be executed in a routine manner (incorporated into the physician's normal decision-making) or it may occur in a more formal setting as outlined in the next section.

Hypothesis Testing and Evaluation Phase

This phase assumes some prior experience or instruction on the use of the HELP compiler and the FIND sistem. Therefore, depending on the depth of the investigation, this phase may be optional. The student uses the HELP compiler to articulate to the computer the logic and the data parameters to be used as criteria for a particular decision, based on analysis of the data and the resulting hypotheses. This process transforms the deductions made from the hypotheses that were tested into an algorithm that the computer can evaluate. The resulting logic or algorithm is called a HELP sector. The clinical computer uses this logic to evaluate patients who are currently hospitalized, making note of each patient who satisfies the logic specified in the HELP sector.

After the computer has made decisions using this logic on hospitalized patients for a few days, the student may begin to measure the effectiveness and accuracy of his logic. To do this, the student uses the FIND system. FIND is a system designed to rapidly find and display patients who satisfy the logic in specified HELP sectors. Using the FIND system, students can follow the patients who satisfy their decision-making HELP sectors, monitor their progress, and evaluate how well their logic performs. This important step completes the cycle of

analyzing data on individual patients, forming generalizations applicable to and verified by populations, and then applying the knowledge gained to make decisions in the individual case. It is important to point out that this approach will aid the user in identifying both the true positives and false positives. It does not separate out the false negatives.

CHAPTER III

Procedures for Evaluation of CAI Systems

The proliferation of CAI systems has prompted limited investigations concerning the effectiveness of these systems. The difficulties of actually obtaining measurable results with teaching systems are well-known (Rubin, et al., 1975). Because of the complexities involved in such evaluations, researchers have generally employed less rigorous evaluative techniques, such as cost-effectiveness and speed of student learning (Molnar, 1976). Much of the evaluation reported in the literature is subjective at best. For example, some authors consider such questions as "Will students use the system?" or "What has been the user response?" A research team at Stanford recently developed a CAI system designed to teach the computer language, BASIC. Their evaluation simply indicated that student reaction was favorable. Similarly the evaluation cites some student complaints about problems with the system (Barr et al., 1975). This section presents an overview and the rationale for the evaluative procedures employed in (1) frame-oriented medical CAI systems; (2) nonauthored CAI systems; and (3) the system described in this thesis.

Frame-Oriented Medical CAI Systems

Methods used to evaluate frame-oriented medical CAI systems

noted above have not been significantly different from those of the Stanford study. For instance, in the network experiment sponsored by the Lister Hill National Center for Biomedical Communications (LHNCBC) referred to in Chapter I (Rubin, 1976) faculty members were asked to evaluate selected CAI courses as part of the experiment. Their evaluations were subjective as revealed by their reports. One group said the ". . . program was by far the most popular program among the student users. 'It was referred to as a great learning experience and challenge. . . . It was fun and educational.'" (Rubin et al., 1975, p. 38) Another group attempted a more rigorous evaluation of another course by trying to correlate student performance on the board scores with CAI usage, but technical difficulties were encountered and "the only data produced were subjective." (Rubin et al., 1975, p. 45).

A bibliography on medical CAI systems, in preparation for publication, by Cochran and Stolurow (1977) has been received via a personal communication. It includes reports on many systems, some of which have some form of an evaluation. The systems reported on may be divided into three groups: (1) systems that are demonstrations only--no evaluation is reported; (2) systems with demonstration and and a subjective evaluation; and (3) systems demonstrated with some form of more rigorous evaluation. In this bibliography there were 56 papers reporting on systems in group 1 above (no evaluation, only demonstration). Twelve articles reported on system demonstrations with only subjective evaluations. These evaluations consisted primarily of users' attitudes, comments, and the general acceptability

of each system. There were ten systems that reported both a demonstration and some type of formal evaluation. Nearly all used a pre- and posttest evaluation technique.

Nonauthored CAI Systems

Methods used to evaluate nonauthored CAI systems are discussed in this section. The nonauthored systems are far fewer and much more complex than the frame-oriented systems. Because of the increased complexities of these systems the authors reporting on them have placed the most emphasis on internal algorithms and technological achievements. Consequently, all such systems identified have given only subjective evaluations, if any were presented at all. For example, one nonauthored CAI system attempted an evaluation by a user attitude survey sheet (Koffman et al., 1975). Another did not perform any evaluation but was only a demonstration (Uttal et al., 1970). A third system reported that students generally found the system to be "helpful and an enjoyable way to learn how to employ classroom concepts in solving problems." (Koffman, 1973, p. 188) The students were also asked to voice their complaints, and the major ones are included as part of the evaluation. This evaluation indicated all participants wanted to continue using the system.

Nonauthored Medical CAI System as Described in This Thesis

This thesis gives an account of the demonstration of a nonauthored medical CAI system that has been developed. A similar approach for presentation has been taken here as has been noted in the

two previous sections. That is, the system has been tested and demonstrated, using student subjects, and evaluated, using data from the demonstrations and attitude appraisals of the participants. The teaching sessions with each user were taped so that subsequent analysis could be conducted. This facilitated not only the capture of user comments and recommendations but also the timing of certain parameters worthy of study. The next section presents an overview of the demonstrations and the details of the evaluation.

Details of Demonstration of Nonauthored Medical CAI System

After some initial testing of the system using a variety of subjects, a medical subject for the demonstration was selected and a data base generated. Only four participants were identified and committed to take part. The teaching sessions took the following form: (a) Introduction to the concept of studying empirical data on populations to arrive at a generalization that may be used in medical decision-making at the individual patient level. (b) The notion of populations and variables were presented to the participants with a list of both as they existed before the session commenced. (c) The participants were shown the capabilities of the system by actually demonstrating the various options available to the user. This included demonstrating the statistical capabilities. (d) The participants were then asked questions of some clinical relevance to get them started with the investigation. The pursuit of the answers to one question frequently generated others worth investigating. The goal was to explore a new mode of answering questions leading to relevant empirical data for

medical decision-making.

The remainder of this chapter presents some characteristics and conclusions about the system. These are listed and the details of each are presented in the order shown: (1) setting up the data bases; (2) a profile of the participants; (3) observations on learning to use the statistical capabilities of the system; (4) the audit trail files; (5) responses from the participants; (6) requested improvements in the system; (7) observations on the course selected participants pursued during their investigations; and (8) competencies required of the instructor and student.

Setting Up the Data Bases

The data bases consisted of populations and variables and a patient data disc. Before the demonstrations began, a single data base was generated. Originally it consisted of one population, 16 variables, and the patient data on disc. Table 3.1 is a brief description of the original, or parent population. Others were added as the demonstrations continued.

Table 3.2 is a sample of the variable printout. This was an investigation of an anemic group of patients. The criterion for inclusion in the data base was an admitting hematocrit ≤ 34. The method of population and variable generation is included in Chapter II. The following gives some of the times required to generate the data base. The transfer of patient data from the monthly archival tapes to the subtape (outlined in Chapter II) took place in about 12 minutes. This subtape (containing only patient data for patients with a hematocrit ≤ 34) was

Table 3.1. The Original, or Parent Population

Index	Number in Group	Message
1	503	PATIENTS ADMITTED WITH A HCT < OR = 34 SEPT - DEC 76

Table 3.2. The Original, or Parent Variables

Index	Number in Group	Message
1	503	PATIENTS ADMITTED WITH A HCT < OR = 34 SEPT - DEC 76, MIN IRON CONTENT
2	503	PATIENTS ADMITTED WITH A HCT < OR = 34 SEPT - DEC 76, MAX IBC
3	503	PATIENTS ADMITTED WITH A HCT < OR = 34 SEPT - DEC 76, RETICULOCYTE COUNT MAX
4	503	PATIENTS ADMITTED WITH A HCT < OR = 34 SEPT - DEC 76, MCV AT TIME OF LOWEST HCT
5	503	PATIENTS ADMITTED WITH A HCT < OR = 34 SEPT - DEC 76, MCH AT TIME OF LOWEST HCT
6	503	PATIENTS ADMITTED WITH A HCT < OR = 34 SEPT - DEC 76, MCHC AT TIME OF LOWEST HCT
7	503	PATIENTS ADMITTED WITH A HCT < OR = 34 SEPT - DEC 76, MIN HGB
8	503	PATIENTS ADMITTED WITH A HCT < OR = 34 SEPT - DEC 76, MIN HCT
9	503	PATIENTS ADMITTED WITH A HCT < OR = 34 SEPT - DEC 76, MIN RBC
10	503	PATIENTS ADMITTED WITH A HCT < OR = 34 SEPT - DEC 76, MIN BILIRUBIN
11	503	PATIENTS ADMITTED WITH A HCT < OR = 34 SEPT - DEC 76, MIN WBC
12	503	PATIENTS ADMITTED WITH A HCT < OR = 34 SEPT - DEC 76, MAX WBC
13	503	PATIENTS ADMITTED WITH A HCT < OR = 34 SEPT - DEC 76, BONE MARROW CELLULAR, MAX
14	503	PATIENTS ADMITTED WITH A HCT < OR = 34 SEPT - DEC 76, BONE MARROW HYPOCELLULAR, MAX
15	503	PATIENTS ADMITTED WITH A HCT < OR = 34 SEPT - DEC 76, BONE MARROW HYPERCELLULAR, MAX
16	503	PATIENTS ADMITTED WITH A HCT < OR = 34 SEPT - DEC 76, M. E. RATIO

then used to create the population, which required 90 seconds to complete. Transfer of patient data to the patient data disc required 2 hours and 20 minutes. The generation of the FIND pointers required less than 20 minutes.

The generation of the 16 variables was accomplished in approximately 40 minutes. The transfer of the patient data from tape to the patient disc, the generation of the FIND pointers and the variables was accomplished during the early morning hours when the other demands on the system were insignificant. The same tasks attempted during normal working hours could require two to five times longer to complete.

The explanation in Chapter II under data base generation strongly suggested that careful consideration be given to the selection of the populations and variables so that subsequent passes through the patient data disc might be avoided. Even with careful planning, 11 additional populations and 16 additional variables had to be generated (using HELP), after the initial data base had been generated. This literally halted some teaching sessions because questions had been asked for which data were not available. The questions could not be resolved until the further generation was completed.

Tables 3.3 and 3.4 list these additional populations and variables that were subsequently generated via HELP logic. The questions that required this additional data will be discussed later. Some of these additional variables required two to three times the time to generate as it took to create the original 16. This is understandable because of the competition involved with other uses for accessing the disc drives, as these were completed often during regular working hours.

Table 3.3. Additional Populations Generated via HELP Logic.

Number in Group	Message
9	*PATIENTS ADMITTED WITH A HCT < OR = 34 SEPT - DEC 76, PATIENTS WITH LOW SERUM IRON < 50, BUT HI IBC*
70	*PATIENTS WITH A SERUM IRON VALUE, IRON AND IBC MEASURED*
40	*PATIENTS WITH A SERUM IRON VALUE, (C) LOW IRON, NORMAL IBC*
1	*PATIENTS WITH A SERUM IRON VALUE, (B) NORMAL IRON, ELEVATED IBC*
19	*(D) NORMAL IRON, NORMAL IBC*
101	*PATIENTS WITH A RETIC COUNT*
9	*PATIENTS WITH A RETIC COUNT, RETIC BEFORE AND AFTER IRON*
108	*PATIENTS ADMITTED WITH A HCT < OR = 34 SEPT - DEC 76, MALES*
394	*PATIENTS ADMITTED WITH A HCT < OR = 34 SEPT - DEC 76, FEMALES*
113	*PATIENTS ADMITTED WITH A HCT < OR = 34 SEPT - DEC 76, PATIENT DISCHARGED WITH DX DELIVERY*
370	*PATIENTS ADMITTED WITH A HCT < OR = 34 SEPT - DEC 76, NOT DELIVERY, LABOR PATIENT*

Table 3.4. Example of Additional Variables Generated via HELP Logic.

Number in Group	Message
51	*IRON LESS THAN 50, PROBABILITY OF LOW SERUM IRON*
19	*PATIENT WITH SERUM IRON > 50, PROBABILITY OF LOW SERUM IRON*
9	*PATIENTS WITH A RETIC COUNT, RETIC BEFORE AND AFTER IRON, MIN RETIC BEFORE ADMINISTERING FE*
9	*PATIENTS WITH A RETIC COUNT, RETIC BEFORE AND AFTER IRON, MAX RETIC AFTER ADMINISTERING FE*
90	*PATIENTS ADMITTED WITH A HCT < OR = 34 SEPT - DEC 76, PATIENT RECEIVING IRON, AVE HGB FROM 3RD TO 9*
503	*PATIENTS ADMITTED WITH A HCT < OR = 34 SEPT - DEC 76, UNITS OF BLOOD RECEIVED*
209	*PATIENTS WHO RECEIVED BLOOD, UNITS OF WHOLE BLOOD RECEIVED*
209	*PATIENTS WHO RECEIVED BLOOD, UNITS OF PACKED RBC*
209	*PATIENTS WHO RECEIVED BLOOD, UNITS WHOLE BLOOD + RBC*
503	*PATIENTS ADMITTED WITH A HCT < OR = 34 SEPT - DEC 76, MAX CREATININE VALUE*
503	*PATIENTS ADMITTED WITH A HCT < OR = 34 SEPT - DEC 76, TREND IN WBC*
503	*PATIENTS ADMITTED WITH A HCT < OR = 34 SEPT - DEC 76, TREND IN POLYS*
503	*PATIENTS ADMITTED WITH A HCT < OR = 34 SEPT - DEC 76, TREND IN LYMPHS*
503	*PATIENTS ADMITTED WITH A HCT < OR = 34 SEPT - DEC 76, PATIENT AGE*
8	*MACROCYTIC ANEMIA PATIENTS (MALES AND NON OB FEMALES), SERUM FOLATE*
8	*MACROCYTIC ANEMIA PATIENTS (MALES AND NON OB FEMALES), FOLIC ACID ADMINISTERED*

A second data base was created at the request of one of the participants. It was interesting to note, however, that other participants became interested and used it. The other data base consisted of patients admitted to the hospital with a blood urea nitrogen (BUN) greater than 20. The parent populations contained 594 patients. Two subpopulations were also created: males and females. There were 15 variables generated originally in this data base. Tables 3.5 and 3.6 are the original populations and variables in this data base. Generation times experienced for this data base were generally comparable to those in the anemia data base with two exceptions. The transfer of patient data to the patient data disc and the variable generation times exceeded that of the same operations in the other data base by approximately 15 minutes and 10 minutes respectively. This generation required the processing of 91 more patients than did the other. With this second data base, additional populations and variables were generated later using HELP, just as occurred with the anemia data base.

It is impossible to anticipate all the variables and populations that will be needed from the original data base before a session begins. One question answered generates other questions, the answers for which frequently require additional data, and the data are available only via techniques which use the HELP logic.

Profile of the Participants

The selection process for participants was quite arbitrary; some were specifically asked to participate and others became involved after hearing about the system.

Table 3.5. Original Populations for BUN Study

Index	Number in Group	Message
1	524	PATIENTS WHOSE 1ST BUN WAS > 20
2	325	MALES
3	242	FEMALES

Table 3.6. Variables Generated Original BUN Populations

Index	Number in Group	Message
1	594	PATIENTS WHOSE 1ST BUN WAS > 20, MCV AT TIME OF LOWEST HTC
2	594	PATIENTS WHOSE 1ST BUN WAS > 20, MIN HCT
3	594	PATIENTS WHOSE 1ST BUN WAS > 20, UNITS OF BLOOD RECEIVED
4	594	PATIENTS WHOSE 1ST BUN WAS > 20, UNITS OF WHOLE BLOOD RECEIVED
5	594	PATIENTS WHOSE 1ST BUN WAS > 20, UNITS OF PACKED RBC
6	594	PATIENTS WHOSE 1ST BUN WAS > 20, UNITS WHOLE BLOOD + RBC
7	594	PATIENTS WHOSE 1ST BUN WAS > 20, PATIENT AGE
8	594	PATIENTS WHOSE 1ST BUN WAS > 20, MIN HCT
9	594	PATIENTS WHOSE 1ST BUN WAS > 20, MIN BUN
10	594	PATIENTS WHOSE 1ST BUN WAS > 20, MIN POTASSIUM
11	594	PATIENTS WHOSE 1ST BUN WAS > 20, MAX CREATININE
12	594	PATIENTS WHOSE 1ST BUN WAS > 20, MAX BUN
13	594	PATIENTS WHOSE 1ST BUN WAS > 20, GLUCOSE AT TIME OF MAX BUN
14	594	PATIENTS WHOSE 1ST BUN WAS > 20, URIC ACID NEAREST 1ST BUN
15	594	PATIENTS WHOSE 1st BUN WAS > 20, MAX POTASSIUM

Many physicians and other health professionals have seen the system but data were collected using only 10 persons. A broad spectrum of users was considered essential to get an evaluation from faculty and students at many levels. Figure 3.1 presents the full spectrum and number of participants.

Only four of the participants had experience with statistics prior to the demonstrations. The first-year residents included three in family and community medicine and two internal medicine residents. One of the internal medicine residents was very familiar with anemia (one of the data bases studied) as he had presented a seminar on the

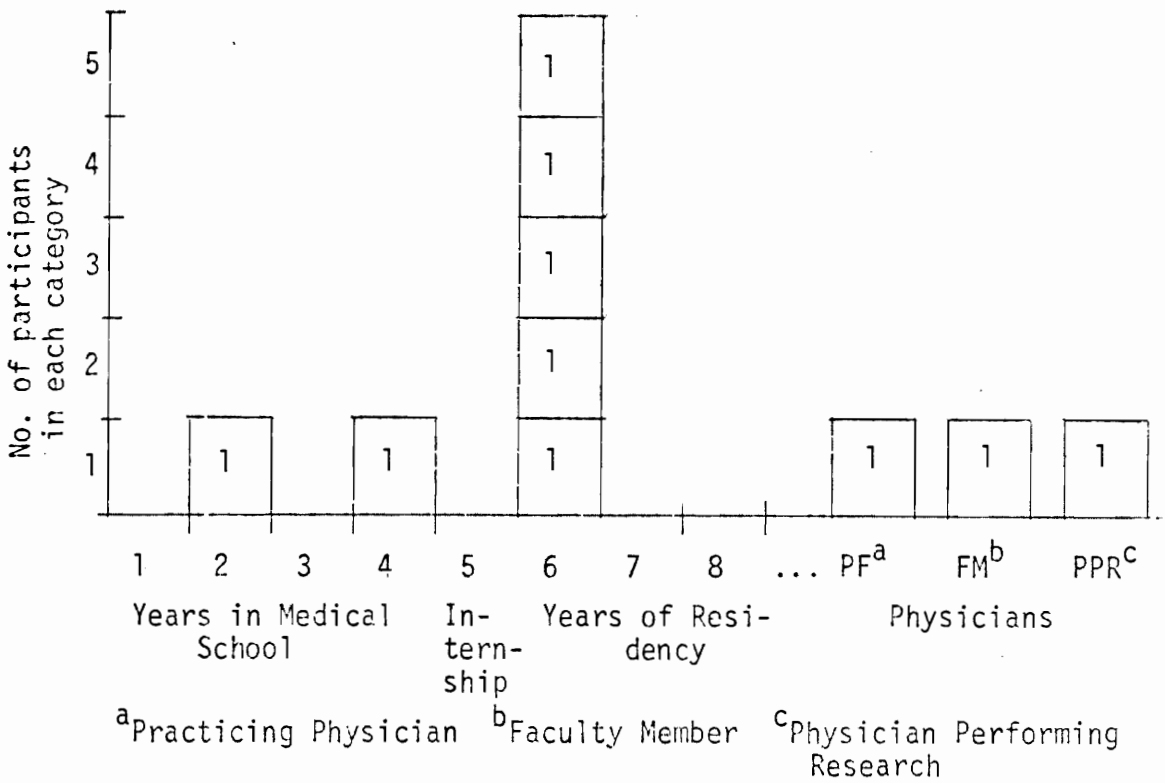

Fig. 3.1. Spectrum and number of participants.

subject to the house staff two weeks before becoming involved in this study. The practicing physician is a hematologist and the faculty member is a cardiologist. The research physician is pursuing opthalmological research studies.

Observations on Learning to Use the Statistical Capabilities of the System

The following reviews the principal statistical tests available to the users and gives an account of their responses to this portion of the system.

The system permits the user to compare data statistically by using three tests. The t test is a parametric statistic and assumes a normal distribution. This test is to be utilized when studying populations with an N less than 30. The other tests include the Kolmogorov-Smirnov and the Mann-Whitney. These are nonparametric tests which do not assume a normal distribution. These nonparametric tests are used as tests for significance. For example, a p = .005 for the Kolmogorov-Smirnov test means there are only five chances in 1000 that the sets of samples were drawn from the same population. Students were taught to rely primarily on the p value from the Kolmogorov-Smirnov test of significance. These tests, aided by the histograms that were displayed before performing the tests for significance, gave the users confidence that they could perform a relatively complex analysis with great ease. Though they did not have an appreciation for the theoretical basis or the derivation of such statistics, with few examples and minimal explanation they proceeded without major difficulty.

In contrast to the comparison mechanisms just presented, the

users were greatly aided by the scatter plots produced when using regression analysis, and by the correlation coefficients to determine the correlation of selected variables. All these statistical analysis mechanisms existed in the original STRATO system.

Before seeing the system's statistical capabilities demonstrated, nearly all participants expressed some apprehension about their using statistics because of a void in their training in this discipline. This is demonstrated in Figure 3.2. A "Y" in the box represents the participant expressing apprehension concerning statistics; an "N" represents participants who expressed no such apprehension.

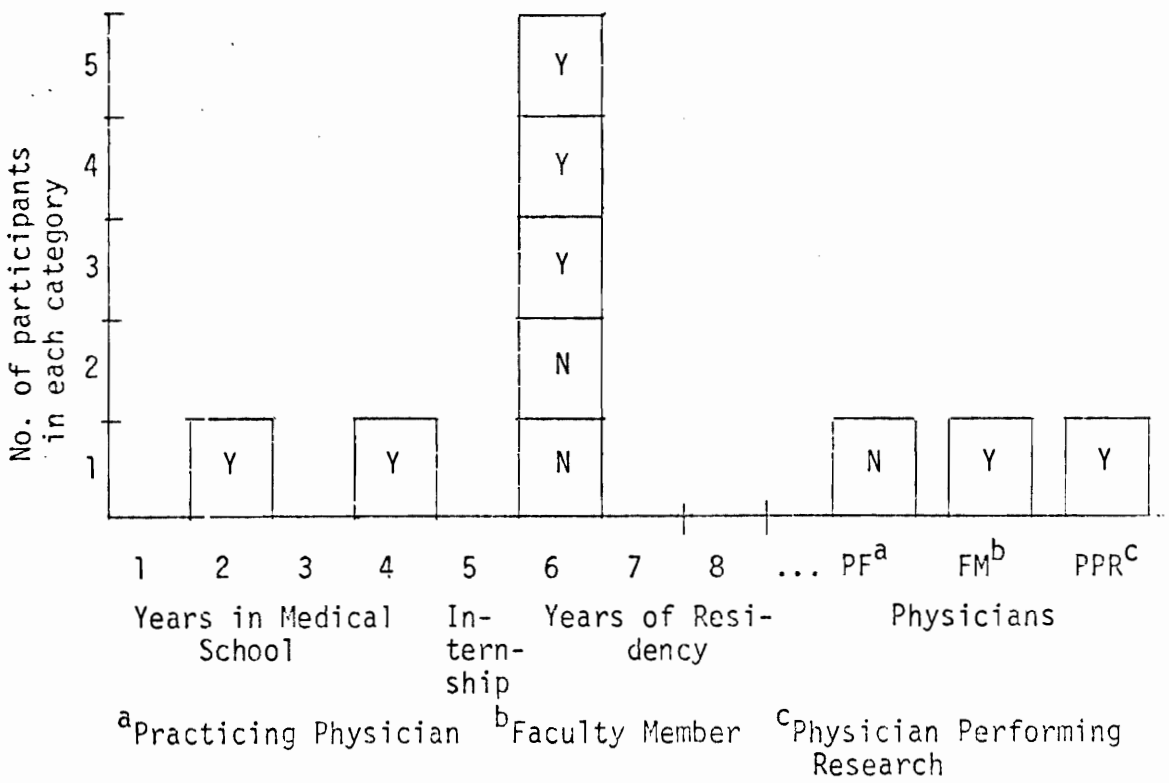

Fig. 3.2. Participants expressing apprehension concerning the use of statistics. ("Y" = apprehensive; "N" = not apprehensive)

Though nearly all participants expressed apprehension concerning the use of statistics, all used the system with ease by the end of one session. All but one participant who expressed apprehension at the outset later commented on the ease of utilizing this feature of the system.

The statistical capabilities were introduced by first explaining the appropriate use of each statistic, followed by examples of comparisons and correlations. This process took approximately five minutes generally, though the explanation for one participant lasted 15 minutes. All participants used the statistical components, as demonstrated by an average of seven comparisons and nine correlations per session.

To summarize, nearly all participants expressed apprehension about using statistics due to a void in their training; however, the participants later demonstrated confidence in their abilities by using the statistical tests frequently in the course of their investigations. They were not intimidated by this aspect after one session.

The Audit Trail Files

One of the problems in evaluating any system is how to capture the data. With a CAI system, the ideal is to design the system so that the evaluation data capturing algorithms are an integral part of the system itself. That is, the system should automatically capture the data desired without manual interference or in some cases even human perception. A plan for how to do this in this system was unfortunately not conceived until about half-way through the

demonstration. However, the plan was quickly implemented and limited data were automatically collected on the last few participants. Further refinements to the data capture algorithms were made subsequent to the final demonstrations, so that future use of the system for research or teaching will reflect even more details of investigations by the users. The data capture technique took the form of an audit trail. The audit trail displays the details of the sequence of options selected by the participant and the results of selecting each option. When coupled with the list of populations and variables the audit trail portrays a good picture of the course pursued by the participant.

The following are some audit trail examples from investigations by a participant. They illustrate how the audit trail in conjunction with the list of populations and variables may be used to review participants' activities.

Example 1. Table 3.7 is a list of the populations. Table 3.8 is a list of the variables, and Table 3.9 is a listing of the audit trail for one of the participants. Notice that the audit trail indicates the time and date of the teaching session and permits the user to append any appropriate message for the session. In this case the message indicates that this was the first session with Dr. ---.

This audit trail seems to show that the physician was comparing the HCT with itself; however, the tape of the session indicates that the participant detected that variables 2 and 8 were redundant and these first two items in the audit trail reflect the fact that he

Table 3.7. List of Populations in Example 1

Index	Number in Group	Message
1	524	PATIENTS WHOSE 1ST BUN WAS > 20
2	325	MALES
3	242	FEMALES
4	17	PATIENTS ON DIGITALIS TYPE MEDICATIONS
5	55	PATIENTS ON GENTAMICIN
6	379	PATIENTS ON DIURETICS
7	1	DIABETIC PATIENTS IN KETO-ACIDOSIS
8	13	HYPERTENSIVE PATIENTS
9	594	PATIENTS IN ASCENDING ORDER BY AGE
10	594	PATIENTS WITH HIGHEST CREATININES FIRST
11	539	PATIENTS NOT ON GENTAMICIN
12	55	PATIENTS ON GENTA ORDERED FROM HIGHEST TO LOWEST BUN
13	54	DISCHARGED DEAD
14	215	PATIENTS NOT ON DIURETICS
15	424	NOT ON DIG
16	2	ON DIG AND GENTAMYCIN
17	389	NOT ON DIG OR GENT

Table 3.8. List of Variables in Example 1

Index	Number in Group	Message
1	594	PATIENTS WHOSE 1ST BUN WAS > 20, MCV AT TIME OF LOWEST HCT
2	594	PATIENTS WHOSE 1st BUN WAS > 20, MIN HCT
3	594	PATIENTS WHOSE 1ST BUN WAS > 20, UNITS OF BLOOD RECEIVED
4	594	PATIENTS WHOSE 1ST BUN WAS > 20, UNITS OF WHOLE BLOOD RECEIVED
5	594	PATIENTS WHOSE 1ST BUN WAS > 20, UNITS OF PACKED RBC
6	594	PATIENTS WHOSE 1ST BUN WAS > 20, UNITS WHOLE BLOOD + RBC
7	594	PATIENTS WHOSE 1ST BUN WAS > 20, PATIENG AGE
8	594	PATIENTS WHOSE 1ST BUN WAS > 20, MIN HCT
9	594	PATIENTS WHOSE 1ST BUN WAS > 20, MIN BUN
10	594	PATIENTS WHOSE 1ST BUN WAS > 20, MIN POTASSIUM
11	594	PATIENTS WHOSE 1ST BUN WAS > 20, MAX CREATININE
12	594	PATIENTS WHOSE 1ST BUN WAS > 20, MAX BUN
13	594	PATIENTS WHOSE 1ST BUN WAS > 20, GLUCOSE AT TIME OF MAX BUN
14	594	PATIENTS WHOSE 1ST BUN WAS > 20, URIC ACID NEAREST 1ST BUN
15	594	PATIENTS WHOSE 1ST BUN WAS > 20, MAX POTASSIUM
16	594	PATIENTS IN ASCENDING ORDER BY AGE
17	325	MAX BUN FOR MALES
18	325	MIN BUN FOR MALES
19	242	MAX BUN FOR FEMALES
20	242	MIN BUN FOR FEMALES
21	55	MAX BUN FOR PATIENTS ON GENTAMICIN
22	539	MAX BUN FOR PATIENTS NOT ON GENTAMICIN
23	215	MAX BUN FOR THOSE NOT ON DIURETICS
24	379	MAX BUN FOR PATIENTS ON DIURETICS

Table 3.9 Listing of Audit Trail for One of Participants, Example 1

```
9:10  8/3/77   FIRST SESSION WITH DR. . . .
CONTROL PATIENTS WHOSE 1ST BUN WAS  > 20, MIN HCT
TEST PATIENTS WHOSE  1ST BUN WAS  > 20, MIN HCT

CONTROL PATIENTS WHOSE 1ST BUN WAS  > 20, MIN HCT
TEST PATIENTS WHOSE 1ST BUN WAS  > 20, MIN HCT

CONTROL MAX BUN FOR MALES
TEST  MAX BUN FOR FEMALES

CONTROL MAX BUN FOR MALES
TEST MAX BUN FOR FEMALES
KULMOGOROV-SMIRNOV P = 0.12109
MANN-WHITNEY TWO TAIL:  0.44435

CONTROL MIN BUN FOR MALES
TEST MIN BUN FOR FEMALES

CONTROL MAX BUN FOR PATIENTS NOT ON GENTAMICIN
TEST MAX BUN FOR PATIENTS ON GENTAMICIN
KULMOGOROV-SMIRNOV P = 0.00000
MANN-WHITNEY TWO TAIL:  0.00000

CONTROL PATIENTS IN ASCENDING ORDER BY AGE
TEST PATIENTS IN ASCENDING ORDER BY AGE

CONTROL MAX BUN FOR THOSE NOT ON DIURETICS
TEST MAX BUN FOR PATIENTS ON DIURETICS
KOLMOGOROV-SMIRNOV P = 0.06642
MANN-WHITNEY TWO TAIL:  0.29747

CONTROL PATIENTS WHOSE 1ST BUN WAS  > 20, MAX BUN
TEST PATIENTS WHOSE 1ST BUN WAS  > 20, MAX CREATININE

X AXIS PATIENTS WHOSE 1ST BUN WAS  > 20, MAX BUN
Y AXIS PATIENTS WHOSE 1ST BUN WAS  > 20, MAX CREATININE
CORRELATION COEFFICIENT R = .26
CORRELATION COEFFICIENT R = .52
```

was comparing the data in each variable to see if they were the same data. This was examined twice. This duplication occurred due to an oversight during the data base generation. It was not really a significant error.

It is easy to follow the path of the investigation from the audit trail alone. Briefly it included the following:

1. He checked the comparison of the maximum BUN values for the males and females. (First he checked the range of values and the averages; then he performed two statistical tests of their significance. This could have been done in one step but he chose not to.)

2. He performed a similar comparison for the minimum BUN values for each group.

3. He compared the maximum BUN for patients receiving and not receiving gentamicin and found that they had significantly different maximum BUN values.

4. He then asked for a distribution of the ages of the 594 patients.

5. Next, he compared the maximum BUN for patients receiving and not receiving diuretics. There was no significant difference in the maximum BUN for patients in either group.

6. He correlated the maximum BUN with the maximum creatinine but first made an error by specifying that a comparison was to be made rather than a regression. The regression showed no correlation (notice $R = .26$, then it improved only slightly, $R = .52$) after some outliers were removed.

This illustrates that from the data in the populations,

variables, and audit trail, it is possible to reconstruct the sequence of steps in the investigations and the outcomes of each step. After completing all the demonstrations, the audit trail algorithms were modified to reflect additional information resulting from the comparisons and regressions. For example, the audit trail now reflects the slope equation (i.e., y = mx + b) by supplying the values for each variable. These are supplied by the system as well as such things as the correlation coefficient, etc.

The example in Table 3.10 serves to illustrate the data now collected automatically in the audit trail file during each session. A comparison of this data with that in Table 3.9 illustrates that now an even more accurate picture can be reconstructed not only of the events in the session but the results of each step. This may be used by the investigator as an aid in seeing relationships among selected parameters and provide insight for strategies of analysis in future investigations.

Table 3.10. An Example of Data Now Collected Automatically in Audit Trail File

```
X AXIS MAX BUN FOR PATIENTS NOT ON GENTAMICIN
Y AXIS MAX BUN FOR PATIENTS ON GENTAMICIN
Y = .14        X = 44.48
CORRELATION COEFFICIENT R = .07
STANDARD DEVIATION FROM REGRESSION S(Y,X) - 32.79
STANDARD ERROR OF REGRESSION COEFFICIENT S(B) = .25
SIGNIFICANCE OF CORRELATION COEFFICIENT P = .59

CONTROL MAX BUN FOR PATIENTS NOT ON GENTAMICIN
TEST MAX BUN FOR PATIENTS ON GENTAMICIN
CONTROL MEAN = 34.02
TEST MEAN = 51.83
CONTROL S.D.   21.14
TEST S.D. 31.21
```

Responses from the Participants

The comments from the users are classified as relating to the features of the system or the potential uses of the system. This section paraphrases most of the user comments in an attempt to give the essence of the more noteworthy spontaneous responses.

Comments Relating to the Features of the System. One participant noted that it would take literally weeks of manual data gathering and manipulation to perform the analysis procedures we were conducting routinely with this system. Another participant commented on the flexibility of the system as he had used it for the anemic data base and the elevated BUN group. He noted how it did not seem to be restricted to a limited number of subjects.

After using the system for two teaching sessions, one resident requested that a terminal connection be made available in his clinic area at a remote site. He had previously noted that the system could be used in an independent study mode. Four other participants requested an opportunity to use the system in this mode which may be feasible (especially with the audit trail capabilities), but this aspect was not investigated.

Nearly all participants commented on the advantage and ease of first displaying a histogram or a correlation plot, and then being able to easily rank the patients and begin examining the details of individual patient records. In this way additional insight was gained for other factors which might be influencing the particular parameter under investigation.

One resident wanted to eliminate patients who had been admitted due to injuries in an accident. His concern was to eliminate the effects of trauma in the current investigation. Unfortunately it would have required the completion of another search to generate such a list. This general problem is presented in the next section.

One medical student suggested that the real advantage of the system was in the ease of generation of new populations and variables from the existing ones. This feature permitted rapid investigation for validating or eliminating a hypothesis. Another resident noted that this system permitted teaching by exclusions and inference. He had read an article (Platt, 1964), suggesting that strong inference and the scientific method be used more liberally in the sciences. His particular interest was that this process be used more in medical decision-making and that this system could be a significant aid to direct clinicians in that direction.

One faculty member was very skeptical about the system when first seeing it. He made negative comments concerning the statistical components in the system. His concern was that students with limited statistical background would not be able to use the system. While it is true some basic knowledge of statistics is required it has not proved to be a problem during the demonstrations. This problem was discussed earlier in this chapter. The same person later, after viewing more demonstrations, stated that by using this system he could see some relationships he had never thought about before. His final remarks after the second session were very positive.

Comments Relating to Potential Uses of the System. One medical student noted that the same physician name kept appearing with patients who were seriously anemic yet appropriate lab tests and thorough work-ups were apparently not being done. After seeing several such cases, the student suggested that the Professional Standards Review Organization (PRSO) might be interested. This has been the subject of the research work of Louis Ricord of the University of Utah Department of Medical Biophysics and Computing (MBC).

The same student suggested that this system also be used by researchers to study diseases that are not presently well understood. A resident said that the study he was pursuing with this system was publishable and suggested that a comparison be made with the data he was using against schemes published in various hematological journals. He suggested that this system might be used to validate or to take issue with the published data, thus contributing to medical knowledge and eventually medical decision-making.

One student suggested that this system be used in a laboratory setting for a decision-making theory class aimed at medical students.

Requested Improvements in the System

Almost without exception each user requested that the system display variable values in a more readable format. Currently the system permits only integer values for variables. For example a hematocrit might be 41.5 but the system displays such a hematocrit as 415. The user must be aware of the normal units of the variable being studied and then mentally set the decimal point appropriately.

One user suggested that a new subpopulation be generated. The problem was to generate a subpopulation using a parent population and another subpopulation. The following Venn diagrams will help explain the problem (see Fig. 3.3). The entire population is

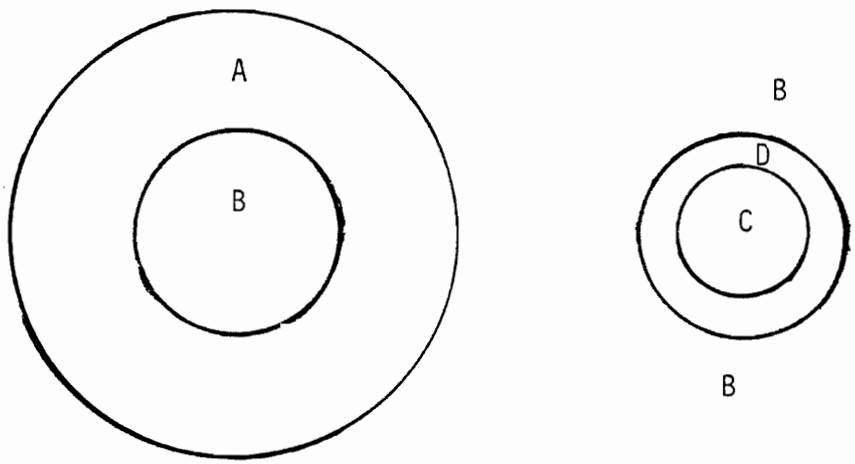

Fig. 3.3. Venn diagrams using a parent population and another subpopulation.

represented by the circle A. The females are represented by circle B. The females (circle B, at the right) are further divided into the females admitted for labor and delivery. The other females (portion of circle B labeled D) are referred to as the non-OB females. These are the complement of C.

The user wanted to create a subpopulation of non-OB females. There already existed subpopulations for females and the OB females. The complement feature was not available but was added immediately. Until it was added the only generation mechanism was the search of the disc using HELP logic.

One resident suggested studying the effects of administering gentamicin for varied lengths of time, on selected parameters. This suggestion could not be serviced quickly because of the time required to generate new variables and populations. The user was suggesting that a faster means be provided to honor such requests.

Two users suggested that a document be created either interactively or in a hardcopy form that would provide examples of the features in the system and how to use it. One suggested a programmed text type of document so that users might access the system more freely in an independent study mode.

Observations Selected Participants Pursued During Investigations

This section presents a general analysis of the teaching sessions as well as details of the investigations of selected participants.

Analysis. The system overview consisted of introducing the participant to the concept of populations and variables, statistical capabilities, and the empirical approach to medical decision-making. This overview was conducted with each participant. Twelve minutes were required to complete the overview on the average for nine participants. One session required over 70 minutes because of more intense questioning by one participant.

Two important measures of acceptability appear to be (1) the number of participants requesting an opportunity for multiple sessions; and (2) the number requesting an opportunity for independent study.

Figure 3.4 presents the number of sessions conducted for each participant. The number in each box indicates the number of teaching sessions the participant engaged in. These teaching sessions averaged nearly 95 minutes each. The longest session was slightly more than three hours and the shortest was 55 minutes. What was especially interesting was that three of the residents who were especially reticent to participate initially not only stayed more than an hour in the first session, but all three requested additional sessions. Two of these continued to request opportunities to participate in future sessions (beyond this testing and demonstration stage) to pursue investigations on existing and additional data bases. Had time permitted, these requests might have been satisfied.

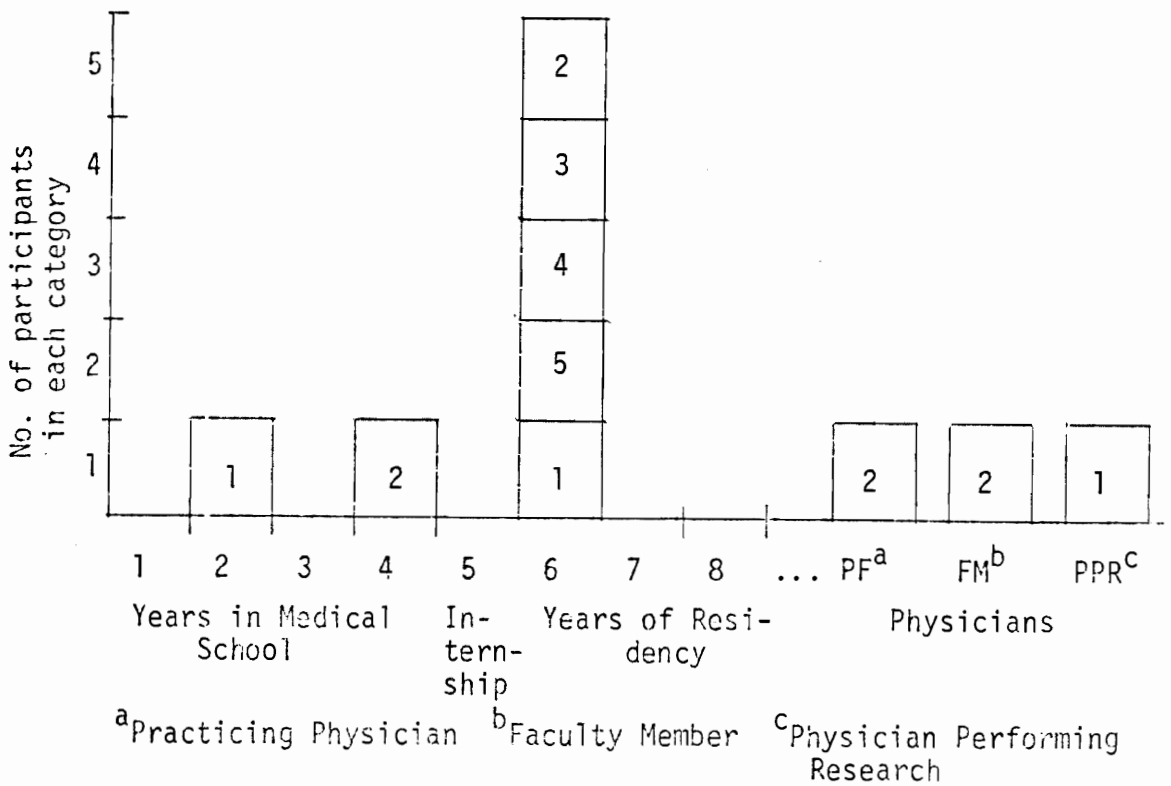

Fig. 3.4. Number of sessions conducted for each participant.

The other measure referred to was the request for an opportunity to use the system to pursue independent study. Four participants requested an opportunity to do so. Two of these actually performed a portion of their investigations in this mode. This assumes that the users can manipulate the system according to their individual preferences. To do so required that the users watched closely and then practiced with the aid of an experienced user. One participant who did not actually use the system in the independent mode but suggested it as an option, noted that he thought the system directed the user sufficiently that the user could acquire the required competencies in 15 minutes. Such a statement is probably excessively optimistic (one hour may be more realistic) but it is encouraging to see that nearly half of the users felt the system was sufficiently easy to use to operate in this mode.

Some of the participants are obviously more prepared to pursue meaningful independent investigations than others. For example, it is very clear that the level of participants' knowledge of the subjects determines the sophistication of the objectives pursued during an investigation. To illustrate, part of a medical student's audit trail appears in Table 3.11. Careful study of this partial audit trail and the tape reveal that the participant was in search of what may be called "low level" data. This analysis is not intended to demean the pursuit of the investigation. This example shows that the "low level" data for decision-making may be derived by asking, "What is the average minimum hematocrit and MCV values for OB patients?" This type of data, though information, is not as clinically significant as other

Table 3.11. Part of a Medical Student's Audit Trail

```
20:15    7/27/77     SESSION WITH . . .

CONTROL MIN HCT FOR THE OB PATIENTS
TEST MIN HCT FOR THE OB PATIENTS

CONTROL MIN MCV FOR OBS
TEST MIN MCV FOR OBS

CONTROL MIN MCV FOR OBS
TEST MIN MVC FOR OBS

CONTROL MAX BILI FOR FEMALES
TEST MAX BILI FOR FEMALES

X AXIS MIN MCV FOR MACROCYTIC PTS
Y AXIS HCT OF MACROCYTIC PTS
CORRELATION COEFFICIENT R = .10

CONTROL MIN MCV FOR MACROCYTIC PTS
TEST MIN MCV FOR MACROCYTIC PTS

CONTROL MIN HCT FOR PTS WITH WBC > 5000
TEST MIN HCT IN PTS WITH WBC  < 5000
KOLMOGOROV-SMIRNOV P = 0.00000
MANN-WHITNEY TWO TAIL:  0.00000
```

data equally available through the same system. This is why emphasis has been placed on guidance from an expert in the subject under investigation.

In contrast to the above example, consider the data in Table 3.12, which illustrates a portion of another audit trail from a more experienced clinician. The question sequence was: (1) Do patients on gentamicin have a significantly different maximum BUN value than those not taking the drug? (2) Was the maximum BUN affected by the administration of diuretics? (3) Did digitalis affect the renal funtional indicator--creatinine levels, etc." These questions led to a more meaningful and clinically relevant result as shall be pointed out later.

These examples support the premise that the system permits the experienced and inexperienced investigator to pursue data on questions at their current level and advance to increased levels of complexity as needed. The demand for additional data seems to increase proportionately to the level of the investigator's sophistication. For example, the hematology resident asked some very sophisticated questions to be answered from the data base. These questions could not be answered until 13 new variables had been extracted from the patient data, using HELP logic.

Sessions with Selected Participants. One participant investigated the relationships of serum iron values and iron binding capacity in anemic patients. The hematology texts indicate that a serum iron less than 50 mg/dl is considered low, while an iron binding capacity

Table 3.12. Audit Trail from a More Advanced Participant

9128 8/4/77

CONTROL MAX BUN FOR PATIENTS NOT ON GENTAMICIN
TEST MAX BUN FOR PATIENTS ON GENTAMICIN
KOLMOGOROV-SMIRNOV P = 0.00000

CONTROL MAX BUN FOR THOSE NOT ON DIURETICS
TEST MAX BUN FOR PATIENTS ON DIURETICS
T TEST P = 0.02747
KOLMOGOROV-SMIRNOV P - 0.06642
MANN-WHITNEY TWO TAIL: 0.29747

CONTROL MAX CR. IN PTS NOT ON DIG.
TEST MAX CREAT. IN PATIENTS ON DIG
KOLMOGOROV-SMIRNOV P = 0.00000
MANN-WHITNEY TWO TAIL: 0.00000

NOTE: The comparison of BUN levels for patients on/not on gentamicin indicates that the BUN levels were significantly different between the two populations.

The comparison of BUN levels for patients receiving/not receiving diuretics shows no significant difference.

The creatinine levels of patients on/not on digitalis were significantly different by both tests.

(IBC) greater than 410 mg/dl is above the normal threshold. By using the interactive population and variable generation techniques, two populations were created: (1) patients with a serum iron < 50 mg/dl and (2) patients with an IBC > 410 mg/dl. By using HELP logic, each of these was further divided into two groups, as shown in the 2 x 2 contingency table (see Figure 3.5). A study of this contingency table reveals that a physician ordering one test could predict the outcome of the other variable based on the single test ordered. Notice that if only the IBC were ordered and within normal limits, one could predict with a 40-to-19 ratio, or about 2-to-1, that the serum iron would be normal also. If these tests are reversed in order and the serum iron is normal, one may predict with a 19-to-1 certainty that the IBC is also normal. This analysis was especially exciting to one resident who felt a commitment to decision analysis processes for medical decision-making.

Much of the time in these sessions was spent in arriving at information which had potential input into decision-making algorithms.

	Low fe	Normal fe
Elevated IBC	9	1
Normal IBC	40	19

Fig. 3.5. 2 x 2 contingency table

One item of information particularly new and interesting to all but the hematology experts is presented below:

A study of those patients receiving the most units of blood revealed that for the most part they were leukemic patients who commonly received up to eight units of platelets in a short period of time. For example one patient received 107 units of blood; the next highest was 99 units of blood, of which most were platelets. There were two OB patients among those receiving extraordinarily high units of blood; they received 70 and 66 units respectively. The open-heart surgery patients were next in order of total units received. At first, these high figures for units of blood transfused appeared to be in error but the atcual patient records confirmed them to be accurate.

One of the residents suggested a comparison of the maximum bilirubin for the males and females. The comparison showed that the males exhibited an elevated maximum bilirubin over the females. The participant said that there does not appear to be any reason to believe that males have any higher incidence of liver disease than do the females, other than the fact that the male population may have included a larger number of alcoholics than appeared in the female group. Creatinine for the males and females was compared next and the males had a significantly elevated creatinine above that of the females. The females in this study, of course, included OB patients admitted with a hematocrit less than 34. This led to the idea of eliminating the OB patients so that we might compare only the males and the non-OB females for their maximum bilirubin and creatinine values. Even with the OB females removed, the males still exhibited significantly elevated bilirubin

and creatinine values. This led to the question, "Are the patients with the elevated bilirubin values in the males the same patients who have the elevated creatinine levels? This test was performed and showed that the patients with the high creatinines and the high bilirubins were not necessarily the same patients. It was then suggested that the sole criteria for being included in the study was that the patient's admitting hematocrit was less than or equal to 34 and that it was a far more significant drop in the hematocrit for males to be at the level of 34 than it was for the females. Therefore, the males must have been sicker to even be included in the study.

To test this hypothesis, the participants proposed the following: Since the average hematocrit for males is 45 and the basis for being admitted to the study was having a hematocrit of 34, this is a drop of eleven points or approximately 25% of their normal hematocrit. Therefore, assuming a female's normal hematocrit is something around 40, let's drop their hematocrit level by 25% also. This equals ten points from the average female's hematocrit. Consequently, another subset of the non-OB females was created consisting of those females who had a hematocrit less than or equal to 30. The creatinine and bilirubin values were generated for this new subset of females and compared against that of the males. Again, the males had significantly higher maximum bilirubin and creatinine values than did the females. There was considerable discussion subsequent to this examination pertaining to what could be responsible for these observations.

The answers eventually arrived at were (1) creatinine is elevated in the males because it reflects the muscle mass of the body,

and (2) the elevated bilirubin in males indicates that the males in this study had a higher incidence of liver disease than did the females. Many complex and divergent pathways were investigated before arriving at these rather simple conclusions. This investigation was especially interesting because it caused the participants to consider interrelationships and data that needed investigation before arriving at their decision after a series of exclusions.

One of the most productive investigations dealt with the effect of gentamicin on renal function. The maximum BUN level for patients receiving gentamicin was significantly different from that of patients not receiving the medication. Diuretics did not affect the BUN levels of the patients. There were 55 patients who received gentamicin. Figures 3.6 and 3.7 illustrate two HELP sectors used for extracting creatinine values from the patient data.

The HELP sector shown in Figure 3.6 is designed to extract from the patient data the maximum creatinine value for a patient before receiving gentamicin. The maximum creatinine value retrieved will be stored as a variable associated with each patient in a population. The first line of the sector is the message or name of the variable being created. The final evaluation (FVAL) is the maximum serum creatinine value for the patient, extracted using all of the statements in the HELP sector through and including statement C. The next line A indicates that the patient record is to be searched for the first dosage of gentamicin. Gentamicin is an antiinfective drug among the group of drugs known as aminoglycosides. Statement B instructs the computer to abort further searching if the patient did not receive gentamicin. Statement C is

```
Maximum serum creatinine before gentamicin therapy
FVAL    C
A   Anti-infectives, (A) aminoglycosides, (B) gentamicin===
      MGM 1st
B   Exist if (not A) exit
C   (A) SMA-12, Creatinine (X10) MG%, (B) SMA-12, specimen
      type Boolean (A NE Q) and (B EQ 18) maximum from 4
      months before event A to 1 minute before event A
```

Fig. 3.6. A HELP sector for extracting maximum serum creatinine values before gentamicin therapy.

```
Maximum serum creatinine after gentamicin therapy
FVAL    B
A   Anti-infectives, (A) aminoglycosides, (B) gentamicin ===
      MGM, (C) DC date to get C last
B   (A) SMA-12, creatinine (X10) MG%, (B) SMA-12, specimen
      type Boolean (A NE Q) and ( EQ 18) maximum from 1
      minuted after event A
```

Fig. 3.7. A HELP sector for extracting maximum serum creatinine values after gentamicin therapy is terminated.

more complex and consists of five parts. Each part is presented in the following discussion:

At (A) in statement C the value for creatinine is specified as one compoment from the SMA-12 lab test. At (B) the specimen type is requested. Parts (A) and (B) are followed by Boolean logic. For example, the creatinine value (from Part A) cannot be zero, meaning we must have a positive value for creatinine. Similarly, the specimen type (from Part B) must be a blood sample (a specimen type of 18 means it is a blood sample). The fourth portion indicates the maximum creatining value is to be retrieved. The fifth portion indicates the time

limit restrictions. It specifies that the maximum creatinine value is to be sought from the patient data in the period from four months before receiving the first dosage of gentamicin (event at statement A) to one minute before receiving the drug. When the computer has finished evaluating line C, it will have retrieved the maximum serum creatinine value for the patient before receiving gentamicin. This value is referenced in the FVAL statement.

Figure 3.7 illustrates a second HELP sector designed to extract the maximum serum creatinine value after gentamicin therapy has terminated. Since there is great similarity in the logoc for the two HELP sectors in these figures, an explanation of only the distinguishing characteristics of the sector shown in Figure 3.7 follows.

Statement A directs the computer to retrieve the discontinuation date (DC) of the last gentamicin therapy. Statement B directs the computer to extract the maximum creatine value from the patient record beginning at one minute following the termination of gentamicin therapy.

The creatinine levels were measured before and after receiving gentamicin, and found to be significantly different from each other. This was not surprising because it generally is clinically believed that gentamicin has a potentially derogatory effect on renal function. What is surprising was that when the trends of creatinine level before and after receiving gentamicin were compared, the patients actually exhibited the highest creatinine levels before receiving the medication. This was considered a very significant discovery, worthy of note for the clinicians' decision-making.

Competencies Required of Instructor and Student

The comments in this section are for the most part based on analysis of the tapes, though a few of the observations were suggested by the participants. The competencies for the instructor are presented first. The order of presentation of each item has no special significance.

The instructor must be an expert in the subject being studied. This seems obvious but experience showed that the more knowledgeable the expert the more meaningful the session. This does not preclude the possibility for independent study.

Ideally, the teacher should understand the theoretical basis for the statistical tests available. For instance, he should know that parametric tests assume a normal distribution (i.e., t test is for small normally distributed populations), and that the Kolmogorov-Smirnov test for significance makes no assumption about the distribution. The faculty member must be committed to the notion of arriving at a diagnostic decision using more formal and quantitative techniques. An ability to ask clinically-relevant questions is perhaps one of the most important qualities of an instructor using this system. The one leading the investigation must ask questions to direct the students that they may "discover" for themselves the answers. There is almost an overwhelming urge to immediately give the answers and an effort must be exerted to refrain from doing so. With this important concept it reinforces the first competency noted above. The more the instructor knows of the interrelationships of parameters in the subject matter, the better

the questions he can ask. Asking the right questions assumes the instructor has the session mapped out so that it focuses on the objectives to be taught.

The students gain the most when they have a thorough knowledge of physiological systems and an understanding of the interrelationships of each. In particular the students should have reviewed the physiology pertaining to the subject being investigated before using the system. The students should have some clinical experience with patients relating to the subject being investigated. In general, the value of the system seems to be proportional to the extent of clinical experience in the subject area. This is not to say that the inexperienced participant is eliminated from using the system, but that the depth of learning is more significant with more clinical experience. Unlike the frame-oriented systems that provide remedial feedback or the quick "correct answer," this system does not "spoon feed" answers to the participants.

The system's structure encourages the student to stretch his reasoning and creative capacities, by providing clinical data and many flexible analytical capabilities. The type and implications of the analysis are as far-reaching as the student is able to conceive and the data base permits. Therefore some degree of creativity and extensive reasoning capabilities are required to realize the potential for which the system was designed. This characteristic cannot be overemphasized.

When students work in groups, the pooling of ideas through discussion may create a more productive learning environment. This was noted especially when three residents worked together in one session.

CHAPTER IV

Significant Features and Contributions of This Nonauthored Medical CAI System

Most CAI systems, including the nonauthored systems, are only used once by a student for one subject area (e.g., learning a programming language syntax). If a student does return for additional sessions it is usually for review purposes only. For example, using conventional frame-oriented CAI systems, a student will rarely use the system more than twice for the same course, because of the boredom involved in seeing the identical material repeated. In order for a student to view additional related material an "author" must dedicate many hours to create new courseware for the student. Even with the nonauthored systems previously presented, the material is subject-specific so that no options exist to move to other subjects of interest in the same discipline. The SCHOLAR system for example could not readily present information on European countries without additional data collection and enlargement of the semantic network.

The demonstrations of this system have shown that it is not subject-specific nor are extensive hours of authoring required when a new subject area is requested. However, a significant amount of computer time is consumed in generating the data base for a new subject area. Though other subjects could have been studied, time constraints would not permit the investigation of all the additional

subject areas suggested by the participants.

When a new subject area for study is selected, the time required to generate the data base for it is usually less than the number of hours spent in using the system for investigations. Most CAI systems report author/lesson ratios on the order of 50:1, 150:1, 400:1 (i.e., 150 hours of authoring time to provide enough material to keep the student active at the terminal for one hour) while this system actually showed a reversal of that trend. For example, it may require five to eight hours to (1) determine which populations and variables are required, and (2) extract the data base. Only a small part of the generation time involves the faculty members; most involves only a computer technician. Using the data base the students could spend two to three times the data base set-up time pursuing their investigations. This is dependent on the depth of the investigations.

The existence of this system itself is significant because of its generative characteristics. It contributes by filling part of the void in nonauthored systems and responds to the call by CAI experts, who have requested that more emphasis be placed on the development of nonauthored systems. At the same time medical CAI experts are suggesting added emphasis be placed on clinically-relevant CAI. This nonauthored medical CAI system was conceived in response to both of these demands. It is clinically relevant as evidenced by the response of the participants in the demonstrations.

The system permits interactive creation of lists of populations and variables as well as flexible statistical analysis. Though these and other attributes make it a powerful tool for analysis, its purpose

is not primarily to teach that data are distributed in some particular way or that two variables are correlated with a certain correlation coefficient. To analyze the data in such a manner is like memorizing facts and figures for decision-making. Its more comprehensive purpose is to provide a tool that encourages the student to practice and improve his skills in inductive reasoning, which extends into other kinds of data. Using these skills the student can study the relationships of selected laboratory tests with diseases and explore the relative frequencies of disease manifestations. Further the student can use these skills to explore the data and to develop diagnostic and prognostic decision criteria leading to hypotheses which can be treated against the data base.

Because this nonauthored medical CAI system uses HELP logic as an integral part, it provides an opportunity for a trained participant to formalize the decision-making logic so that it may actually be tested against currently hospitalized patients. Unfortunately, none of the participants was able to attain the level of sophistication needed to create the logic into a formal HELP sector. Had time been allocated to studying the HELP system, the participants doubtless could have accomplished some formal logic definitions. The problem was simply insufficient training in using the HELP system itself.

The demonstration revealed that clinicians not only will use the system once but will request opportunities to do so repeatedly. It also showed that it is possible and desirable in some cases to use the system for independent investigations. The fact that the system is self-documenting via the population/variable and audit trail

printouts make this an even more viable alternative.

This nonauthored medical CAI system runs under the control of the STRATO system. The University of Utah Department of Medical Biophysics and Computing has demonstrated a commitment to the maintenance of STRATO. This virtually assures the availability of this CAI system. Part of the design philosophy was to integrate the system with an ongoing effort so that it might continue to be available to interested clinicians. Further details on this subject and possible extensions to the system appear in the next section.

Possible Future Enhancements Worthy of Consideration

The participants' comments outlined in Chapter III should perhaps receive priority for implementation. These would include a conversion algorithm for making the variable values more readable, and a programmed introduction to the generative medical CAI system.

At times it is enlightening to see an actual patient's file from the medical records department. The system is designed to blank out all patient names, while reviewing patient data, to maintain patient privacy. Therefore, currently it is impossible for the user to identify any patient by name. Consequently, any particular patient file cannot be retrieved from medical records. It would therefore be convenient to provide an additional option whereby the user can specify a patient number which would be displayed automatically in the medical records deparment. Persons in medical records could then respond by making the patient record available for retrival by a responsible person.

Some advanced CAI systems provide a touch panel display permitting the user to touch with a finger or light pen, objects, words, or other material being displayed on the screen. With such an option available to users of the nonauthored medical system, they could view a scattergram and circle or touch a patient or sets of patients on the scattergram. The patient(s) so designated could then be set aside into populations for more extensive examination. The capability to select patients from the scattergram is possible with the current system, but the procedure to accomplish it is not nearly as simple as that proposed here.

A most exciting extension of this system would be to routinely transfer all patient data to random access devices (800 megabyte discs, magnetic domains, etc.) adding it to the archival data base already there. Inverted files could be created for this large data base to improve access times. Using this expanded data base a physician could create a data base consisting of X patients with characteristics similar to those of a specific patient now hospitalized. An automatic interface could be created to select the set of patients to be included in the population, directly from the specified patient's history, sex, disease state, laboratory values, medications, and other critical parameters. This would permit the clinician to simply specify the current patient and a few other data items to initiate the generation of a set of patients "like" the current one.

This would permit decision-making using actual clinical data on perhaps hundreds of similar patients to the current patient. The

cost of maintaining such a data base would be significant, but it would provide an analytical capability for decision-making with significant potential. Such a system that is available to the clinician and is simple to operate should provide a wealth of relevant data ultimately leading to better decision-making and better care.

If such a system were operational it would be used because it meets the immediate needs of the clinician. One usually studies that which will pay off in treatment of current patient problems. This was especially true of the demonstrations in this study. The most enthusiastic participants designed their investigations to be relevant to their current needs. These ideas all require a significant commitment of funds and time for fruition. A more realistic extension with possibilities for achievement appears below.

Participants and an observer (a faculty member in the College of Medicine at the University of Utah) suggested that a course be organized to teach formal systems for medical decision-making. Furthermore they suggested that such a course offer practical applications via laboratory sessions using this nonauthored medical CAI system. This is only being discussed presently but it could obviously be an immediate and realistic extension of the system.

REFERENCES

Abrahamson, S., Denson, J. S., and Wolf, R. M. Effectiveness of a simulator in training anesthesiology residents. Journal of Medical Education, 44 (June 1969):515-519.

Barr, A., Beard, M., and Atkinson, R. C. A rationale and description of a CAI program to teach the BASIC programming language, Instructional Science, 4 (Spring 1975):1-31.

Brigham, C., Kamp, M. C., and Cross, R. Guide to computer assisted instruction in the health sciences. Springfield, Virginia: National Technical Information Service, 1972.

Carbonell, J. R., AI in CAI: An artificial intelligence approach to computer-assisted instruction, IEEE Trans. on Man-Machine Systems, Vol. MMS-11, 4 (December 1970):190-202.

Cochran, T., and Stolurow, K. A. C. A bibliography of medical CAI systems. Received via a personal communication from the authors both of whom are at SUNY, Stoneybrook, New York.

Cornfield, J., et al. Multigroup diagnosis of electrocardiograms. Computers and Biomedical Research, 6 (February 1973):97-120.

Geyman, J. P. Toward the definition of family practice--A quantum jump. The Journal of Family Practice, 3 (January 1976):22-68.

Gorry, G. A., and Barnett, G. O. Experience with a model of sequential diagnosis. Computers and Biomedical Research, 1 (May 1968):490-507.

Hickey, Albert E. Research guidelines for computer-assisted instruction. Newburyport, Massachusetts: By the author, 42 Pleasant Street, 1974.

Kamp, M. C. Index to computerized teaching in the health sciences. Los Angeles: Association for Development of Computer-Based Instructional Systems, 1975.

Koffman, Elliot B. Design techniques for generative computer-assisted instructional systems. IEEE Transactions on Education, Vol. E-16 (November 1973):182-189.

Koffman, E. B., and Blount, S. E. Artificial intelligence and automatic programming in CAI. Artificial Intelligence, 6 (Fall 1975):215-234.

Ledley, R. S., and Lusted, L. B. Reasoning foundation of medical diagnosis: Symbolic logic, probability, and value theory aid our understanding of how physicians reason. Science, 130 (July 1959):9-21.

McMullen, D. W. Generative CAI: Procedures and prospects. Educational Technology, 14 (February 1974):27-30.

Molnar, Andrew R. The use of computers in education. Technological Horizons in Education Journal, 3 (February 1976):18-23.

Moore, D. H., II. Combining linear and quadratic discriminants. Computers and Biomedical Research, 6 (October 1973):422-429.

Peele, H. A., and Riseman, E. M. "Four faces of HAL: A framework for using artificial intelligence techniques in computer-assisted instruction. IEEE Transactions on Systems, Man, and Cybernetics, SMC-5 (May 1975):375-380.

Platt, J. R. Strong inference. Science, 146 (No. 3642) (16 October 1964):347-352.

Pryor, T. A., et al. HELP: A computer system for medical decision-making. Computer Magazine, 8 (January 1975):34-38.

Reed, F. C., et al. A model for the development of computer assisted instruction programs. Educational Technology, 14 (March 1974): 12-20.

Roid, G. H. Selecting CAI author languages to solve instructional problems. Educational Technology, 14 (May 1974):29-31.

Rubin, M. L. Computer-assisted instruction mixed ratings for the network? Biomedical Communications, 4 (July 1976):10-12.

Rubin, M. L., Hunter, B., and Kretsch, M. Evaluation of the experimental CAI network (1973-1975) of the Lister Hill National Center for Biomedical Communications. Birmingham, Alabama: National Library of Medicine, Human Resources Research Organization (HumRRO), 1975.

Uttal, William R., Rogers, Mirriam, Hieronymous, Ramelle, and Pasich, Timothy. Generative computer-assisted instruction in analytic geometry. Newburyport, Massachusetts: Entelek, Inc., 1970.

Warner, H. R., Olmsted, C. M., and Ruthorford, B. D. HELP: A program for medical decision-making. Computers and Biomedical Research, 5 (February 1972):65-74.

Warner, H. R., Woolley, F. R., and Kane, R. L. Computer-assisted instruction for teaching clinical decision-making. Computers and Biomedical Research, 7 (December 1974):564-574.

Warner, H. R., et al. A mathematical approach to medical diagnosis application to congenital heart disease. Journal of the American Medical Association, 177 (July 1961):177-183.

Wenzel, B. Instructional innovations in clinical applications of computers and computer-assisted instruction. An unpublished summary received via personal communication, 1976.

VITA

Name	Richard Stephen Dick
Birthdate	April 21, 1945
Education	Brigham Young University (Provo, Utah) 1963-1964, 1966-1968. B.S. Zoology/Computer Science
	Case Western Reserve University (Cleveland, Ohio) 1970-1971. M.L.S. Library Science/Operations Research
	University of Utah (Salt Lake City, Utah) 1975-1978
Professional Positions	Boeing Computer Services, McLean, Virginia, Senior Systems Analyst. September 1977 to present.
	Eccles Health Sciences Library, University of Utah. Research faculty appointment for the development of medical computer assisted instruction (CAI) June 1974 to September 1975.
	Sperry UNIVAC Corporation, Salt Lake City, Communications and Terminals Division. Systems programmer. June 1971 to June 1974.
	Case Western Reserve University Center for Documentation and Communications Research. Programmer/ Analyst. September 1970 to May 1971.
	United States Army, Fort Ord, California. Programmer rank E4. Full-time work at Monterey, California. October 1968 to June 1970
	Eccles Health Sciences Library consultation, investigation, and evaluations.
Scholarship	NLM-PHS scholarship to study library science at Case Western Reserve University
Fellowship	University of Utah College of Medicine, study of Medical Biophysics and Computing.
Copyright	"A Computerized System for the Medical Library Exchange" designed and implemented on a UNIVAC 1108 in 1968.

Computer Languages Used	ALGOL COBOL FORTRAN BAL DML DMS GPSS BASIC
Computer System/ Hardware	UNIVAC 1108 and 1110 EXEC 8 IBM 360/67 and 50 OS/MVT IBM 74040 Eclipse/CDC 3300 UNIVAC 9000 series DOS/OS
Publication	Communicator, official organ of the University of Utah, briefly described a system I developed while at Eccles--an interactive authoring system which reduces the cost of courseware development significantly both in terms of dollars and time. The system has been used for courseware production at the University of Utah.