

Expert Systems Research: Modeling the Medical Decision Making Process

by

Edward I-i. Shortliffe and Lawrence M. Pagan

Departments of Medicine and Computer Science

Stanford University
Stanford, CA 94305

EXPERT SYSTEMS RESEARCH:
MODELING THE MEDICAL DECISION MAKING PROCESS*

Edward H. Shortliffe, MD, PhD

Lawrence M. Fagan, PhD**

Heuristic Programming Project

Departments of Medicine and Computer Science

Stanford University

Stanford, California 94305

Presented at the
Workshop on Integrated Approaches to Patient Monitoring

University of Florida at Gainesville

5-6 March 1982

*The Ventilator Manager project was supported by NIH Grant GM-24669, awarded to the Institute of Medical Science at Pacific Medical Center (San Francisco). Computing resources were provided by the SUMEX-AIM facility at Stanford University under NIH Grant RR-00785. Dr. Shortliffe is recipient of research career development award LM-00048 from the National Library of Medicine.

**Current address: Phd to MD program, University of Miami School of Medicine, Miami, Florida 33101.

Introduction

During the quarter century since the birth of the branch of computer science known as artificial intelligence (AI), **much** of the research has focused on developing symbolic models of human inference. In the last decade several related AI research themes have come together to form what is now known as "expert systems research." In this paper we review AI and expert systems to acquaint the reader with the field and to suggest ways in which this research will eventually be applied to advanced medical monitoring.

Knowledge Engineering

Artificial intelligence has been described as "the study of ideas that enable computers to do the things that make human beings seem intelligent."* An implicit assumption is that the computer should have the ability to reason symbolically (rather than by combining numbers statistically or by using other numerical manipulations that underlie conventional computer programs). Related assumptions are that intelligent programs should be able to acquire new knowledge and to apply it appropriately; they should also be able to manipulate and communicate ideas.

Of particular relevance to the study of medical inference is AI research into the construction of systems for knowledge-based consultation. "Knowledge" is the key word and must be distinguished from "data." Computers have long been used to store data, but isolated data points do not become knowledge until they have been analyzed and summarized. Accordingly, we suggest that there are at least four types of knowledge that should be distinguished from statistical data. These characterize the information that must be available to an expert system?

1. knowledge derived from data analysis (largely numerical or statistical);
2. judgmental or empirical subjective knowledge -- the kind that experts recognize is based on their own experience but which may be difficult to verify without complex and time-consuming studies;
3. common sense or scientific knowledge -- these kinds of knowledge are often simple facts (e.g., cars are for transportation, Gainesville is in Florida), but are symbolic in nature and must be known by an expert in a field;

4. strategic knowledge or **"self-knowledge"** -- this is the kind of knowledge that often distinguishes an expert from a well-trained novice in a field, e.g., an anesthesiologist's method of selecting an optimal anesthetic agent for a compromised patient or a cardiologist's favorite technique for choosing the diagnostic tests by which to assess a patient with a new complaint of chest pain; thus, **judgment** combined with a "strategic plan" can be used to adapt a program to idiosyncratic situations.

An expert system, then, is a computer program that contains and can apply specialized knowledge. It uses this knowledge to make suggestions to users who may not have the full range of expertise available to the system. The construction of such programs is known as "knowledge engineering"^{4,5} and typically requires close collaboration between human experts in a field and computer scientists familiar with expert systems research. In order to simulate an encounter between a non-expert and a human consultant, an expert system generally must contain all four kinds of knowledge outlined above-and, thus, is more than a conventional data retrieval system.

An expert system functions as an interface between the intended user of the system and the domain expert (or experts) who collaborated in its construction. Since the experts may not be generally available to all who would like their advice, the expert system becomes a surrogate.

During the encounter between an expert and a person seeking advice, there are three types of information transfer:

1. the expert requests information about the case under consideration;
2. the expert offers a recommendation or conclusion based on the available data; and
3. if requested, the expert explains the basis for the final decisions.

Conventional approaches to computer-based consultation typically address only the first two of these items. It is when a system designer wishes to construct programs that can explain the basis for their decisions, in terms that a physician can easily understand, that the techniques of knowledge engineering become particularly pertinent.

Questions Asked By Physicians

In our work designing and building medical expert systems at Stanford over the last decade, we **have** been guided by key questions that are frequently asked by physicians and that help to direct the design of the program:!

1. Do I need this system?
2. Will it help without being dogmatic?
3. Does it justify its recommendations so that I can judge them for myself?
4. Is it fast and easy to use?
5. Is it designed to make me feel comfortable when I use it?

The goal of most such systems is to answer all five questions in the affirmative. We will address these questions indirectly by discussing the major themes in expert systems research. An early expert system, known as MYCIN, is used to illustrate many pertinent principles. We also briefly describe a more recent program, the Ventilator Manager system, that applies knowledge engineering to monitoring patients in the intensive care unit.

The MYCIN System

The MYCIN program was developed at Stanford University in the mid-1970s⁷ and remains a good example of issues in the design and construction of knowledge-based programs. MYCIN was designed to advise physicians regarding the selection of antibiotics for patients with severe infections. When designing MYCIN, we were aware that the need for this kind of consultation system would not guarantee its acceptance by physicians. It was also important that it reach decisions comparable to those of infectious disease experts and that it be able to explain the basis for its decisions. Artificial intelligence techniques seemed to offer solutions to these design considerations!

Issues in Knowledge Engineering

Knowledge Representation

We have described the four kinds of knowledge that an expert system requires to provide specialized consultation. An active area of AI research is developing new techniques to encode and manipulate such symbolic (non-numeric) knowledge. The method used by the MYCIN and Ventilator Manager programs is called production rules (Fig. 1). Other methodologies include frames, semantic networks,* and combinations of these. Most knowledge can be encoded in whatever formalism a designer wishes, but the representation technique must accomplish the following:

1. formulation of explanations that convey a line of reasoning that a user can understand and critique;
2. separation of the domain knowledge from the program itself; in this way, knowledge can be changed without affecting the computer program;
3. a level of performance such that an expert can easily identify and correct faults in the knowledge; and
4. interaction with an inferential method such that the system reliably reaches excellent decisions.

Drawing on past experience with production rules in a large system that reasoned about chemical 'structure', we used production **rules**⁹ to encode the knowledge of infectious disease therapy". A production rule is a conditional statement that indicates the circumstances under which a-particular conclusion may be drawn (Fig. 1). MYCIN contains about 550 rules dealing with the diagnosis and treatment of bacteremia and meningitis. The rules are encoded in a stylized language that **the** computer can interpret, and routines have been written to translate them into English for display to the user as shown in the figure. The strengths of the conclusions are indicated by numerical weights called certainty factors (CFs).¹¹ Most expert systems that deal with uncertain inferences have been forced to incorporate scoring systems to keep track of the weight of evidence favoring competing hypotheses.

As is true in other complex domains, human actions and beliefs in medicine interact with

If: 1) The infection which requires therapy is meningitis,
 and
 2) The patient has evidence of a serious skin or soft
 tissue infection, and
 3) Organisms were not seen on the stain of the culture,
 and
 4) The **type** of infection is bacterial

Then : There is evidence that the organism (other than those
 seen on cultures or smears) which might be causing the
 infection is staphylococcus-coag-pos (.75) or
 streptococcus (.5)

Figure 1: A sample rule from the MYCIN system shows that the premise of the rule is a set of conditions (here the four clauses following the word "If:") that are tested by the program. If the conjunction of the premise conditions is true, the conclusion or action of the rule (the "Then:" clause) is executed and the inference is appropriate.

common sense and with precise knowledge in ways that defy simple delineation. The challenge at this stage in the development of expert systems is, therefore, to constrain the task realistically so that the program is tenable but still allows useful decisions for solving real problems. This has been demonstrated by Clancey in his use of MYCIN for tutoring medical students.* Rules that had been adequate for excellent performance in a consultation system were seriously deficient when they were used for teaching.¹³ We have also found that MYCIN-like rules need substantial modification to deal with inference in settings like medical monitoring where parameters change rapidly and analyses of temporal trends are **crucial**.¹⁴

Knowledge Acquisition

As researchers gain experience in **constructiing** consultation systems, identifying and encoding expert knowledge has been perceived as one of the most complex and arduous tasks encountered. Experts often have difficulty distilling their knowledge, and, because there are other major problems in identifying and encoding knowledge, the most appropriate domains for expert systems research at this time may therefore be fields in which the knowledge is already highly-structured and well-

specified.*.

In the case of MYCIN, we obtained rules regarding antibiotic selection and bacterial identification by talking with experts in infectious disease. We quickly learned that people express their knowledge best in the context of problem solving. Thus, we presented difficult cases to the experts, observed their performance as they gathered information they needed when deciding how to treat, and asked pertinent questions when the experts seemed to make leaps in logic or to ignore certain data. In this way, rules were obtained and then encoded using the formalism of production rules.

One member of our group, intrigued with the idea of a program that could guide this kind of knowledge gathering, developed a program known as TEIRESIAS.¹⁶ It used the rules already known to MYCIN combined with learning strategies to help an expert identify missing or erroneous knowledge and to update MYCIN's rules interactively. TEIRESIAS provides an example of one form of "machine learning" (learning by being told, as opposed to learning by experience or by analogy).

Models of Reasoning

Once knowledge has been captured from experts, books, or other sources and once it has been encoded, the program must have an effective method to search through the knowledge base for relevant facts and to tie them together in ways that simulate human reasoning. There are two issues: 1) control of the reasoning process, and 2) management of uncertainty.

Control of the Reasoning Process

Entire books have been written on techniques for traversing a symbolic search space.¹⁷ The approach we used in MYCIN is known as goal-directed reasoning or backward-chaining. MYCIN's rules are only loosely related to one another before a consultation begins, i.e., the system builder

*One of our present research projects, ONCOCIN,¹⁵ is a cancer chemotherapy consultation program. We chose this field for our current work precisely because the knowledge in the protocols used to guide the treatment of these patients is formalized, written down, and subject to rigorous analysis.

does not tell the program explicitly how or when the rules should interact. As it considers a particular patient, MYCIN selects the relevant rules and chains them together. Two rules chain together if the conclusion of one helps determine the truth value of a condition in the premise of the other; thus, the resulting reasoning network is created dynamically¹

MYCIN reasons backward from its goal of determining therapy for a patient. It starts by considering rules for therapy selection, but the premise of each of those rules in turn sets up new questions or subgoals. These new goals then invoke new rules and a reasoning network is thereby developed. When the truth of a premise is best determined by asking the physician rather than by applying a rule, e.g., to determine the value of a laboratory test likely to be known by the doctor, a question is displayed. The physician responds and the program continues to select additional rules. Once information on the patient is entered, some rules will fail to apply. The rules that are invoked provide a chain of inference specific to the case under consideration.

When reasoning is based on observations rather than goals, problem solving proceeds forward -- from data to conclusions -- so-called "data-driven" reasoning. Many expert systems use this alternative approach, and there are psychological data to suggest that "forward reasoning" more closely approximates the way human beings solve complex problems. As discussed below, the Ventilator Manager (VM) program uses the forward reasoning approach. A series of rules are examined to determine the reliability of incoming data. In successive steps, VM considers the meaning of the measurements in the current context.

Management of Uncertainty

When individual inference steps are not certain, a level of complexity is added to the interpretation of conclusions reached by an intelligent program. Not only are conventions needed to assign weights to the rules, but also, when two or more pieces of evidence support the same conclusion, some mechanism is needed to determine the net strength of the hypothesis. Expert system researchers have found that formal probability theory is less than satisfactory for these

purposes, although some have attempted to adapt it to symbolic reasoning **systems**.¹⁸ This issue has been particularly relevant for MYCIN, where the knowledge expressed in a rule tends to include “suggestive” or “strongly suggestive” evidence for a given conclusion. The need to combine pieces of evidence regarding a single hypothesis, but derived from a number of different rules, requires a numeric system to capture and represent an expert’s measure of belief regarding the inference stated in the rules. Conditional probabilities were not adequate for this purpose,” and we devised instead the system of “certainty factors” mentioned earlier. These numbers lie on a -1 to + 1 scale, with -1 indicating absolute disproof of a hypothesis, + 1 indicating its proof, and 0 indicating the absence of evidence for or against the hypothesis (or evidence equally weighted in both directions). We have described the model’s relationship to formal probability theory, and its methods for combining evidence from diverse sources (rules and user estimates).”

Generation of Good Advice

The ultimate test of the validity of the model of reasoning used in an expert system is its ability to reach accurate conclusions and thereby to give valuable advice. Our discussion of system evaluation below mentions some of the difficulties that arise in attempting to show that a program is performing at the level of an expert. There is a related controversy among expert systems researchers. Some do not believe it will be possible for computer-based consultants to function at the level of human experts until we better understand and model the inferences used by intelligent problem solvers. Others argue that the details of a reasoning model are not important as long as the system reaches good decisions and can explain its reasoning in understandable terms. However, psychological studies of human problem solving, e.g., Elstein’s work in the field of medicine,¹⁹ are being studied increasingly by knowledge engineers as they attempt to develop systems that are more **robust**.

Explanation of Decisions

As we have frequently stressed, a crucial aspect of the interaction between an expert and his client is the ability of the non-expert to request explanations regarding the recommendation received.

Our group recently surveyed several hundred physicians and found that they cited explanation as a principal requirement for a clinically acceptable computer-based consultation? In fact, most physicians felt that explanation is more important than absolute accuracy since they felt the latter is unrealistic. If a program can explain the basis for its conclusions, users can examine the reasoning carefully and decide for themselves whether to follow that advice. We believe this is important for an expert system in medicine; consultation programs are *tools* for use by individuals who are highly trained and unwilling to turn over their decisions to a computer program, regardless of how well it has been validated.

One of the great advantages of the rules used in MYCIN is the development of mechanisms to explain and justify system **performance**. These capabilities also contribute greatly to MYCIN's educational **role**.¹³ Since all questions asked by MYCIN are generated by a rule that is under consideration, and since all rules can be displayed in English, a rule is easily understood by a physician who wants to know why MYCIN has asked a particular question during a consultation. In addition, we have developed simple techniques for understanding free text questions entered by a physician who is obtaining a consultation.* MYCIN can thereby analyze a question and recover the rules used to make specific decisions during the interaction. Despite the power that rules provided for facilitating explanations in an expert system, there are still serious limitations to the MYCIN approach. Many of the problems with explanation parallel those with knowledge acquisition. For example, an expert whose thoughts jump directly from data to therapy provides little structure upon which an explanation can be based. This tendency to skip intermediate concepts prevents the exposition of general principles of action. Clancey has discussed some of these issues in his **thesis**,¹³ and we are continuing research in this **area**.²²

Validation and Evaluation

As expert systems have begun to mature, techniques have become necessary to demonstrate that they can perform as an expert in the field. MYCIN has been evaluated now in three large studies and we have discovered that the design of validation experiments is itself an area of research interest;

a number of specific points are pertinent.

First, any evaluation is difficult because there is so much difference of opinion in medicine, even among experts. Hence, it is unclear how to select a standard by which to measure the system's performance. Actual clinical outcome cannot be used because each patient is treated in only one way and because a poor outcome in a gravely ill patient cannot necessarily be ascribed to poor selection of therapy.

Second, although MYCIN performed at or near expert level in almost all cases, the evaluating experts in an early study had serious reservations about the clinical utility of the program. It is difficult to assess how much of this opinion is due to inadequacies in the knowledge or design of the system and how much to bias against *any* computer-based consultation aid. In a subsequent study, we attempted to eliminate this bias from the study by having the evaluators unaware of which recommendations were from MYCIN and which from physicians.²³ In that setting, MYCIN's recommendations were preferred uniformly or judged equivalent to those of five infectious disease experts.

Eventually, other questions must be answered regarding MYCIN and systems like it. Each question requires its own evaluation. Is the system used? If so, do the users follow the system's advice? If so, does the user benefit from the encounter with the system? Will the system be cost-effective? What are the legal implications in the use of, or failure to use, such systems? The answers to these questions are years away for most consultation systems but ultimately are just as important as whether the methodology leads to accurate and reliable advice.

Generalization

We mentioned earlier the importance of separating the knowledge in an expert system from the program that processes that knowledge and generates the advice. One reason for this separation is the ease with which information can be added or corrected. Many expert systems include "editors" that allow a knowledge engineer to modify the knowledge base without having to change the program

in any way.

There is a second advantage. One can imagine the development of interchangeable knowledge bases, each driven by the same program. Each knowledge base would have to be structured in accordance with the conventions required by the program, but once this were accomplished, a single program could generate advice in a number of different domains. This area of active research is known as "generalization" or the development of "system building tools."^{24, 25} The key idea is to develop a general purpose program that can be used by knowledge engineers to build an expert system in a new domain. Such programs must define conventions for the representation of knowledge and for the inferential models. The builders of systems must in turn comply with these conventions.

Consider, for example, the system building tool that we have constructed from MYCIN. By removing all knowledge of infectious diseases, i.e., all the rules, we were left with a set of programs that we call "Essential MYCIN" or **EMYCIN**.²⁵ EMYCIN can in turn be used to build an expert system in a new domain so long as it is natural to structure the knowledge in terms of production rules. Additional code was added to EMYCIN to allow the system designer to produce a knowledge base quickly and accurately. Several consultation programs have been developed using EMYCIN, including consultants for other medical problems and a consultant for structural engineering design.

An Expert System For Medical Monitoring

The Ventilator Manager (VM) program is an experiment in expert system development that builds on our experience with production rules in the MYCIN system. VM is designed to interpret on-line quantitative data in the intensive care unit (ICU). These data are used to manage post-operative mechanical ventilatory assistance. VM was strongly influenced by the MYCIN architecture outlined earlier, but the program was redesigned to allow for the description of events that change over time.

VM is an extension of a physiologic monitoring system,** and is designed to perform five specialized tasks in the ICU:

1. to detect possible errors in measurement;
2. to recognize untoward events in the patient/machine system and suggest corrective action,
3. to summarize the patient's physiologic status,
4. to suggest adjustments to therapy based on the patient's status over time and long-term therapeutic goals, and
5. to maintain a set of case-specific expectations and goals for future evaluation by the program.

The program interprets physiologic measurements over time and uses a model of intensive care therapies and clinical knowledge about the diagnostic implications of data.

Interpretation of Dynamic Data

Most medical decision making programs, including MYCIN, have based their advice on the data available at one particular time. In actual practice, the clinician receives additional information from tests and observations over time and reevaluates the diagnosis and prognosis of the patient. Both the progression of the disease and the response to previous therapy are important for assessing the patient's situation.

Data are collected in different therapeutic situations, or "contexts." In order to interpret the data properly, VM includes a model of the stages that a patient follows from ICU admission through the end of the critical monitoring phase. The correct interpretation of physiologic measurements depends on knowing which stage the patient is in. The goals for intensive care are also stated in terms of these clinical contexts. The program maintains descriptions of the current and optimal ventilator-y therapies for any given time.

** VM was developed as a collaborative research project between Stanford University and Pacific Medical Center (PMC) in San Francisco. It was tested with patient information acquired from a physiologic monitoring system implemented in the cardiac surgery ICU at PMC and developed by Dr. John Osborn and his colleagues.

IF: Relations about one or more parameters hold

THEN: 1) Make a conclusion. based on these facts:
 2) Make appropriate suggestions to
 clinicians; and
 3) Create new expectations about the
 future values of parameters.

Figure 2: The structure of a prototypical rule from the VM system is similar to the production rules used in MYCIN, with both a premise and conclusion specified. The "conclusion" of a VM rule frequently indicates actions that the system should take, e.g., suggest a change ventilator setting, in addition to inferences that can be drawn.

Knowledge Representation in VM

Knowledge is represented in VM by production rules similar to those used in MYCIN (Fig. 2). In addition to the premise and conclusion (or action) associated with each rule, several other parameters are included: (1) the rule's symbolic name, (2) the rule group, e.g., rules about instrument faults, (3) the main concept (definition) of the rule and (4) all of the therapeutic states in which it makes sense. Fig. 3 shows a sample rule for determining hemodynamic stability.

The VM knowledge base includes rules to support five reasoning steps that recur whenever new data from the monitoring system become available. These are:

1. rules to characterize measured data as reasonable or spurious;
2. rules to determine therapeutic state of the patient (viz., the mode of ventilation);
3. rules to adjust expectations of future values of measured variables when the patient's **state changes**;
4. rules to check physiologic status, including cardiac rate, hemodynamics, ventilation, or oxygenation; and
5. rules to check compliance with long-term therapeutic goals.

Each step is associated with a collection of rules, sorted by the type of conclusions made in the action

STATUS RULE: STABLE-HEMODYNAMICS
 DEFINITION: Defines stable hemodynamics based
 on blood pressures and heart rates.
 APPLIES to patients on VOLUME, CMV, ASSIST,
 T-PIECE
 COMMENT: Look at mean arterial pressure for
 changes in blood pressure and systolic
 blood pressure for maximum pressures.

IF:
 HEART RATE is ACCEPTABLE
 PULSE RATE does NOT CHANGE by 20 beats/minute
 in 15 minutes
 MEAN ARTERIAL PRESSURE is ACCEPTABLE
 MEAN ARTERIAL PRESSURE does NOT CHANGE by 15
 torr in 15 minutes
 SYSTOLIC BLOOD PRESSURE is ACCEPTABLE

THEN :
 The HEMODYNAMICS are STABLE

Figure 3: In a VM interpretation of a rule, the meaning of
 "ACCEPTABLE" varies with the clinical context, i.e., the type of
 ventilatory assistance. VOLUME, CMV, ASSIST & T-PIECE refer to the types
 of ventilation therapies for which VM has been given rule-based
 knowledge.

portion of the rule -- e.g., all rules that determine the validity of the data. Between steps (3) and (4)
 above, a special algorithm is used to compare the validated measurements with the current
 expectations, thus determining whether a measurement can be classified as high or low.

Symbolic Measurement Ranges

Most of the rules symbolically represent the values measured, and the terms "acceptable" or
 "ideal" characterize the appropriate ranges. The meaning of acceptable will change as the patient
 moves from state to state, but the statement of the relation between the physiologic measurements
 remains constant. The use of symbolic statements, e.g., "heart rate is acceptable," allows the
 exposition of common principles of physiologic interpretation in different contexts and minimizes the
 number of rules needed to describe the complexity of the diagnostic situation.

The meaning of the symbolic range is determined by rules that establish expectations about the value of **measured** data. For example, when a patient is taken off the ventilator, the upper limit of acceptability for the expired carbon dioxide measurement is raised. The actual numeric calculation of "expired **PCO2** high" in the premise of any rule changes when the context switches (removal from **ventilatory** support), but the statement of the rules remains the same.. A sample rule that creates these expectations is shown in Fig. 4.

```

INITIALIZING RULE: INITIALIZE-CMV
DEFINITION: Initialize expectations for
             patients on control led mandatory
             ventilation (CMV) therapy
APPLIES to all patients on CMV

IF ONE OF:
    PATIENT TRANSITIONED FROM VOLUME TO CMV
    PATIENT TRANSITIONED FROM ASSIST TO CMV

THEN EXPECT THE FOLLOWING:

```

	[acceptable range]					
	very low	low	[ideal]		high	very high
	---	---	min	max	---	---
MEAN PRESSURE	60	75	80	95	110	120
HEART RATE		60			110	
.				.		
.				.		
EXPIRED						
pCO2	22	28	30	35	42	50

Figure 4: This portion of an initializing rule establishes initial expectations of acceptable and ideal ranges of variables. Not all ranges are defined for each measurement. **PCO2** refers to the carbon dioxide in expired air measured at the mouth by one of the on-line monitoring devices.

Interpreting Rules

The VM rule interpreter is based on the **MYCIN** interpreter. The major changes are: (1) **forward-chaining** (data-driven) invocation of rules as opposed to backward-chaining, (2) checks to see that information acquired in a previous time frame is still valid for making conclusions, and (3) iteration

through appropriate parts of the rule set each time new information is available.

A data-driven approach is necessary to take advantage of the small set of measurement values available in each time frame. This means that the inference works forward from the available information as opposed to **working** backward from a goal and posing questions to the user when no data are available. Because of the demanding nature of the ICU, the system must acquire and interpret data with minimal staff intervention. Therefore, since some facts about patients will not be known to the system, caveats are attached to the suggestions it prints out.

Each of the five groups of rules (corresponding to the five reasoning steps mentioned above) is considered in order. Each rule is examined to determine whether it applies in the current context. For example, rules designed to evaluate “T-piece” breathing status are not examined when the T-piece is not in place. The premise of the rule is examined for validity, and the conclusions are recorded by the program along with expectations on the future ranges of measurement values. Suggestions to clinicians are also printed out.

Often the examination of the premise requires the use of a value acquired earlier in time, e.g., the temperature, which is volunteered to the monitoring system at intervals. The reliability of the stored value is determined by evaluating either a time constant (for variables that predictably change over time) or a rule (for cases in which the assessment of a value’s reliability is dependent upon context-specific information). Associated with each parameter in the system is a specific mechanism for determining its reliability over time. If a measurement is concluded to be spurious or outdated, then it is treated as if it were unknown, requiring alternate methods for determining the status of the patient. Rule invocation is repeated each time that a new set of measurements is available (currently every 2 to 10 minutes).

Identical conclusions made in contiguous time frames are represented by keeping track of the interval specified by the times of the first and last assertion. A list of these intervals summarizes the history of a particular conclusion. The evaluation of a premise such as “Patient hyperventilating for a

30-minute period within the last hour” is made by direct examination of the intervals stored along with conclusions, as opposed to looking at the original measurements. Expectations are associated with the appropriate measurement and are classified by duration and type, e.g., the upper limit of the acceptable range. Expectations can persist for a fixed interval, e.g., “for twenty minutes starting in ten minutes,” or for the duration of one or more clinical situations, e.g., while the patient is on the ventilator.

Comparison of Mycin and VM Design Gods

MYCIN was designed to serve on the ward as an expert consultant for antimicrobial therapy selection. A typical interaction might take place after the patient has been diagnosed and preliminary cultures drawn but little microbiological data are available. In critical situations, a tentative decision about therapy must often be made pending actual culture results. In return for assistance in making this decision, the clinician is asked to spend the small amount of time required to seek a consultation.

The intensive care unit is quite different, however. Continuous surveillance and evaluation of the patient's status is required. The problem is one of making therapeutic adjustments over a long period of time, many of which are minor, such as adjusting the respiratory rate on the ventilator. The main reasons for using VM are to monitor status or to investigate an unusual event. The program must therefore be able to interpret measurements with minimal human participation. When an interaction does take place, e.g., when an unexpected event is noted by the program, it must be terse and concise.

This difference in the timing and style of the man/machine interaction has considerable impact on system design. For example, the VM system must be able to do the following:

1. to reach effective decisions on the presumption that input from a clinician will be brief,
2. to use historical data to determine a clinical situation,
3. to provide advice at any point in the hospital course of the patient,
4. to follow-up on the outcomes of previous therapeutic decisions, and

5. to summarize conclusions made' over time.

VM's environment thus differs from **MYCIN's** in that natural language is an unlikely mode of communication.

A consultation program should also be able to model the changing medical environment so that the program can interpret the available data in context. Areas like infectious disease require an assessment of clinical problems in a variety of changing clinical situations, e.g., "patients after positive or negative culture data are available," "patients who are severely ill but lack culture results," "patients after partial or complete therapy," or "patients with acquired superinfection."

It is also necessary for VM to contain knowledge that can be used to evaluate its therapeutic advice, just as a human consultant follows a case over a period of time. This is complicated by the fact that the user of the system may not follow the **therapy** recommended. If the patient does not react as expected to the given therapy, then the program has to determine what alternate therapeutic steps may be required.

During the implementation of the VM program, we observed many types of clinical behavior that represent a challenge to symbolic modeling. One such behavior is the unwillingness of clinicians to change therapies frequently. After a patient meets the criteria for switching from therapy A to 'B, e.g., **IMV** to T-piece, clinicians tend to allow the patient's status to drop below optimal criteria before returning to therapy A. This was represented in the knowledge base by pairs of therapy selection rules (A to B, B to A) with a grey zone between the two criteria, e.g., "acceptable" limits might be used to suggest going from therapy A to therapy B, whereas "very high" or "very low" limits would be used for going from B to A. If the same limit were used for going in each direction, a small fluctuation of one measurement near a cut-off value would provide very erratic therapy suggestions. A more robust approach would be to make decisions in such situations based upon how long the patient has been in the given state and in accordance with the previous therapy or therapies.

A more detailed discussion of VM is included in a thesis based on the work,¹⁴ and, in another paper, we have described in greater detail the contrasts between MYCIN and this approach to monitoring.”

Advantages of AI for Medical Monitoring

Symbolic Models Provide a Context for Data Interpretation

The MYCIN and VM systems are experiments in how best to build symbolic models for environments in which clinical data are gathered. Both models were designed to provide a context in which numeric data could be interpreted flexibly. The model of “ventilator-y mode” that we developed for VM is used to guide the system’s interpretation of parameters such as heart rate and mean arterial pressure. Similarly, in MYCIN, the clinical context is determined by physical findings and other “soft” data, e.g., MYCIN’s list of previous medications and their time of administration is used to determine how lab findings should be interpreted in the context of “partially treated meningitis.”

Context is somewhat more problematic in VM because we have had to assume that only a small amount of current non-numeric information will be available to describe a patient. The context therefore must be deduced from some of the numerical information, which is, in turn, used to interpret the remaining data in the current or in subsequent time periods. For example, certain sudden changes in airway parameters can generally be interpreted as the context “suctioning the patient” and the interpretation of cardiac and hemodynamic parameters are then keyed into this new clinical situation. This high level model of the patient’s current state can be used to help in the analysis of trends in measurements, e.g., to provide a basis for predicting difficulties, to allow for selective tracking of “problems” as opposed to measurements, and to provide a more clinical orientation for subsequent collection of data.

Symbolic Models Can Simulate the Flexibility of Expert Reasoning

Symbolic processing can merge mathematic analysis with judgment or “intelligence” similar to that of an expert. For example, the VM program often throws out data because the monitor is operating outside its limits of reliability or is improperly connected. When the trend concerning this

time period is examined, the unreliable data are ignored, much as they would be by an experienced clinician observing the same trend information. Similarly, **recognition** that the clinical situation drastically changed in the immediate past can be used to tag as dubious any conclusions made by an averaging function applied to current data. VM uses production rules to determine the utility of “recent” measurements based on how quickly the clinical situation is changing. Just as the operational assumptions for the monitors can be tested, so can any mathematical or statistical procedure. Intelligent selection of a mathematical technique to analyze data typically requires expert knowledge of statistics as well as of the clinical domain. The use of such knowledge when picking and applying statistical tests for the analysis of medical data bases has recently been examined by Blum.²⁸

Symbolic Models Facilitate Statements Of Expectations

The ultimate goal of computer-based monitoring systems is to predict problems before they happen rather than to “alarm” as difficulties arise. Providing a context is the first step towards reaching that goal. VM has a limited mechanism for adjusting the acceptable limits for one or more variables based on recent events interpreted by the program. Other AI programs also address this problem, e.g., programs for military signal analysis and speech understanding can set up “future expectations” and then perform some analysis when future expectations are not realized. Similarly, air traffic control programs can plan ahead in time, can detect conflicts, and can plan alternate traffic routes. This type of computational capability would be especially useful in determining whether the patient’s response to therapy is acceptable.

Symbolic Models Provide an Ability to Select Relevant Data For Analysis

AI allows for a selective evaluation of the measured data. Many approaches to the interpretation of data have concentrated on selecting the “most important variable” or on creating a “figure of merit index” that tries to summarize difficulties into one specific number. Our approach to representing clinical knowledge allows the physician to specify a variety of relations between variables, or between variables and other available information. As always, using knowledge about

the current clinical context can be **used** to guide the evaluation of rules and, thereby, to draw attention to specific relations only when they are important. A single index can be replaced with specific subtopics, e.g., "**hemodynamic** status,*" and with their relations to therapeutic goals, e.g., "the hemodynamics have not been stable long enough to suggest removing ventilatory support." We wish to turn a large amount of data into useful diagnostic conclusions supported by logically referenced specific numerical data.

Summary

Several years of experience with MYCIN have led to an understanding of additional requirements for symbolic processing approaches to medical decision making. These include extending the knowledge base beyond the facts necessary for expert performance, providing a structure for a large number of production rules, and extending the inferential aids to include assistance throughout the patient's clinical course. For decision aids in the intensive care unit, or in other equally dynamic situations, programs cannot depend on interaction with the clinical users. Furthermore, they must handle data that are changing over time and may be missing or spurious. They must also be able to track the patient's status during the course of disease or in response to therapy. The VM program has suggested that knowledge engineering techniques, such as those developed for MYCIN, can be adapted to dynamic clinical settings such as monitoring patients in the intensive care unit.

References

1. Buchanan, B.G., "Research on expert systems," Report **HPP-81-1**, Heuristic Programming Project, Stanford University, February 1981, To appear in *Machine Intelligence 10*, (D. Michie, ed.), **1982**
2. Winston, P.H., *Artificial Intelligence*, Addison-Wesley, Reading, 1977.
3. Shortliffe E.H. Buchanan B.G. Feigenbaum E.A., "Knowledge engineering for medical decision making: a review of computer-based clinical decision aids," *Proceedings of the IEEE*, , 1979, pp. **1207- 1224**.
4. Feigenbaum, E.A., "The art of artificial intelligence: themes and case studies in knowledge engineering," *Proceedings of the 5th International Joint Conference on Artificial Intelligence*, Cambridge, Massachusetts, 1977, pp. **1014- 1029**.
5. Michie, D., "Knowledge engineering," *Cybernetics*, Vol. 2, 1973, pp. 197-200.
6. Shortliffe, E.H., "Consultation systems for physicians: the role of artificial intelligence techniques," *Proceedings Third National Conference of the Canadian Society for Computational Studies of Intelligence*, Canadian Society for Computational Studies of Intelligence, Victoria, British Columbia, May 1980, Also in *Readings in Artificial Intelligence* (B. Webber and N. Nilsson, eds.) Menlo Park: Tioga Publishing Co., 1981.
7. Shortliffe, E.H., *Computer-based medical consultations: MYCIN*, Elsevier/North Holland, New York, 1976.
8. Buchanan B.G. Feigenbaum E.A., "DENDRAL and Meta-DENDRAL: their applications dimension," *Artificial Intelligence*, Vol. 11, 1978, pp. 5-24.
9. Davis R. King J., "An overview of production systems," in *Machine Representation of Knowledge*, E.W. Elcock and D. Michie, eds., Wiley and Sons, New York, 1976.
10. Davis R. Buchanan B.G. Shortliffe E.H., "Production rules as a representation for a knowledge-based consultation program," *Artificial Intelligence*, Vol. 8, 1977", pp. 15-45.
11. Shortliffe E.H. Buchanan B.G., "A model of inexact reasoning in medicine," *Math. Biosci.*, Vol. **23**, 1975, pp. **351-379**.
12. Clancey, W.J., "Tutoring rules for guiding a case method dialogue," *Int. J. Man-Machine Studies*, Vol. 11, 1979, pp. **25-49**.
13. Clancey, W.J., "Transfer of rule-based expertise through a tutorial dialogue," Memo STAN-CS-79-769, Computer Science Department, Stanford University, September 1979, Doctoral Dissertation
14. **Fagan, L.M.**, *VM: Representing time-dependent relations in a clinical setting*, PhD dissertation, Heuristic Programming Project, Stanford University, 1980.
15. Shortliffe E.H. Scott A.C. Bischoff M.B. et al., "ONCOCIN: An expert system for oncology protocol management," *Proceedings of the 7th International Joint Conference on Artificial*

- Intelligence, **IJCAI**, Vancouver, British Columbia, August 1981.
16. Davis, R., "Applications of meta-level knowledge to the construction, maintenance, and use of large knowledge bases," Memo HPP-76-7, Heuristic Programming Project, Stanford University, July 1976, Doctoral Dissertation
 17. **Nilsson**, N.J., Problem solving methods in *artificial* intelligence, McGraw-Hill Book Company, Inc., San Francisco, 1971.
 18. **Duda** R.O. Gaschnig J. Hart P.E. et al., "Development of the PROSPECTOR consultation system for mineral exploration," Final report SRI Projects 5821 and 6415, SRI International, 1978.
 19. Elstein A.S. Shulman L.S. and Sprafka S.A., *Medical problem solving: an analysis of clinical reasoning*, Harvard University Press, Cambridge, 1978.
 20. Teach R.L. Shortliffe E.H., "An analysis of physician attitudes regarding computer-based consultation systems," *Comput. Biomed. Res.*, Vol. 14, 1980, pp. 542-558.
 21. Scott A.C. Clancey W.J. Davis R. Shortliffe E.H., "Explanation capabilities of knowledge-based production systems," *Amer. J. Computational Linguistics*, 1977, .
 22. **Wallis** J.W. Shortliffe E.H., "Explanatory power for medical expert systems: studies in the representation of causal relationships for clinical consultations", Submitted for publication, December 1981.
 23. Yu V.L. **Fagan** L.M. Wraith S.M. et al., "Antimicrobial selection by a computer: a blinded evaluation by infectious disease experts," *JA MA*, Vol. 242, 1979, pp. 1279-1 282.
 24. Nii H.P. Aiello N., "AGE (Attempt to Generalize): A knowledge-based program for building knowledge-based programs," *Proceedings of the 6th International Joint Conference on Artificial Intelligence*, Tokyo, Japan, August 1979, pp. 645-655.
 25. van **Melle**, W., "A domain-independent system that aids in constructing knowledge-based consultation programs," Memo HPP-80-11, Heuristic Programming Project, Stanford University, June 1980.
 26. Osborn J. J. Beaumont J.O. Raison J.C.A. Abbott R.P., "Computation for quantitative on-line measurement in an intensive care ward," in *Computers in Biomedical Research*, R.W. Stacey and B.D. **Waxman**, eds., Academic Press, New York, 1969, pp. 207-237.
 27. **Fagan** L.M. Shortliffe E.H. Buchanan B.G., "Computer-based medical decision making: from MYCIN to VM," *Automedica*, Vol. 3, 1980, pp. 97-1 06.
 28. Blum, R.L., "Discovery, confirmation, and incorporation of causal relationships from a large time-oriented clinical database: the RX project," in *Computers in Biomedical Research*, H.R. Warner, ed., Academic Press, New York, 1982, pp. 164-187.

