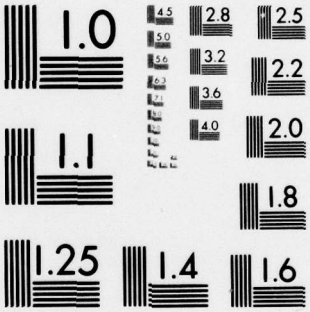
AD-A074 076 STANFORD UNIV CALIF DEPT OF COMPUTER SCIENCE F/G 6/5
KNOWLEDGE ENGINEERING FOR MEDICAL DECISION MAKING: A REVIEW OF --ETC(U)
FEB 79 E H SHORTLIFFE, B G BUCHANAN MDA903-77-C-0322
UNCLASSIFIED STAN-CS-79-723 NL

| OF |
AD
A074076

END
DATE
FILMED
10-79
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

LEVEL II

12

AD A074076

KNOWLEDGE ENGINEERING FOR MEDICAL DECISION MAKING: A Review of Computer-Based Clinical Decision Aids

by

Edward H. Shortliffe, MD, PhD
Bruce G. Buchanan, PhD
Edward A. Feigenbaum, PhD

DDC
RECEIVED
SEP 19 1979
C

STAN-CS-79-723
Feruary 1979

COMPUTER SCIENCE DEPARTMENT
School of Humanities and Sciences
STANFORD UNIVERSITY

DDC FILE COPY

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

79 09 18 210

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

14

1. REPORT NUMBER: STAN-CS-79-723

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

6

4. TITLE (and Subtitle): KNOWLEDGE ENGINEERING FOR MEDICAL DECISION MAKING: A Review of Computer-Based Clinical Decision Aids,

5. TYPE OF REPORT & PERIOD COVERED: technical, February 1979

6. PERFORMING ORG. REPORT NUMBER: STAN-CS-79-723

10

7. AUTHOR(S): Edward H. Shortliffe, M.D., Ph.D.; Bruce G. Buchanan, Ph.D.; Edward A. Feigenbaum, Ph.D.

15

8. CONTRACT OR GRANT NUMBER(S): MDA903-77-C-0322

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

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

11

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

12. REPORT DATE: February 1979

13. NUMBER OF PAGES: 52

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

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

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Reproduction in whole or in part is permitted for any purpose of the U.S. Government.

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

9 Technical rept.

18. SUPPLEMENTARY NOTES

12 53 p.

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

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

(see reverse side)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

094 120

Computer-based models of medical decision making account for a large proportion of clinical computing efforts. This article reviews representative examples from each of several major medical computing paradigms. These include (1) clinical algorithms, (2) clinical databanks that include analytic functions, (3) mathematical models of physical processes, (4) pattern recognition, (5) Bayesian statistics, (6) decision analysis, and (7) symbolic reasoning or artificial intelligence. Because the techniques used in the various systems cannot be examined exhaustively, the case studies in each category are used as a basis for studying general strengths and limitations. It is noted that no one method is best for all applications. However, emphasis is given to the limitations of early work that have made artificial intelligence techniques and knowledge engineering research particularly attractive. We stress that considerable basic research in medical computing remains to be done and that powerful new approaches may lie in the melding of two or more established techniques.

Accession For
NTIS GRA&I
DDC TAB
Unannounced
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or special
A

To Appear in the Proceedings of the IEEE
September 1979

KNOWLEDGE ENGINEERING FOR MEDICAL DECISION MAKING:
A Review of Computer-Based Clinical Decision Aids

Edward H. Shortliffe, MD, PhD
Bruce G. Buchanan, PhD
Edward A. Feigenbaum, PhD

Heuristic Programming Project
Departments of Medicine and Computer Science
Stanford University
Stanford, California 94305

*Please send reprint requests to Dr. Shortliffe at the following address:
Department of Medicine (General Internal Medicine)
Stanford University School of Medicine
Stanford, California 94305

Abstract

Computer-based models of medical decision making account for a large proportion of clinical computing efforts. This article reviews representative examples from each of several major medical computing paradigms. These include (1) clinical algorithms, (2) clinical databanks that include analytic functions, (3) mathematical models of physical processes, (4) pattern recognition, (5) Bayesian statistics, (6) decision analysis, and (7) symbolic reasoning or artificial intelligence. Because the techniques used in the various systems cannot be examined exhaustively, the case studies in each category are used as a basis for studying general strengths and limitations. It is noted that no one method is best for all applications. However, emphasis is given to the limitations of early work that have made artificial intelligence techniques and knowledge engineering research particularly attractive. We stress that considerable basic research in medical computing remains to be done and that powerful new approaches may lie in the melding of two or more established techniques.

KNOWLEDGE ENGINEERING FOR MEDICAL DECISION MAKING: A Review of Computer-Based Clinical Decision Aids

1 Introduction

As early as the 1950's, physicians and computer scientists recognized that computers could assist with clinical decision making [63] and began to analyze medical diagnosis with a view to the potential role of automated decision aids in that domain [61]. Since that time a variety of techniques have been applied, accounting for at least 800 references in the clinical and computing literature [112]. In this article we review several medical decision making paradigms and discuss some issues that account for both the multiplicity of approaches and the limited clinical success of most systems developed to date. Because other authors have reviewed computer-aided diagnosis [47],[92],[114] and the potential impact of computers in medical care [93], our emphasis here is somewhat different. We will focus on the symbolic representation and use of knowledge, termed "knowledge engineering," and the inadequacies of data-intensive techniques which have led to the exploration of novel symbolic reasoning approaches during the last decade.

1.1 Reasons For Attempting Computer-Aided Medical Decision Making

Because of the accelerated growth in medical knowledge, physicians have tended to specialize and to become more dependent upon assistance from other experts when a patient presents with a complex problem outside one's own area of expertise. The primary care physician who first sees a patient has thousands of tests available with a wide range of costs (both fiscal and physical) and potential benefits (i.e., arrival at a correct diagnosis or optimal therapeutic management). Even the experts in a specialized field may reach very different decisions regarding the management of a specific case [131]. Diagnoses that are made, and upon which therapeutic decisions are based, have been shown to vary widely in their accuracy [26],[83],[89]. Furthermore, medical students usually learn about decision making in an unstructured way, largely through observation and by emulating the thought processes they perceive to be used by their clinical mentors [53].

Thus the motivations for attempts to understand and automate the process

of clinical decision making have been numerous [114]. They are directed both at diagnostic models and at assisting with patient management decisions. Among the reasons for introducing computers into such work are the following:

(1) To improve the accuracy of clinical diagnosis through approaches that are systematic, complete, and able to integrate data from diverse sources;
(2) To improve the reliability of clinical decisions by avoiding unwarranted influences of similar but not identical cases (a common source of bias among physicians), and by making the criteria for decisions explicit, and hence reproducible;
(3) To improve the cost efficiency of tests and therapies by balancing the expenses of time, inconvenience, or funds against benefits and risks of definitive actions;
(4) To improve our understanding of the structure of medical knowledge, with the associated development of techniques for identifying inconsistencies and inadequacies in that knowledge; and
(5) To improve our understanding of clinical decision making, in order to improve medical teaching and to make computer programs more effective and easier to understand.

1.2 The Distinction Between Data And Knowledge

The models on which computer systems base their clinical advice range from data-intensive to knowledge-intensive approaches. There are at least four types of knowledge that may be distinguished from pure statistical data:

(1) knowledge derived from data analysis (largely numerical);
(2) judgmental or subjective knowledge;
(3) scientific or theoretical knowledge; and
(4) high-level strategic knowledge or "self-knowledge."

If there is a chronology to the field over the last 20 years, it is that there has been progressively less dependence on "pure" observational data and more emphasis on higher-level symbolic knowledge inferred from primary data. We include with domain knowledge the category of "judgmental knowledge" which reflects the experience and opinions of an expert regarding an issue about which the formal data may be fragmentary or nonexistent. Since many decisions made in clinical medicine depend upon this kind of judgmental expertise, it is not surprising that investigators should begin to look for ways to capture and use

the knowledge of experts in decision making programs. Another reason to move away from purely data-intensive programs is that in medicine the primary data available to decision makers are far from objective [20],[57]. They include subjective reports from patients, and error-prone observations [27]. Also, the terminology used in the reports is not standardized [7] and the classifications often overlap. Thus decision making aids must be knowledgeable about the unreliability of the data [57] as well as the uncertainty of the inference.

For example, data-intensive programs include medical record systems which accumulate large databanks to assist with decision making. There is little knowledge per se in the databank, but there are large amounts of data which can help with decisions and be analyzed to provide new knowledge. A program that retrieves a patient's record for review, or even one that identifies and retrieves the records of similar patients (matching some set of descriptors), is performing a data management task with little reasoning involved [36],[86]. Although there is statistical "knowledge" contained in the conditional probabilities generated from such a databank and utilized for Bayesian analysis, it is all numeric. At the other extreme are systems that encode and use the kind of expert knowledge which cannot be easily gleaned from databanks or literature reviews [75],[102]. Systems that model human reasoning or emphasize education of users tend to fall towards this end of the data-knowledge continuum.

In addition to judgmental and statistical knowledge, there are other forms of information that can play an important role in computer-based clinical decision aids. For example, underlying scientific theories and relationships are often ignored by diagnostic programs but provide the foundation for decisions made by human experts. Consider, for example, the potential utility of techniques that could effectively represent and use the basic knowledge of biochemistry, biophysics, or detailed human physiology. Biomedical modeling research offers some mathematical techniques for encoding such knowledge in certain domains, but symbolic approaches and clinically useful applications are still largely unrealized.

Finally, there is another kind of knowledge used by human decision makers -- an understanding of reasoning processes and strategies themselves. This kind of "high-level" or "meta-level" knowledge, if incorporated into computer programs, may not only heighten their decision making performance but also augment their acceptability to users by making them appear more aware of their own power, strategies, and limitations.

We use the term "knowledge engineering," then, to refer to computer-based symbolic reasoning issues such as knowledge representation, acquisition, explanation, and "self-awareness" or self-modification [19]. It is along these dimensions that knowledge-based programs differ most sharply from conventional calculations. For example, they can solve problems by pursuing a line of reasoning; the individual inference steps and the whole chain of reasoning may also form the basis for explanations of decisions. A major concern in knowledge engineering is clear separation of the medical knowledge in a program from the inference mechanism that applies that knowledge to the data of individual cases. One goal of this paper is to identify, in the strengths and weaknesses of earlier work, those issues which have motivated several current researchers to investigate the automation of clinical decision aids through knowledge engineering.

1.3 Parameters For Assessing Work In The Field

Barriers to successful implementation of computer-based diagnostic systems have been analyzed on several occasions [7],[23],[106] and need not be reviewed here. However, in assessing programs it is pertinent to examine several parameters that affect the success and scope of a particular system in light of its intended users and application. Unfortunately, the medical computing literature has few descriptions of systems for which all the following issues can be assessed.

(1) How accurate is the program?[1]

(2) What is the nature of the knowledge in the system and how is it generated or acquired?

(3) How is the clinical knowledge represented, and how does it facilitate the performance goals of the system described?

(4) How are knowledge and clinical data used and how does this impact on system performance?

(5) Is the system accepted by the users for whom it is intended? Is the interface with the user adequate? Does the system function outside of a research setting and is it suitable for dissemination?

(6) What are the limitations of the approach?

---

[1]Although this is important it is not the only measure of clinical effectiveness. For example, the effects on morbidity, mortality, and length of hospital stay may also be important parameters. As we shall show, few systems have reached a stage of implementation where these parameters could be assessed. Moreover, because of the complexity of the interacting influences that affect the usual measures of outcome, it may be difficult ever to define the marginal benefit of such systems.

An issue we have chosen not to address is the cost of a system, including the size of the required computing resource. Not only is information on this question scanty for most of the programs, but expenses generated in a research and development environment do not realistically reflect the costs one expects from a system once it is operating for service use.

1.4 Overview Of This Paper

An exhaustive review of computer-aided diagnosis will not be attempted in light of the vastness of the field, and we have therefore chosen to present the prominent paradigms by discussing representative examples. In separate sections we give an overview, example, and discussion of (1) clinical algorithms, (2) databank analysis, (3) mathematical models, (4) pattern recognition, (5) Bayesian analysis, (6) decision theory, and (7) symbolic reasoning. We close each section by identifying the range of applications for which the approach appears most appropriate, the limitations of the approach, and the ways in which symbolic reasoning techniques may strengthen the approach by improving its performance or acceptability.

The seven principal examples we have selected are not necessarily the best nor the most successful; however, they illustrate the issues we wish to discuss within the major paradigms. We have also referenced other closely related systems, so the bibliography should guide the reader to more details on particular topics. Any attempt to categorize programs in this way is inherently fraught with problems in that several systems draw upon more than one paradigm. Thus we have occasionally felt obligated to simplify a topic for clarity in light of the overall purposes of this review and the limitations of the space available to us.

Because we are only interested here in decision making tools for use by clinicians, we have chosen to disregard systems that are designed primarily for use by researchers [39],[50], [65],[90]. Furthermore, we shall not discuss biomedical engineering applications of computers, such as advanced automated instrumentation techniques (e.g., computerized tomography[2]) or signal processing techniques (e.g., programs for EKG analysis [79] or patient monitoring [116]). Because they do not explicitly make inferences, we have also omitted programs designed largely for data storage and retrieval with the actual analysis and decision making left to the clinician [36],[58],[124]. We have also chosen to

---

[2]See Kak's article in this issue of the PROCEEDINGS.

discuss working computer programs rather than unimplemented theories or early reports of work in progress.

2 Clinical Algorithms and Automation

2.1 Overview

Clinical algorithms, or protocols, are flowcharts to which a diagnostician or therapist can refer when deciding how to manage a patient with a specific clinical problem [97]. Such protocols usually allow decisions to be made by carefully following the simple branching logic, although there are built-in safeguards whereby referrals to experts are made if a patient is unusually complex. The value of a protocol depends upon the infrequency with which such referrals are made, so it is important to design algorithms that reflect an appropriate balance between safety and efficiency. In general, algorithms have been designed by expert physicians for use by paramedical personnel who have been entrusted with the performance of certain routine clinical-care tasks[3]. The methodology has been developed in part because of a desire to define basic medical logic concisely so that detailed training in pathophysiology would not be necessary for ancillary practitioners. Experience has shown that intelligent high school graduates, selected in large part because of poise and warmth of personality, can provide excellent care guided by protocols after only four to eight weeks of training. This care has been shown to be equivalent to that given by physicians for the same limited problems, and to be accepted by physicians and patients alike for such diverse clinical situations as diabetes management [56],[66], pharyngitis [38], headache [37], and other disease categories [104],[110].

The role of the computer in such applications has been limited, however. In fact, several groups initially experimented with computer representation of the algorithms but have since abandoned the efforts and resorted to prepared paper forms [56],[110]. In these cases the computer had originally guided the physician assistant's collection of data and had specified precisely what decisions should be made or actions taken, in accordance with the clinical algorithm. However, since the algorithmic logic is generally simple, and can

---

[3]Clinical algorithms have also been prepared for use by physicians themselves, but Grimm has found that they are generally less well-accepted by doctors [38]. He showed, however, that physician performance could improve when protocols were used in certain settings.

often be represented on a single sheet of paper, the advantages of an automated approach over a manual system have not been clearly demonstrated. In one study Vickery showed that supervising physicians could detect no significant difference between the performance of physicians' assistants using automated versus manual systems, although the computer system entirely eliminated errors in data collection (since it demanded all relevant data at the appropriate time) [110]. Furthermore, the computer could not, of course, decide whether the actual observations entered by the physicians' assistant were correct; yet this kind of inaccuracy was one of the most common reasons that supervisors found an assistant's performance unsatisfactory.

There are two other ways in which the computer has been used in the setting of clinical algorithms. First, mathematical techniques have been used to analyze signs and symptoms of diseases and thereby to identify those that should most appropriately be referenced in corresponding clinical algorithms [30],[55],[113]. The process for distilling expert knowledge in the form of a clinical algorithm can be an arduous and imperfect one [97]; formal techniques to assist with this task may prove to be very valuable.

Some researchers in this area also use computers to assist with clinical care audit by comparing actual actions taken by a physicians' assistant with those recommended by the algorithm itself. Sox et al. [104] have described a system in which the assistant's checklist for a patient encounter was sent to a central computer and analyzed for evidence of deviation from the accepted protocol. Computer-generated reports then served as feedback to the physicians' assistant and to the supervising physician.

2.2 Example

We have selected for discussion a project that differs from those previously cited in that (1) computer techniques are still being used, and (2) the clinical algorithms are designed for use by primary care physicians themselves. This is the cancer chemotherapy system developed in Alabama by Mesel et al. [70]. The algorithms were developed to allow private practitioners, at a distance from the regional tertiary-care center, to manage the complex chemotherapy for their cancer patients without routinely referring them to the central oncologists. Mesel et al. have described a "consultant-extender system" that enables the primary physician to treat patients with Hodgkin's Disease under the supervision of a regional specialist. Five

oncologists developed a care protocol for the treatment of Hodgkin's Disease, and this algorithm was placed on-line. Once patients had agreed to participate in the study, their private physicians would prepare "encounter forms" at the time of each office visit. These forms would document pertinent interval history, physical findings, and lab data, as well as chemotherapy administered. The form would then be sent to the regional center where it was analyzed by the computer and a customized clinical algorithm was produced to assist the private physician with the management of that patient during the next appointment. Thus the computer program would take into account the ways in which the individual patient's disease might progress or improve and would prepare an appropriate clinical algorithm. This protocol was sent back to the physician in time for it to be available at the next office visit. The private practitioner was encouraged to call the regional specialist directly if the protocol seemed in some way inadequate or additional questions arose. The authors present data suggesting that their system was well-accepted by physicians and patients, and that excellent care was delivered[4]. Retrospective review of cases that were treated at the referral center itself, but without the use of the protocols, showed a 16% rate of variance from the management guidelines specified in the algorithms; there was no such variance when the protocols were followed. Thus algorithms may be effective tools for the administration of complex specialized therapy in circumstances such as those described[5].

2.3 Discussion of the Methodology

Although clinical algorithms are among the most widespread and best accepted of the decision aids described in this article, the simplicity of their logic makes it clear why the technique cannot be effectively applied in most medical domains. Decision points in the algorithms are generally binary (i.e., a given sign or symptom is either present or absent), and there tend to be many circumstances that can arise for which the user is advised to consult the supervising physician (or specialist). Thus the difficult decision tasks are left to experts, and there is generally no formal algorithm for managing the case from that point on. It is precisely the simplicity of the algorithmic

---

[4]This is an interesting result in light of Grimm's experience mentioned in footnote 3. One possible explanation is that physicians were more accepting of the algorithmic approach in Mesel's case because it allowed them to perform tasks that they would previously not have been able to undertake.

[5]More recently the Alabama group has reported similar success implementing a consultant-extender system for adjuvant chemotherapy in breast carcinoma [129].

logic, and the safeguard of the supervising expert, which has permitted many algorithms to be represented on one or two sheets of paper and has obviated the need for direct computer use in most of the systems. The contributions of clinical algorithms to the distribution and delivery of health care, to the training of paramedics, and to quality care audit, have been impressive and substantial. However, the approach is not suitable for extension to the complex decision tasks to be discussed in the following sections.

3 Databank Analysis for Prognosis and Therapy Selection

3.1 Overview

Automation of medical record keeping and the development of computer-based patient databanks have been major research concerns since the earliest days of medical computing. Most such systems have attempted to avoid direct interaction between the computer and the physician recording the data, with the systems of Weed [123],[124] and Greenes [36] being notable exceptions. Although the earliest systems were designed merely as record-keeping devices, there have been several recent attempts to create programs that could also provide analyses of the information stored in the computer databank. Some early systems [36],[52] had retrieval modules that identified all patient records matching a Boolean combination of descriptors; however, further analyses of these records for decision making purposes was left to the investigator. Weed has not stressed an analytical component in his automated problem-oriented record [124], but others have developed decision aids which use medical record systems fashioned after his [103].

The systems for databank analysis all depend on the development of a complete and accurate medical record system. Once such a system is developed, a number of additional capabilities can be provided: (1) correlations among variables can be calculated, (2) prognostic indicators can be measured, and (3) the response to various therapies can be compared. A physician faced with a complex management decision can look to such a system for assistance in identifying patients in the past who had similar clinical problems and can then see how those patients responded to various therapies. A clinical investigator keeping the records of his study patients on such a system can use the program's statistical capabilities for data analysis. Hence, although these applications are inherently data-intensive, the kinds of "knowledge" generated by specialized

retrieval and statistical routines can provide valuable assistance for clinical decision makers. For example, they help avoid the inherent biases of anecdotal experience, such as occur when an individual practitioner bases decisions primarily on personal encounters with one or two patients having a rare disease or complex of symptoms.

There are many excellent programs in this category, one of which is discussed in some detail in the next section. Several others warrant mention, however. The HELP System at the University of Utah [117],[119],[120] uses a large data file on patients in the Latter-Day Saints Hospital. Clinical experts formulate specialized "HELP sectors" which are collections of logical rules that define the criteria for a particular medical decision. These sectors are developed by an interactive process; the expert proposes important criteria for a given decision and is provided with actual data regarding that criterion (based on relevant patients and controls from the computer databank). The criteria in the sector are thus adjusted by the expert until adequate discrimination is made to justify using the sector's logic as a decision tool[6]. The sectors are then used for a variety of tasks throughout the hospital.

Another system of interest is that of Feinstein et al. at Yale [21], in which physicians interact with the system to request assistance in estimating prognosis and guiding management for patients with lung cancer. Similarly, Rosati et al. have developed a system at Duke University which uses a large databank on patients who have undergone coronary arteriography [88]. New patients can be matched against those in the databank to help determine patient prognosis under a variety of management alternatives.

3.2 Example

One of the most successful projects in this category is the ARAMIS system of Fries at Stanford University [24]. The approach was designed originally for use in an outpatient rheumatology clinic, but then broadened to a general clinical database system, the Time-Oriented Databank (TOD) [126],[127], so that it could be transferred to clinics in oncology, metabolic disease, cardiology, endocrinology, and certain pediatric subspecialties. All clinic records are kept in a tabular format in which a column in a large table indicates a specific clinic visit and the rows indicate the relevant clinical parameters that are

---

[6]This process might be seen as a technique to assist with the formulation of clinical algorithms as discussed in the previous section. Another approach using databank analysis for algorithm development is described in [30].

being followed over time. These charts are maintained by the physicians seeing the patient in clinic, and the new column of data is later transferred to the computer databank by a transcriptionist; in this way time-oriented data on all patients are kept current. The defined database (clinical parameters to be followed) is determined by clinical experts, and in the case of rheumatic diseases has now been standardized on a national scale [41].

The information in the databank can be used to create a prose summary of the patient's current status, and there are graphical capabilities which can plot specific parameters for a patient over time [126]. However, it is in the analysis of stored clinical experience that the system has its greatest potential utility [25]. In addition to performing search and statistical functions such as those developed in databank systems for clinical investigation [50],[65], ARAMIS offers a prognostic analysis for a new patient when a management decision is to be made. Using the consultative services of the Stanford Immunology Division, an individual practitioner may select clinical indices for his patient that he would like matched against other patients in the databank. It is imperative that such indices be selected wisely and hence with expert advice; the Stanford immunologists have found that the best descriptors for characterizing patients are often different from those that a novice chooses to use. Based on two to five such descriptors, the computer locates relevant prior patients and prepares a report outlining their prognosis with respect to a variety of endpoints (e.g., death, development of renal failure, arthritic status, pleurisy). Therapy recommendations are also generated on the basis of a response index that is calculated for the matched patients. A prose case analysis for the physician's patient can also be generated; this readable document summarizes the relevant data from the databank and explains the basis for the therapeutic recommendation.

The rheumatologic databank generated under ARAMIS has now been expanded to involve a national network of immunologists who are accumulating time-oriented data on their patients. This national project seeks in part to obtain enough data so that groups of retrieved patients will be sizable, thereby controlling for some observer variability and making the system's recommendations more statistically defensible.

3.3 Discussion of the Methodology

Databank analysis systems have powerful capabilities to offer to the

individual clinical decision maker. Furthermore, medical computing researchers recognize the potential value of large databanks in supporting many of the other decision making approaches discussed in subsequent sections. There are important additional issues regarding databank systems:

(1) Data acquisition remains a major problem. Many systems have avoided direct physician-computer interaction but have then been faced with the expense and errors of transcription. The developers of one well-accepted record system still express their desire to implement a direct interface with the physician for these reasons, although they recognize the difficulties encountered in encouraging direct use of a computer system by doctors [107].

(2) Analysis of data in the system can be complicated by missing values that frequently occur, outlying values, and poor reproducibility of data across time and among physicians. Conversely, the system can itself be used to identify questionable values of tests or observations.

(3) The decision aids provided tend to emphasize patient management rather than diagnosis. Feinstein's system [21] is only useful for patients with lung cancer, for example, and the ARAMIS prognostic routines, which are designed for patient management, assume that the patient's rheumatologic diagnosis is already known.

(4) There is no formal correlation between the way expert physicians approach patient management decisions and the way the programs arrive at recommendations. Feinstein and Koss felt that the acceptability of their system would be limited by a purely statistical approach, and they therefore chose to mimic human reasoning processes to a large extent [59], but their approach appears to be an exception.

(5) Data storage space requirements can be large since the decision aids of course require a comprehensive medical record system as a basic component.

Slamecka has distinguished between structured and empirical approaches to clinical consulting systems [103], pointing out that databanks provide a largely empirical basis for advice whereas structured approaches rely on judgmental knowledge elicited from the literature or from experts. It is important to note, however, that judgmental knowledge is itself based on empirical information. Even an expert's "intuitions" are based on observations and "data collection" over years of experience. Thus one might argue that large, complete, and flexible databanks <u>could</u> form the basis for large amounts of

judgmental knowledge that we now have to elicit from other sources. Some researchers have indicated a desire to experiment with methods for the automatic generation of medical decision rules from databanks, and one component of the research on Slamecka's MARIS system is apparently pointed in that direction [103]. Indeed, some of the most exciting and practical uses of large databanks may be found precisely at the interface with those knowledge engineering tasks that have most confounded researchers in medical symbolic reasoning [5].

4 Mathematical Models of Physical Processes

4.1 Overview

Pathophysiologic processes can be well-described by mathematical formulae in a limited number of clinical problem areas. Such domains have lent themselves well to the development of computer-based decision aids since the issues are generally well-defined. The actual techniques used by such programs tend to reflect the details of the individual applications, the most celebrated of which have been in pharmacokinetics (specifically digitalis dosing), acid-base/electrolyte disorders, and respiratory care [69].

It is important that cooperating experts assist with the definition of pertinent variables and the mathematical characterization of the relationships among them. The computer program requests the relevant data, makes the appropriate computations, and provides a clinical analysis or recommendation for therapy. Some of the programs have also involved branched-chain logic to guide decisions about what further data are needed for adequate analysis[7].

Programs to assist with digitalis dosing have gradually introduced broader medical knowledge over the last ten years. The earliest work was Jelliffe's [48] and was based upon his considerable experience studying the pharmacokinetics of the cardiac glycosides. His computer program used mathematical formulations based on parameters such as therapeutic goals (e.g., desired predicted blood levels), body weight, renal function, and route of administration. In one study he showed that computer recommendations reduced the frequency of adverse digitalis reactions from 35% to 12% [49]. Later,

---

[7] "Branched-chain" logic refers to mechanisms by which portions of a decision network can be considered or ignored depending upon the data on a given case. For example, in an acid-base program the anion gap might be calculated and a branch-point could then determine whether the pathway for analyzing an elevated anion gap would be required. If the gap were not elevated, that whole portion of the logic network could be skipped.

another group revised the Jelliffe model to permit a feedback loop in which the digitalis blood levels obtained with initial doses of the drug were considered in subsequent therapy recommendations [78],[96]. More recently, a third group in Boston, noting the insensitivity of the first two approaches to the kinds of nonnumerical observations that experts tend to use in modifying digitalis therapy, augmented the pharmacokinetic model with a patient-specific model of clinical status [35]. Running their system in a monitoring mode, in parallel with actual clinical practice on a cardiology service, they found that each patient in the trial in whom toxicity developed had received more digitalis than would have been recommended by their program.

4.2 Example

Perhaps the best known program in this category is the interactive system developed at Boston's Beth Israel Hospital by Bleich. Originally designed as a program for assessment of acid-base disorders [2], it was later expanded to consider electrolyte abnormalities as well [3],[4]. The knowledge in Bleich's program is a distillation of his own expertise regarding acid-base and electrolyte disorders. The system begins by collecting initial laboratory data from the physician seeking advice on a patient's management. Branched-chain logic is triggered by abnormalities in the initial data so that only the pertinent sections of the extensive decision pathways created by Bleich are explored. The approach is therefore similar to the flowcharting techniques used by the clinical algorithms of Section 2, but it involves more complex mathematical relationships than algorithms typically do. Essentially all questions asked by the program are numerical laboratory values or "yes-no" questions (e.g., "Does the patient have pitting edema?"). Depending upon the complexity and severity of the case, the program eventually generates an evaluation note that may vary in length from a few lines to several pages. Included are suggestions regarding possible causes of the observed abnormalities and suggestions for correcting them. Literature references are also provided with the recommendations.

Although the program was made available at several East Coast institutions, few physicians accepted it as an ongoing clinical tool. Bleich points out that part of the reason for this was the system's inherent educational impact; physicians simply began to anticipate its analysis after they had used it a few times [3][8].

---

[8] More recently he has been experimenting with the program operating as a monitoring system, thereby avoiding direct interaction with the physician.

The system's lack of sustained acceptance by physicians is probably due to more than its educational impact, however. For example, there is no feedback in the system; every patient is seen as a new case and the program has no concept of following a patient's response to prior therapy. Furthermore, the program generates differential diagnosis lists but does not pursue specific etiologies; this can be particularly bothersome when there are multiple coexistent disturbances in a patient and the program simply suggests parallel lists of etiologies without noting or pursuing the possible interrelationships.

Finally, the system is highly individualized in that it contains only the parameters and relationships that Bleich specifically thought were important to include in the logic network. Of course human consultants also give personalized advice which may differ from that obtained from other experts. However, a group of researchers in Britain [85] who compared Bleich's program to four other acid-base/electrolyte systems, found total agreement among the programs in only 20% of test cases when these systems were asked to define the acid-base disturbance and the degree of compensation present. Their analysis does not reveal which of the programs reached the correct decision, however, and it may be that the results are more an indictment of the other four programs than a valid criticism of the advice from Bleich's acid-base component.

4.3 Discussion of the Methodologies

The programs mentioned in this section differ from one another in several respects, and each tends to overlap with other paradigms we have discussed. Bleich's program, for example, is essentially a complicated clinical algorithm interfaced with mathematical formulations of electrolyte and acid-base pathophysiology. As such it suffers from the weaknesses of all algorithmic approaches, most importantly its highly structured and inflexible logic which is unable to contend with circumstances not specifically anticipated in the algorithm. The digitalis dosing programs all draw on mathematical techniques from the field of biomedical modeling [40], but have recently shown more reliance on methods from other areas as well. In particular these have included symbolic reasoning methods that allow clinical expertise to be encoded and used in conjunction with mathematical techniques [35]. The Boston group that developed this most recent digitalis program is interested in similarly developing an acid-base/electrolyte system so that judgmental knowledge of experts can be interfaced with the mathematical models of pathophysiology[9].

---

[9]This project was described by Prof. Peter Szolovits, of MIT's clinical

There is also a large research community of mathematicians who attempt to understand and characterize physical processes by devising simulation models [40]. Although such models are largely empirical and have generally not found direct application in clinical medicine, their research role may eventually be broadened to provide practical decision aids through interfaces with the other paradigms described in this review.

The major strength of mathematical models is their ability to capture mathematically sound relationships in a concise and efficient computer program. However, the major limitation, as with most of the paradigms discussed here, is that few areas of medicine are amenable to firm, quantitative description. Because the accuracy of the results depend on correct identification of relevant parameters, the precision and certainty of the relationships among them, and the accuracy of the techniques for measuring them, mathematical models have limited applicabiity at present. Furthermore, those domains that do lend themselves to mathematical description may still benefit from interactions with symbolic reasoning techniques, as has been demonstrated in the digitalis therapy adviser [35].

5 Statistical Pattern Recognition Techniques

5.1 Overview

Pattern recognition techniques define the mathematical relationship between measurable features and classifications of objects [15],[51]. In medicine, the presence or absence of each of several signs and symptoms in a patient may be definitive for the classification of the patient as "abnormal" or into the category of a specific disease. They are also used for prognosis [1], or predicting disease duration, time course, and outcomes. These techniques have been applied to a variety of medical domains, such as image processing and signal analysis, in addition to computer-assisted diagnosis.

In order to find the diagnostic pattern, or discriminant function, the method requires a training set of objects, for which the correct classification is already known, as well as reliable values for their measured features. If the form and parameters are not known for the statistical distributions underlying the features, then they must be estimated. Parametric techniques

---

decision making group, during a workshop on artificial intelligence in medicine at the University of Tokyo in November 1978.

focus on learning the parameters of the probability density functions, while non-parametric (or "distribution-free") techniques make no assumptions about the form of the distributions. After training, then, the pattern can be compared to new, unclassified objects to aid in deciding the category to which the new object belongs[10].

There are numerous variations on this general approach, most notably in the mathematical techniques used to extract characteristic measurements (the features) and to find and refine the pattern classifier during training. For example, linear regression analysis is a commonly used technique for finding the coefficients of an equation that defines a recurring pattern or category of diagnostic or prognostic interest. A class of patients can be described by a feature vector $X = [x_1, x_2, \ldots, x_n]$ (where x_i is one of n descriptive variables). The goal is to produce an equation relating the posterior probabilities[11] of each diagnostic class to the feature vector through a set of n coefficients (a_i)[12]:

$$P(D_i|X) = a_1x_1 + a_2x_2 + \ldots + a_nx_n$$

Recent work emphasizes structural relationships among sets of features more than statistical ones.

Three of the best known training criteria for the discriminant function are:

(a) least-squared-error criterion: choose the function that minimizes the squared differences between predicted and observed measurement values;

(b) clustering criterion: choose the function that produces the tightest clusters;

(c) Bayes' criterion: choose the function that has the minimum cost associated with incorrect diagnoses[13].

---

[10]It is possible to detect patterns, even without a known classification for objects in the training set, with so-called "unsupervised" learning techniques. Also, it is possible to work with both numerical and non-numerical measurements.

[11]The posterior probability of a diagnostic class, represented as $P(D_i|X)$, is the probability that a patient falls in diagnostic category D_i given that the feature vector X has been observed.

[12]See [62] for a study in which the coefficients are reported because of their medical import.

[13]This is one of many uses of Bayes' Theorem, a definitional rule that relates posterior and prior probabilities. For an overview of its use as a diagnostic rule (as opposed to a training criterion) and a definition of the formula, see Section 6.

Ten commonly used mathematical models based on these criteria have been shown to produce remarkably similar diagnostic results for the same data [7].

5.2 Example

There are numerous papers on uses of pattern recognition methods in medicine. Armitage [1] discusses three examples of prognostic studies, with an emphasis on regression methods. Goldwyn et al. [31] discuss uses of cluster analysis. One recent diagnostic application by Patrick [73] uses Bayes' criterion to classify patients having chest pains into three categories: D_1: acute myocardial infarction (MI); D_2: coronary insufficiency; and D_3: non-cardiac causes of chest pain. The need for early diagnosis of heart attacks without laboratory tests is a prevalent problem, yet physicians are known to misclassify about one third of the patients in categories D_1 and D_2 and about 80% of those in D_3. In order to determine the correct classification, each patient in the training set was classified after 3 days, based on laboratory data including electrocardiogram (ECG) and blood data (cardiac enzymes). There remained some uncertainty about several patients with "probable MI." Seventeen variables were selected from many: 9 features with continuous values (including age, heart rates, white blood count, and hemoglobin) and 8 features with discrete values (sex and 7 ECG features).

The training data were measurements on 247 patients. The decision rule was chosen using Bayes' theorem to compute the posterior probabilities of each diagnostic class given the feature vector X ($X = [x_1, x_2, \ldots, x_{17}]$. Then a decision rule was chosen to minimize the probability of error by adjusting the coefficients on the feature vector X such that for the correct class D_i:

$$P(D_i|X) = MAX[P(D_1|X), P(D_2|X), P(D_3|X)]$$

The class conditional probability density functions must be estimated initially, and the performance of the decision rule depends on the accuracy of the assumed model.

Using the same 247 patients for testing the approach, the trained classifier averaged 80% correct diagnoses over the three classes, using only data available at the time of admission. Physicians, using more data than the computer, averaged only 50.5% correct over these three categories for the same patients. Training the classifier with a subset of the patients, and using the remainder for testing, produced nearly as good results.

5.3 Discussion of the Methodology

The number of reported medical applications of pattern recognition techniques is large, but there are also numerous problems associated with the approach. The most obvious difficulties are choosing the set of features in the first place, collecting reliable measurements on a large sample, and verifying the initial classifications among the training data. Current techniques are inadequate for problems in which trends or movement of features are important characteristics of the categories. Also the problems for which existing techniques are accurate are those that are well characterized by a small number of features ("dimensions of the space").

As with all techniques based on statistics, the size of the sample used to define the categories is an important consideration. As the number of important features and the number of relevant categories increase, the required size of the training set also increases. In one test [7], pattern classifiers trained to discriminate among 20 disease categories from 50 symptoms were correct 51% - 64% of the time. The same methods were used to train classifiers to discriminate between 2 of the diseases, from the same 50 symptoms, and produced correct diagnoses 92% - 98% of the time.

The context in which a local pattern is identified raises problems related to the issue of utilizing medical knowledge. It is difficult to find and use classifiers that are best for a small decision, such as whether an area of an X-ray is inside or outside the heart, and integrate those into a global classifier, such as one for abnormal heart volume.

Accurate application of a classifier in a hospital setting also requires that the measurements in that clinical environment are consistent with the measurements used to train the classifier initially. For example, if diseases and symptoms are defined differently in the new setting, or if lab test values are reported in different ranges, or different lab tests used, then decisions based on the classification are not reliable.

Pattern recognition techniques are often misapplied in medical domains in which the assumptions are violated. Some of the difficulties noted above are avoided in systems that integrate structural knowledge into the numerical methods and in systems that integrate human and machine capabilities into single, interactive systems. These modifications will overcome one of the major difficulties seen in completely automated systems, that of providing the system with good "intuitions" based on an expert's *a priori* knowledge and experience [51].

6 Bayesian Statistical Approaches

6.1 Overview

More work has been done on Bayesian approaches to computer-based medical decision making than on any of the other paradigms we have discussed. The appeal of Bayes' Theorem[14] is clear: it potentially offers an exact method for computing the probability of a disease based on observations and data regarding the frequency with which these observations are known to occur for specified diseases. In several domains the technique has been shown to be exceedingly accurate, but there are also several limitations to the approach which we discuss below.

In its simplest formulation, Bayes' Theorem can be seen as a mechanism to calculate the probability of a disease, in light of specified evidence, from the *a priori* probability of the disease and the conditional probabilities relating the observations to the diseases in which they may occur. For example, suppose disease D_i is one of n mutually exclusive diagnoses under consideration and E is the evidence or observations supporting that diagnosis. Then if $P(D_i)$ is the *a priori* probability of the ith disease [15]:

$$P(D_i|E) = \frac{P(D_i)\ P(E|D_i)}{\sum_{j=1}^{n} P(D_j)\ P(E|D_j)}$$

The theorem can also be represented or derived in a variety of other forms, including an odds/likelihood ratio formulation. We cannot include a full discussion here, but any introductory statistics book or Lusted's volume [64] presents the subject in considerable detail.

Among the most commonly recognized problems with the utilization of a Bayesian approach is the large amount of data required to determine all the conditional probabilities needed in the rigorous application of the formula. Chart review or computer-based analysis of large databanks occasionally allows most of the necessary conditional probabilities to be obtained. A variety of additional assumptions must be made. For example: (1) the diseases under consideration are assumed mutually exclusive and exhaustive (i.e., the patient is assumed to have one of the n diseases), (2) the clinical observations are

---

[14] also often referred to as Bayes' rule, discriminant, or criterion

[15] Here $P(D_i|E)$ is the probability of the ith disease *given* that evidence E has been observed; $P(E|D_i)$ is the probability that evidence E will be observed in the setting of the ith disease.

assumed to be conditionally independent over a given disease[16], and (3) the incidence of the symptoms of a disease is assumed to be stationary (i.e., the model does not allow for changes in disease patterns over time).

One of the earliest Bayesian programs was Warner's system for the diagnosis of congenital heart disease [115]. He compiled data on 83 patients and generated a symptom-disease matrix consisting of 53 symptoms (attributes) and 35 disease entities. The diagnostic performance of the computer, based on the presence or absence of the 53 symptoms in a new patient, was then compared to that of two experienced physicians. The program was shown to reach diagnoses with an accuracy equal to that of the experts. Furthermore, system performance was shown to improve as the statistics in the symptom-disease matrix stabilized with the addition of increasing numbers of patients.

In 1968 Gorry and Barnett pointed out that Warner's program had required making all 53 observations for every patient to be diagnosed, a situation which would not be realistic for many clinical applications. They therefore used a modification of Bayes' Theorem in which observations are considered sequentially[17]. Their computer program analyzed observations one at a time, suggested which test would be most useful if performed next, and included termination criteria so that a diagnosis could be reached, when appropriate, without needing to make all the observations [32]. Decisions regarding tests and termination were made on the basis of calculations of expected costs and benefits at each step in the logical process[18]. Using the same symptom-disease matrix developed by Warner, they were able to attain equivalent diagnostic performance using only 6.9 tests on average[19]. They pointed out that, because the costs of medical tests may be significant (in terms of patient discomfort, time expended, and financial expense), the use of inefficient testing sequences should be regarded as ineffective diagnosis. Warner has also more recently included Gorry and Barnett's sequential diagnosis approach in an application regarding structured patient history-taking [118].

---

[16]The purest form of Bayes' Theorem allows conditional dependencies, and the order in which evidence is obtained, to be explicitly considered in the analysis. However, the number of required conditional probabilities is so unwieldy that conditional independence of observations, and non-dependence on the order of observations, is generally assumed [108].

[17]A similar approach was devised in Russia at approximately the same time by Vishnevskiy and associates. Their analyses, and a summary of the impressive amount of statistical data they have amassed, are contained in [111].

[18]See the decision theory discussion in Section 7.

[19]Tests for determining attributes were defined somewhat differently than they had been by Warner. Thus the maximum number of tests was 31 rather than the 53 observations used in the original study.

The medical computing literature now includes many examples of Bayesian diagnosis programs, most of which have used the nonsequential approach, in addition to the necessary assumptions of symptom independence and mutual exclusiveness of disease as discussed above. One particularly successful research effort has been chosen for discussion.

6.2 Example

Since the late 1960's deDombal and associates, at the University of Leeds (England), have been studying the diagnostic process and developing computer-based decision aids using Bayesian probability theory. Their area of investigation has been gastrointestinal diseases, originally acute abdominal pain [12] with more recent analyses of dyspepsia [44] and gastric carcinoma [134].

Their program for assessment of acute abdominal pain was evaluated in the emergency room of their affiliated hospital [12]. Emergency physicians filled out data sheets summarizing clinical and laboratory findings on 304 patients presenting with abdominal pain of acute onset. The data from these sheets became the attributes that were subjected to Bayesian analysis; the required conditional probabilities had been previously compiled from a large group of patients with one of seven possible diagnoses[20]. Thus the Bayesian formulation assumed each patient had one of these diseases and would select the most likely on the basis of recorded observations. Diagnostic suggestions were obtained in batch mode and did not require direct interaction between physician and computer; the program could generate results in from 30 seconds to 15 minutes depending upon the level of system use at the time of analysis [43]. Thus the computer output could have been made available to the emergency room physician, on average, within 5 minutes after the data form was completed and handed to the technician assisting with the study.

During the study [12], however, these computer-generated diagnoses were simply saved and later compared to (a) the diagnoses reached by the attending clinicians, and (b) the ultimate diagnosis verified at surgery or through appropriate tests. Although the clinicians reached the correct diagnosis in only 65%-80% of the 304 cases (with accuracy depending upon an individual's training and experience), the program was correct in 91.8% of cases. Furthermore, in 6 of the 7 disease categories the computer was proved more

---

[20] appendicitis, diverticulitis, perforated ulcer, cholecystitis, small bowel obstruction, pancreatitis, and non-specific abdominal pain.

likely than the senior clinician in charge of a case to assign the patient to the correct disease category. Of particular interest was the program's accuracy regarding appendicitis, a diagnosis which is often made incorrectly. In no cases of appendicitis did the computer fail to make the correct diagnosis, and in only six cases were patients with non-specific abdominal pain incorrectly classified as having appendicitis. Based on the actual clinical decisions, however, over 20 patients with non-specific abdominal pain were unnecessarily taken to surgery for appendicitis, and in six cases patients with appendicitis were "watched" for over eight hours before they were finally taken to the operating room.

These investigators also performed a fascinating experiment in which they compared the program's performance based on data derived from 600 real patients, with the accuracy the system achieved using "estimates" of conditional probabilities obtained from experts [60][21]. As discussed above, the program was significantly more effective than the unaided clinician when real-life data were used. However, it performed significantly <u>less</u> well than clinicians when expert estimates were used. The results supported what several other observers have found, namely that physicians often have very little idea of the "true" probabilities for symptom-disease relationships.

Another Leeds study of note was an analysis of the effect of the system on the performance of clinicians [13]. The trial we have mentioned that involved 304 patients was eventually extended to 552 before termination. Although the computer's accuracy remained in the range of 91% throughout this period, the performance of clinicians was noted to improve markedly over time. Fewer negative laparotomies were performed, for example, and the number of acute appendices that perforated (ruptured) also declined. However, these data slowly returned towards baseline after the study was terminated, suggesting that the constant awareness of computer monitoring and feedback regarding system performance had temporarily generated a heightened awareness of intellectual processes among the hospital's surgeons.

6.3 Discussion of the Methodology

The ideal matching of the problem of acute abdominal pain and Bayesian analysis must be emphasized; the technique cannot necessarily be as effectively

---

[21] Such estimates are referred to as "subjective" or "personal" probabilities, and some investigators have argued that they should be used in Bayesian systems when formally derived conditional probabilities are not available [64].

applied in other medical domains where the following limitations of the Bayesian approach may have a greater impact.

(1) The assumption of conditional independence of symptoms usually does not apply and can lead to substantial errors in certain settings [72]. This has led some investigators to seek new numerical techniques that avoid the independence assumption [8]. If a pure Bayesian formulation is used without making the independence assumption, however, the number of required conditional probabilities becomes prohibitive for complex real world problems [108].

(2) The assumption of mutual exclusiveness and exhaustiveness of disease categories is usually false. In actual practice concurrent and overlapping disease categories are common. In deDombal's system, for example, many of the abdominal pain diagnoses missed were outside the seven "recognized" possibilities; if a program starts with an assumption that it need only consider a small number of defined likely diagnoses, it will inevitably miss the rare or unexpected cases (precisely the ones with which the clinician is most apt to need assistance).

(3) In many domains it may be inaccurate to assume that relevant conditional probabilities are stable over time (e.g., the likelihood that a particular bacterium will be sensitive to a specific antibiotic). Furthermore, diagnostic categories and definitions are constantly changing, as are physicians' observational techniques, thereby invalidating data previously accumulated[22]. A similar problem results from variations in a priori probabilities depending upon the population from which a patient is drawn[23]. Some observers feel that these are major limitations to the use of Bayesian techniques [16].

In general, then, a purely Bayesian approach can so constrain problem formulation as to make a particular application unrealistic and hence unworkable. Furthermore, even when diagnostic performance is excellent such as in deDombal's approach to abdominal pain evaluation, clinical implementation and system acceptance will generally be difficult. Forms of representation that allow explanation of system performance in familiar terms (i.e., a more

---

[22]Although gradual changes in definitions or observational techniques may be statistically detectable by database analysis, a Bayesian analysis that uses such data is inevitably prone to error.

[23]deDombal has examined such geographic and population-based variations in probabilities and has reported early reports of his analysis [14].

congenial interface with physician users) will heighten clinical acceptance; it is at this level that Bayesian statistics and symbolic reasoning techniques may most beneficially interact.

7 Decision Theoretical Approaches

7.1 Overview

Bayes' Theorem is only one of several techniques used in the larger field of decision analysis, and there has recently been increasing interest in the ways in which decision theory might be applied to medicine and adapted for automation. Several excellent reviews of the field are available in basic reviews [45], textbooks [84], and medically-oriented journal articles [67],[94],[109]. In general terms, decision analysis can be seen as any attempt to consider values associated with choices, as well as probabilities, in order to analyze the processes by which decisions are made or should be made. Schwartz identifies the calculation of "expected value" as central to formal decision analysis [94]. Ginsberg contrasts medical classification problems (e.g., diagnosis) with broader decision problems (e.g., "What should I do for this patient?"), and asserts that most important medical decisions fall in the latter category and are best approached through decision analysis [29].

The following topics are among the central issues in the field:

(1) Decision Trees. The decision making process can be seen as a sequence of steps in which the clinician selects a path through a network of plausible events and actions. Nodes in this tree-shaped network are of two kinds: decision nodes, where the clinician must choose from a set of actions, and chance nodes, where the outcome is not directly controlled by the clinician but is a probabilistic response of the patient to some action taken. For example, a physician may choose to perform a certain test (decision node) but the occurrence or nonoccurrence of complications may be largely a matter of statistical likelihood (chance node). By analyzing a difficult decision process before taking any actions, it may be possible to delineate in advance all pertinent chance and decision nodes, all plausible outcomes, plus the paths by which these outcomes might be reached. Furthermore, data may exist to allow specific probabilities to be associated with each chance node in the tree.

(2) Expected Values. In actual practice physicians make sequential decisions based on more than the probabilities associated with the chance node that follows. For example, the best possible outcome is not necessarily sought if the costs associated with that "path" far outweigh those along alternate pathways (e.g., a definitive diagnosis may not be sought if the required testing procedure is expensive or painful and patient management will be unaffected; similarly, some patients prefer to "live with" an inquinal hernia rather than undergo a surgical repair procedure). Thus, anticipated "costs" (financial, complications, discomfort, patient preference) can be associated with the decision nodes. Using the probabilities at chance nodes, the costs at decision nodes, and the "value" of the various outcomes, an "expected value" for each pathway through the tree (and in turn each node) can be calculated. The ideal pathway, then, is the one which maximizes the expected value.

(3) Eliciting Values. Obtaining from physicians and patients the costs and values they associate with various tests and outcomes can be a formidable problem, particularly since formal analysis requires expressing the various costs in standardized units. One approach has been simply to ask for value ratings on a hypothetical scale, but it can be difficult to get the physician or patient to keep the values[24] separate from their knowledge of the probabilities linked to the associated chance nodes. An alternate approach has been the development of lottery games. Inferences regarding values can be made by identifying the odds, in a hypothetical lottery, at which the physician or patient is indifferent regarding taking a course of action with certain outcome and betting on a course with preferable outcome but with a finite chance of significant negative costs if the "bet" is lost. In certain settings this approach may be accepted and provide important guidelines in decision making [77].

(4) Test Evaluation. Since the tests which lie at decision nodes are central to clinical decision analysis, it is crucial to know the predictive value of tests that are available. This leads to consideration of test sensitivity, specificity, receiver operator characteristic curves, and sensitivity analysis. Such issues are discussed by Komaroff in this issue of the Proceedings [57] and have also been summarized elsewhere in the clinical literature [68].

---

[24] also termed "utilities" in some references; hence the term "utility theory" [84].

Many of the major studies of clinical decision analysis have not specifically involved computer implementations. Schwartz et al. examined the workup of renal vascular hypertension, developing arguments to show that for certain kinds of cases a purely qualitative theoretical approach was feasible and useful [94]. However, they showed that for more complex clinically challenging cases the decisions could not be adequately sorted out without the introduction of numerical techniques. Since it was impractical to assume that clinicians would ever take the time to carry out a detailed quantitative decision analysis by hand, they pointed out the logical role for the computer in assisting with such tasks and accordingly developed the system we discuss as an example below [33].

Other colleagues of Schwartz at Tufts have been similarly active in applying decision theory to clinical problems. Pauker and Kassirer have examined applications of formal cost-benefit analysis to therapy selection [74] and Pauker has also looked at possible applications of the theory to the management of patients with coronary artery disease [76]. An entire issue of the New England Journal of Medicine has also been devoted to papers on this methodology [46].

7.2 Example

Computer implementations of clinical decision analysis have appeared with increasing frequency since the mid-1960's. Perhaps the earliest major work was that of Ginsberg at Rand Corporation [28], with more recent systems reported by Pliskin and Beck [80] and Safran et al. [91].

We will briefly describe here the program of Gorry et al., developed for the management of acute renal failure [33]. Drawing upon Gorry's experience with the sequential Bayesian approach previously mentioned [32], the investigators recognized the need to incorporate some way of balancing the dangers and discomforts of a procedure against the value of the information to be gained. They divided their program into two parts: phase I considered only tests with minimal risk (e.g., history, examination, blood tests) and phase II considered procedures involving more risk and inconvenience. The phase I program considered 14 of the most common causes of renal failure and used a sequential test selection process based on Bayes' Theorem and omitting more advanced decision theoretical techniques [32]. The conditional probabilities used were subjective estimates obtained from an expert nephrologist and were

therefore potentially as problematic as those discussed by Leaper et al. [60] (see Section 6.2). The researchers found that they had no choice but to use expert estimates, however, since detailed quantitative data were not available either in databanks or the literature.

It is in the phase II program that the methods of decision theory were employed because it was in this portion of the decision process that the risks of procedures became important considerations. At each step in the decision process this program considers whether it is best to treat the patient immediately or to first carry out an additional diagnostic test. To make this decision the program identifies the treatment with the highest current expected value (in the absence of further testing), and compares this with the expected values of treatments that could be instituted if another diagnostic test were performed. Comparison of the expected values are made in light of the risk of the test in order to determine whether the overall expected value of the test is greater than that of immediate treatment. The relevant values and probabilities of outcomes of treatment were obtained as subjective estimates from nephrologists in the same way that symptom-disease data had been obtained. All estimates were gradually refined as they gained experience using the program, however.

The program was evaluated on 18 test cases in which the true diagnosis was uncertain but two expert nephrologists were willing to make management decisions. In 14 of the cases the program selected the same therapeutic plan or diagnostic test as was chosen by the experts. For three of the four remaining cases the program's decision was the physicians' second choice and was, they felt, a reasonable alternative plan of action. In the last case the physicians also accepted the program's decision as reasonable although it was not among their first two choices.

7.3 Discussion of the Methodology

The excellent performance of Gorry's program, despite its reliance on subjective estimates from experts, may serve to emphasize the importance of the clinical analysis that underlies the decision theoretical approach. The reasoning steps in managing clinical cases have been dissected in such detail that small errors in the probability estimates are apparently much less important than they were for deDombal's purely Bayesian approach [60]. Gorry suggests this may be simply because the decisions made by the program are based

on the combination of large aggregates of such numbers, but this argument should apply equally for a Bayesian system. It seems to us more likely that distillation of the clinical domain in a formal decision tree gives the program so much more knowledge of the clinical problem that the quantitative details become somewhat less critical to overall system operation. The explicit decision network is a powerful knowledge structure; the "knowledge" in deDombal's system lies in conditional probabilities alone and there is no larger scheme to override the propagation of error as these probabilities are mathematically manipulated by the Bayesian routines.

The decision theory approach is not without problems, however. Perhaps the most difficult problem is assigning numerical values (e.g., dollars) to a human life or a day of health, etc. Some critics feel this is a major limitation to the methodology [120]. Overlapping or coincident diseases are also not well-managed, unless specifically included in the analysis, and the Bayesian foundation for many of the calculations still assumes mutually exclusive and exhaustive disease categories. Problems of symptom conditional dependence still remain, and there is no easy way to include knowledge regarding the time course of diseases. Gorry points out that his program was also incapable of recognizing circumstances in which two or more actions should be carried out concurrently. Furthermore, decision theory per se does not provide the kind of focusing mechanisms that clinicians tend to use when they assume an initial diagnostic hypothesis in dealing with a patient and discard it only if subsequent data make that hypothesis no longer tenable. Other similar strategies of clinical reasoning are becoming increasingly well-recognized [53] and account in large part for the applications of symbolic reasoning techniques to be discussed in the next section.

8 Symbolic Reasoning Approaches

8.1 Overview

In the early 1970's researchers at several institutions simultaneously began to investigate potential clinical applications of symbolic reasoning techniques drawn from the branch of computer science known as artificial intelligence (AI). The field is well-reviewed in a recent book by Winston [128]. The term "artificial intelligence" is generally accepted to include those computer applications that involve symbolic inference rather than strictly

numerical calculations. Examples include programs that reason about mineral exploration, organic chemistry, or molecular biology; programs that converse in English and understand spoken sentences; and programs that generate theories from observations.

Such programs gain their power from qualitative, experiential judgments, codified in so-called "rules-of-thumb" or "heuristics", in contrast to numerical calculation programs whose power derives from the analytical equations used. The heuristics focus the attention of the reasoning program on parts of the problem that seem most critical and parts of the knowledge base that seem most relevant. They also guide the application of the domain knowledge to an individual case by deleting items from consideration as well as focusing on items. The result is that these programs pursue a line of reasoning as opposed to following a sequence of steps in a calculation. Among the earliest symbolic inference programs in medicine was the diagnostic interviewing system of Kleinmuntz [54]. Other early work included Wortman's information processing system, the performance of which was largely motivated by a desire to understand and simulate the psychological processes of neurologists reaching diagnoses [130].

It was a landmark paper by Gorry in 1973, however, that first critically analyzed conventional approaches to computer-based clinical decision making and outlined his motivation for turning to newer symbolic techniques [34]. He used the acute renal failure program discussed in Section 7.2 [33] as an example of the problems arising when decision analysis is used alone. In particular, he analyzed some of the cases on which the program had failed but the physicians considering the cases had performed well. His conclusions from these observations include the following four points.

(1) Clinical judgment is based less on detailed knowledge of pathophysiology than it is on gross chunks of knowledge and a good deal of detailed experience from which rules of thumb are derived.

(2) Clinicians know facts, of course, but their knowledge is also largely judgmental. The rules they learn allow them to focus attention and generate hypotheses quickly. Such heuristics permit them to avoid detailed search through the entire problem space.

(3) Clinicians recognize levels of belief or certainty associated with many of the rules they use, but they do not routinely quantitate or use these certainty concepts in any formal statistical manner.

(4) It is easier for experts to state their rules in response to perceived misconceptions in others than it is for them to generate such decision criteria a priori.

In the renal failure program medical knowledge had been embedded in the structure of the decision tree. This knowledge was never explicit, and additions to the experts' judgmental rules had generally required changes to the tree itself.

Based on observations such as those above, Gorry identified at least three important problems for investigation:

(1) Medical Concepts. Clinical decision aids had traditionally had no true "understanding" of medicine. Although explicit decision trees had given the decision theory programs a greater sense of the pertinent associations, medical knowledge and the heuristics for problem solving in the field had never been explicitly represented nor used. So-called "common sense" was often clearly lacking when the programs failed, and this was often what most alienated potential physician users.

(2) Conversational Capabilities. Both for capturing knowledge from collaborating experts, and for communicating with physician users, Gorry argued that further research on the development of computer-based linguistic capabilities was crucial.

(3) Explanation. Diagnostic programs had seldom emphasized an ability to explain the basis for their decisions in terms understandable to the physician. System acceptability was therefore inevitably limited; the physician would often have no basis for deciding whether to accept the program's advice, and might therefore resent what could be perceived as an attempt to dictate the practice of medicine.

Gorry's group at MIT and Tufts developed new approaches to examining the renal failure problem in light of these observations [75].

Due to the limitations of the older techniques, it was perhaps inevitable that some medical researchers would turn to the AI field for new techniques. Major research areas in AI include knowledge representation, heuristic search, natural language understanding and generation, and models of thought processes -- all topics clearly pertinent to the problems we have been discussing. Furthermore, AI researchers were beginning to look for applications to which they could apply some of the techniques they had developed in theoretical

domains. This community of researchers has grown in recent years, and a recent issue of Artificial Intelligence was devoted entirely to applications of AI to biology, medicine, and chemistry [105][25].

Among the programs using symbolic reasoning techniques are several systems that have been particularly novel and successful. At the University of Pittsburgh, Pople and Myers have developed a system called INTERNIST that assists with test selection for the diagnosis of all diseases in internal medicine [81]. This awesome task has been remarkably successful to date, with the program correctly diagnosing a large percentage of complex cases selected from clinical pathologic conferences in the major medical journals[26]. The program uses a hierarchic disease categorization, an ad hoc scoring system for quantifying symptom-disease relationships, plus some clever heuristics for focusing attention, discriminating between competing hypotheses, and diagnosing concurrent diseases [82]. The system currently has an inadequate human interface, however, and is not yet implemented for clinical trials.

Weiss, Kulikowski, and Amarel (Rutgers University) and Safir (Mt. Sinai Hospital, New York City) have developed a model of reasoning regarding disease processes in the eye, specifically glaucoma [125]. In this specialized application area it has been possible to map relationships between observations, pathophysiologic states, and disease categories. The resulting causal associational network (termed CASNET) forms the basis for a reasoning program that gives advice regarding disease states in glaucoma patients and generates management recommendations. The system is undergoing evaluation by a nationwide network of ophtholomologists but is not yet offered for routine clinical use.

For the AI researchers the question of how best to manage uncertainty in medical reasoning remains a central issue. The programs mentioned have developed ad hoc weighting systems and avoided formal statistical approaches. Others have turned to the work of statisticians and philosophers of science who have devised theories of approximate or inexact reasoning. For example, Wechsler [122] describes a program that is based upon Zadeh's fuzzy set theory [133], and Shortliffe and Buchanan [101] have turned to confirmation theory for their model of inexact reasoning.

---

[25]Many of the systems which use AI techniques for medical decision making were developed on the SUMEX-AIM computing resource, a nationally shared system devoted entirely to applications of AI to the biomedical sciences. The SUMEX-AIM computer is physically located at Stanford University but is used by researchers nationwide via connections to computer networks. The resource is funded by the Division of Research Resources, Biotechnology Branch, National Institutes of Health.

[26]Data communicated by Drs. Pople and Myers at the Fourth Annual A.I.M. Workshop, Rutgers University, June 1978.

8.2 Example

The symbolic reasoning program selected for discussion is the MYCIN System at Stanford University [102]. The researchers cited a variety of design considerations which motivated the selection of AI techniques for the consultation system they were developing [99]. They primarily wanted it to be useful to physicians and therefore emphasized the selection of a problem domain in which physicians had been shown to err frequently, namely the selection of antibiotics for patients with infections. They also cited human issues that they felt were crucial to make the system acceptable to physicians:

(1) it should be able to explain its decisions in terms of a line of reasoning that a physician can understand;
(2) it should be able to justify its performance by responding to questions expressed in simple English;
(3) it should be able to "learn" new information rapidly by interacting directly with experts;
(4) its knowledge should be easily modifiable so that perceived errors can be corrected rapidly before they recur in another case; and
(5) the interaction should be engineered with the user in mind (in terms of prompts, answers, and information volunteered by the system as well as by the users).

All these design goals were based on the observation that previous computer decision aids had generally been poorly accepted by physicians, even when they were shown to perform well on the tasks for which they were designed. MYCIN's developers felt that barriers to acceptance were largely conceptual and could be counteracted in large part if a system were perceived as a clinical tool rather than a dogmatic replacement for the primary physician's own reasoning.

Knowledge of infectious diseases is represented in MYCIN as production rules, each containing a "packet" of knowledge obtained from collaborating experts [102][27]. A production rule is simply a conditional statement which relates observations to associated inferences that may be drawn. For example, a MYCIN rule might state that "if a bacterium is a gram positive coccus growing in chains, then it is apt to be a streptococcus." MYCIN's power is derived from such rules in a variety of ways:

---

[27] Production rules are a technique frequently employed in AI research [9] and effectively applied to other scientific problem domains [6].

(1) it is the program that determines which rules to use and how they should be chained together to make decisions about a specific case[28];

(2) the rules can be stored in a machine-readable format but translated into English for display to physicians;

(3) by removing, altering, or adding rules, the system's knowledge structures can be rapidly modified without explicitly restructuring the entire knowledge base; and

(4) the rules themselves can often form a coherent explanation of system reasoning if the relevant ones are translated into English and displayed in response to a user's question.

Associated with all rules and inferences are numerical weights reflecting the degree of certainty associated with them. These numbers, termed certainty factors, form the basis for the system's inexact reasoning [101]. They allow the judgmental knowledge of experts to be captured in rule form and then used in a consistent fashion.

The MYCIN System has been evaluated regarding its performance at therapy selection for patients with either septicemia [132] or meningitis [131]. The program performs comparably with experts in these two task domains, but as yet it has no rules regarding the other infectious disease problem areas. Further knowledge base development will therefore be required before MYCIN is made available for clinical use; hence questions regarding its acceptability to physicians cannot yet be assessed. However, the required implementation stages have been delineated [100], attention has been paid to all the design criteria mentioned above, and the program does have a powerful explanation capability [95].

8.3 Discussion of the Methodology

Whereas the computations used by the other paradigms mostly involve straightforward application of well-developed computing techniques, artificial intelligence methods are largely experimental; new approaches to knowledge representation, language understanding, heuristic search, and the other symbolic reasoning problems we have mentioned are still needed. Thus the AI programs tend to be developed in research environments where short term practical results are unlikely to be found. However, out of this research are emerging techniques

---

[28] The control structure used is termed "goal-oriented" and is similar to the consequent-theorems used in Hewitt's PLANNER [42].

for coping with many of the problems encountered by the other paradigms we have discussed. AI researchers have developed promising methods for handling concurrent diseases [82],[125], assessing the time course of disease [18], and acquiring adequate structured knowledge from experts [11]. Furthermore, inexact reasoning techniques have been developed and implemented [101] (although they tend to be justified largely on intuitive grounds). In addition, the techniques of artificial intelligence provide a way to respond to many of Gorry's observations regarding the three major inadequacies of prior paradigms as described in Section 8.1: (1) the medical AI programs all tend to stress the representation of medical knowledge and a sense of understanding the underlying concepts; (2) many of them have conversational capabilities which draw on language processing research; and (3) explanation capabilities have been a primary focus of systems such as MYCIN.

Szolovits and Pauker have recently reviewed some applications of AI to medicine and have attempted to weigh the successes of this young field against the very real problems that lie ahead [108]. They identify several deficiencies of current systems. For example, termination criteria are still poorly understood. Although INTERNIST can diagnose simultaneous diseases, it also pursues all abnormal findings to completion, even though a clinician often ignores minor unexplained abnormalities if the rest of a patient's clinical status is well understood. In addition, although some of these programs now cleverly mimic the reasoning styles observed in experts [17],[53], it is less clear how to keep the systems from abandoning one hypothesis and turning to another one as soon as new information suggests another possibility. Programs that operate this way appear to digress from one topic to another -- a characteristic that decidedly alienates a user regardless of the validity of the final diagnosis or advice.

Still largely untapped is the power of an AI program to understand its own knowledge base, i.e., the structure and content of the reasoning mechanisms as well as of the medical facts. In effect, AI programs have the ability to "know what they know", the best working example of which can be found in the prototype system named Teiresias [10]. Because such programs can reason about their own knowledge, they have the power to encode knowledge about strategies, e.g., when to use and when to ignore specific items of medical knowledge and which leads to follow up on. Such "meta-level" knowledge offers a new dimension to the design of "intelligent assistant" programs which we predict will be exploited in medical decision making systems of the future.

9 Conclusions

This review has shown that there are two recurring questions regarding computer-based clinical decision making:

(1) Performance: How can we design systems that reach better, more reliable decisions in a broad range of applications, and

(2) Acceptability: How can we more effectively encourage the use of such systems by physicians or other intended users?

We shall summarize these points separately by reviewing many of the issues common to all the paradigms discussed in this paper.

9.1 Performance Issues

Central to assuring a program's adequate performance is a matching of the most appropriate technique with the problem domain. We have seen that the structured logic of clinical algorithms can be effectively applied to triage functions and other primary care problems, but they would be less naturally matched with complex tasks such as the diagnosis and management of acute renal failure. Good statistical data may support an effective Bayesian program in settings where diagnostic categories are small in number, non-overlapping, and well-defined, but the inability to use qualitative medical knowledge limits the effectiveness of the Bayesian approach in more difficult patient management or diagnostic environments. Similarly, mathematical models may support decision making in certain well-described fields in which observations are typically quantified, and related by functional expressions, but in which the knowledge is typically limited to numerical encoding. These examples, and others, demonstrate the need for thoughtful consideration of the technique most appropriate for managing a clinical problem. In general the simplest effective approach is to be preferred[29], but acceptability issues must also be considered as discussed below.

As researchers have ventured into more complex clinical domains, a number of difficult problems have tended to degrade the quality of performance of computer-based decision aids. Significant clinical problems require large knowledge bases that contain complex interrelationships including time and

---

[29] It is also always appropriate to ask whether computer-based approaches are needed at all for a given decision making task. For all but the most complex clinical algorithms, for example, the developers have tended to discard computer programs. Similarly, Schwartz et al. pointed out that decision analyses can often be successfully accomplished in a qualitative manner using paper and pencil [94].

functional dependencies. The knowledge of such domains is inevitably open-ended and incomplete, so the knowledge base must be easily extensible. Not only does this require a flexible representation of knowledge, but it encourages the development of novel techniques for the acquisition and integration of new facts and judgments. Similarly, the inexactness of medical inference must somehow be represented and manipulated within effective consultation systems. As we have discussed, all these performance issues are important knowledge engineering research problems for which artificial intelligence already offers promising new methods.

It is also important to consider the extent to which a program's "understanding" of its task domain will heighten its performance, particularly in settings where knowledge of the field tends to be highly judgmental and poorly quantified. We use the term "understanding" here to refer to a program's ability to reason about, as well as reason with, its medical knowledge base. This implies a substantial amount of judgmental or structural knowledge (in addition to data) contained within the program. Analyses of human clinical decision making [17],[53] suggest that as decisions move from simple to complex, a physician's reasoning style becomes less algorithmic and more heuristic, with qualitative judgmental knowledge and the conditions for invoking it coming increasingly into play. Furthermore, the performance of complex decision aids will also be heightened by the representation and utilization of high level "meta-knowledge" that permits programs to understand their own limitations and reasoning strategies. In order to design medical computing programs with these capabilities, the designers themselves will have to become cognizant of "knowledge engineering" issues. It is especially important that they find effective ways to match the knowledge structures they use to the complexity of the tasks their programs are designed to undertake.

9.2 Acceptability Issues

A recurring observation as one reviews the literature of computer-based medical decision making is that essentially none of the systems has been effectively used outside of a research environment, <u>even when its performance has been shown to be excellent</u>! This suggests that it is an error to concentrate research primarily on methods for improving the computer's decision making performance when clinical impact depends on solving other problems of acceptance as well. There are some data [106] to support the extreme view that

the biases of medical personnel against computers are so strong that systems will inevitably be rejected, regardless of performance. However, we are beginning to see examples of applications in which initial resistance to automated techniques has gradually been overcome through the incorporation of adequate system benefits [121].

Perhaps one of the most revealing lessons on this subject is an observation regarding the system of Mesel et al. [70] described in Section 2.2. Despite documented physician resistance to clinical algorithms in other settings [38], the physicians in Mesel's study accepted the guidance of protocols for the management of chemotherapy in their cancer patients. It is likely that the key to acceptance in this instance is the fact that these physicians had previously had no choice but to refer their patients with cancer to the tertiary care center in Birmingham where all complex chemotherapy was administered. The introduction of the protocols permitted these physicians to undertake tasks <u>that they had previously been unable to do</u>. It simultaneously allowed maintenance of close doctor-patient relationships and helped the patients avoid frequent long trips to the center. The motivation for the physician to use the system is clear in this case. It is reminiscent of Rosati's assertion that physicians will first welcome computer decision aids when they become aware that colleagues who <u>are</u> using them have a clear advantage in their practice [87].

A heightened awareness of "human engineering" issues among medical computing researchers will also make computers more acceptable to physicians by making the programs easier and more pleasant to use. Fox has recently reviewed this field in detail [22]. The issues range from the mechanics of interaction with the computer (e.g., using display terminals with such features as light pens, special keyboards, color, and graphics) to the features of the program that make it appear as a helpful tool rather than a complicating burden. Also involved, from both the mechanical and global design sides, is the development of flexible interfaces that tailor the style of the interaction to the needs and desires of individual physicians.

Adequate attention must also be given to the severe time constraints perceived by physicians. Ideally they would like programs to take no more time than they currently spend when accomplishing the same task on their own. Time and schedule pressures are similarly likely to explain the greater resistance to automation among interns and residents than among medical students or practicing physicians in Startsman's study [106].

The issue of a program's "self-knowledge" impacts on the acceptance of consultation systems in much the same way as it does upon program performance. Decision makers in general, and physicians in particular, will place more trust in systems that appear to understand their own limitations and capabilities, and that know when to admit ignorance of a problem area or inability to support any conclusion regarding an individual patient. Moreover, physicians will have a means for checking up on these automated assistants if the programs have an ability to explain not only the reasoning chain leading to their decisions but also their problem solving strategies. High-level knowledge, including a sense of scope and limitations, may thus allow a program to know enough about itself to prevent its own misuse. Furthermore, since systems that are not easily modifiable tend not to be accepted, meta-level knowledge about representation and interconnections within the knowledge base may help overcome the problem of programs becoming tied too closely to a store of knowledge that is regionally or temporally specific. It is therefore important to stress that considerations such as those we have mentioned here may argue in favor of using symbolic reasoning techniques even when a somewhat less complex approach might have been adequate for the decision task itself.

9.3 Summary

In summary, the trend towards increased use of knowledge engineering techniques for clinical decision programs stems from the dual goals of improving the performance and increasing the acceptance of such systems. Both acceptability and performance issues must be considered from the outset in a system's design because they dictate the choice of methodology as much as the task domain itself does. As greater experience is gained with these techniques, and as they become better known throughout the medical computing community, it is likely that we will see increasingly powerful unions between symbolic reasoning and the alternate paradigms we have discussed. One lesson to be drawn lies in the recognition that much basic research remains to be done in medical computing, and that the field is more than the application of established computing techniques to medical problems.

Acknowledgments

We wish to thank R. Blum, L. Fagan, J. King, J. Kunz, H. Sox, and G. Wiederhold for their thoughtful advice in reviewing earlier drafts of this paper. We are also grateful to Dr. Herbert Sherman and the reviewers for their constructive suggestions regarding revisions.

References

1. Armitage, P. and Gehan, E.A. "Statistical methods for the identification and use of prognostic factors." Int. J. Cancer, 13, pp. 16-36, (1974).

2. Bleich, H.L. "Computer evaluation of acid-base disorders." J. Clin. Invest. 48, pp. 1689-1696 (1969).

3. Bleich, H.L. "The computer as a consultant." N. Eng. J. Med. 284, pp. 141-147 (1971).

4. Bleich, H.L. "Computer-based consultation: electrolyte and acid-base disorders." Amer. J. Med. 53, pp. 285-291 (1972).

5. Blum, R.L. and Wiederhold, G. "Inferring knowledge from clinical data banks: utilizing techniques from artificial intelligence," Proc. 2nd Ann. Symp. on Comput. Applic. in Med. Care, IEEE, Washington D.C., November 1978, pp. 303-307.

6. Buchanan, B.G. and Feigenbaum, E.A. "Dendral and Meta-Dendral: their applications dimension." Artificial Intelligence 11, pp. 5-24 (1978).

7. Croft, D.J. "Is computerized diagnosis possible?" Comput. Biomed. Res. 5, pp. 351-367 (1972).

8. Cumberpatch, J. and Heaps, H.S. "A disease-conscious method for sequential diagnosis by use of disease probabilities without assumption of symptom independence." Int. J. Biomed. Comput. 7, pp. 61-78 (1976).

9. Davis, R. and King, J. "An overview of production systems." In Machine Representation of Knowledge (E.W. Elcock and D. Michie, eds.), New York: Wiley, 1976.

10. Davis, R. Applications Of Meta-Level Knowledge To The Construction, Maintenance, And Use Of Large Knowledge Bases, Heuristic Programming Project, Stanford University, Memo HPP-76-7, July 1976.

11. Davis, R. "Interactive transfer of expertise: acquisition of new inference rules." Proc. of 5th Intl. Joint Conf. on Artif. Intelligence, Cambridge, Mass, 1977.

12. deDombal, F.T., Leaper, D.J., Staniland, J.R., et al. "Computer-aided diagnosis of acute abdominal pain." Brit. Med. J. 2, pp. 9-13 (1972).

13. deDombal, F.T., Leaper, D.J., Horrocks, J.C., et al. "Human and computer-aided diagnosis of abdominal pain: further report with emphasis on performance of clinicians." Brit. Med. J. 1, pp. 376-380 (1974).

14. deDombal, F.T. and Gremy, F. (eds.) Decision Making And Medical Care: Can Information Science Help?, Amsterdam: North-Holland Publishing Co., 1976.

15. Duda, R.O. and Hart, P.E. Pattern Classification and Scene Analysis. New York: Wiley, 1973.

16. Edwards, W. "N=1: diagnosis in unique cases." In Computer Diagnosis And Diagnostic Methods, (J.A. Jacquez, ed.), Springfield, Ill.: Charles C. Thomas, 1972, pp. 139-151.

17. Elstein, A.S., Shulman, L.S., and Sprafka, S.A. Medical Problem Solving: An Analysis of Clinical Reasoning. Cambridge, Mass.: Harvard Univ. Press, 1978.

18. Fagan, L.M. Knowledge Engineering For Dynamic Clinical Settings: Giving Advice In The Intensive Care Unit. Heuristic Programming Project, Stanford University (Doctoral Dissertation), 1979.

19. Feigenbaum, E.A. "The art of artificial intelligence: themes and case studies of knowledge engineering." AFIPS Conference Proc., NCC 1978. Vol. 47. Montvale, N.J.: AFIPS Press, 1978, p. 227.

20. Feinstein, A.R. "Quality of data in the medical record." Comput. Biomed. Res. 3, pp. 426-435 (1970).

21. Feinstein, A.R., Rubinstein, J.F., and Ramshaw, W.A. "Estimating prognosis with the aid of a conversational mode computer program." Anns. Int. Med. 76, pp. 911-921 (1972).

22. Fox, J. "Medical computing and the user." Int. J. Man-Machine Studies 9, pp. 669-686 (1977).

23. Friedman, R.B. and Gustafson, D.H. "Computers in clinical medicine: a critical review." Comput. Biomed. Res. 8, pp. 199-204 (1977).

24. Fries, J.F. "Time-oriented patient records and a computer databank." J. Amer. Med. Assoc. 222, pp. 1536-1542 (1972).

25. Fries, J.F. "A data bank for the clinician?" (editorial). N. Eng. J. Med. 294, pp. 1400-1402 (1976).

26. Garland, L.H. "Studies on the accuracy of diagnostic procedures." Amer. J. Roentgen. 82, pp. 25-38 (1959).

27. Gill, P.W., Leaper, D.J., Guillou, P.J., et al. "Observer variation in clinical diagnosis: a computer-aided assessment of its magnitude and importance." Meth. Inform. Med.. 12, pp. 108-113 (1973).

28. Ginsberg, A.S. Decision Analysis in Clinical Patient Management With an Application to the Pleural Effusion Syndrome. The Rand Corporation, R-751-RC/NLM, July 1971.

29. Ginsberg, A.S. "The diagnostic process viewed as a decision problem." In Computer Diagnosis and Diagnostic Methods, (J.A. Jacquez, ed.), Springfield, Ill.: Charles C. Thomas, 1972.

30. Gleser, M.A. and Collen, M.F. "Towards automated medical decisions." Comput. Biomed. Res. 5, pp. 180-189 (1972).

31. Goldwyn, R.M., Friedman, H.P., and Siegel, J.H. "Iteration and

interaction in computer data bank analysis: as case study in the physiologic classification and assessment of the critically ill." Comput. Biomed. Res. 4, pp. 607-622 (1971).

32. Gorry, G.A. and Barnett, G.O. "Experience with a model of sequential diagnosis." Comput. Biomed. Res. 1, pp. 490-507 (1968).

33. Gorry, G.A., Kassirer, J.P., Essig, A., and Schwartz, W.B. "Decision analysis as the basis for computer-aided management of acute renal failure." Amer. J. Med 55, pp. 473-484 (1973).

34. Gorry, G.A. "Computer-assisted clinical decision making." Meth. Inform. Med. 12, pp. 45-51 (1973).

35. Gorry, G.A., Silverman, H., and Pauker, S.G. "Capturing clinical expertise: a computer program that considers clinical responses to digitalis." Amer. J. Med 64, pp. 452-460 (1978).

36. Greenes, R.A., Barnett, G.O., Klein, S.W., et al. "Recording, retrieval, and review of medical data by physician-computer interaction." N. Eng. J. Med. 282, pp. 307-315 (1970).

37. Greenfield, S., Komaroff, A.L., and Anderson, H. "A headache protocol for nurses: effectiveness and efficiency." Arch. Intern. Med. 136, pp. 1111-1116 (1976).

38. Grimm, R.H., Shimoni, K., Harlan, W.R., and Estes, E.H. "Evaluation of patient-care protocol use by various providers." N. Eng. J. Med. 292, pp. 507-511 (1975).

39. Groner, G.F., Clark, R.L., Berman, R.A., and De Land, E.C. "BIOMOD - an interactive computer graphics system for modeling." Proc. Fall Joint Computer Conference, pp 369-378, 1971.

40. Groth, T. "Biomedical modelling." In Medinfo-77, Amsterdam: North-Holland Publishing Co., 1977, pp. 775-784.

41. Hess, E.V. "A uniform database for rheumatic diseases." Arthritis and Rheumatism 19, pp. 645-648 (1976).

42. Hewitt, C. Description and Theoretical Analysis (Using Schemata) of PLANNER: A Language for Proving Theorems and Manipulating Models in a Robot. Ph.D. Dissertation, Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Mass., 1972.

43. Horrocks, J.C., McCann, A.P., Staniland, J.R., et al. "Computer-aided diagnosis: description of an adaptable system, and operational experience with 2,034 cases." Brit. Med. J. 2, pp. 5-9 (1972).

44. Horrocks, J.C., and deDombal, F.T. "Computer-aided diagnosis of dyspepsia." Amer. J. Diges. Dis. 20:397-406 (1975).

45. Howard, R. A. (ed.). "Special Issue on Decision Analysis." IEEE Transactions on Systems, Science and Cybernetics, Vol SSC-4(3), September 1968.

46. Inglefinger, F.J. "Decision in medicine" (editorial). N. Eng. J. Med. 293, pp. 254-255 (1975).

47. Jacquez, J.A. Computer Diagnosis and Diagnostic Methods, Springfield, Ill.: Charles C. Thomas, 1972.

48. Jelliffe, R.W., Buell, J., Kalaba, R., et al. "A computer program for digitalis dosage regimens." Math. Biosci. 9, pp. 179-193 (1970).

49. Jelliffe, R.W., Buell, J., and Kalaba, R. "Reduction of digitalis toxicity by computer-assisted glycoside dosage regimens." Anns. Int. Med. 77, pp. 891-906 (1972).

50. Johnson, D.C. and Barnett, G.O. "MEDINFO - a medical information system." Comput. Prog. in Biomed. 7, pp. 191-201 (1977).

51. Kanal, L.N. "Patterns in Pattern Recognition: 1968-1974," IEEE Trans. on Information Theory, vol. IT-20, no. 6 (1974).

52. Karpinski, R.H.S. and Bleich, H.L. "MISAR: a miniature information storage and retrieval system." Comput. Biomed. Res. 4, pp. 655-660 (1971).

53. Kassirer, J.P. and Gorry, G.A. "Clinical problem solving: a behavioral analysis." Anns. Int. Med. 89, pp. 245-255 (1978).

54. Kleinmuntz, B. and McLean, R.S. "Diagnostic interviewing by digital computer." Behav. Sci. 13, pp. 75-80 (1968).

55. Knapp, R.G., Levi, S., Lurie, D., and Westphal, M. "A computer-generated diagnostic decision guide: a comparison of statistical diagnosis and clinical diagnosis." Comput. Biol. Med. 7, pp. 223-230 (1977).

56. Komaroff, A.L., Black, W.L., Flatley, M., et al. "Protocols for physician assistants: management of diabetes and hypertension." N. Eng. J. Med. 290:307-312 (1974).

57. Komaroff, A.L. "Medical data collection: hard decisions from soft data." IEEE Proceedings, This issue, (1979).

58. Korein, J., Lyman, M., and Tick, J.L. "The computerized medical record," Bulletin New York Academy of Medicine, 47, pp. 824-826 (1971).

59. Koss, N. and Feinstein, A.R. "Computer-aided prognosis: II. development of a prognostic algorithm." Arch. Intern. Med. 127, pp. 448-459 (1971).

60. Leaper, D.J., Horrocks, J.C., Staniland, J.R., and deDombal, F.T. "Computer-assisted diagnosis of abdominal pain using 'estimates' provided by clinicians." Brit. Med. J. 4, pp. 350-354 (1972).

61. Ledley, R.S. and Lusted, L.B. "Reasoning foundations of medical diagnosis." Science 130, pp. 9-21 (1959).

62. Levi, S., Frant, J.R., Westphal, M.C., and Lurie, D. "Development of a decision guide - optimal discriminations for meningitis determined by statistical analysis." Meth. Inform. Med. 15, pp. 87-90 (1976).

63. Lipkin, M. and Hardy, J.D. "Mechanical correlation of data in differential diagnosis of hematologic diseases." J. Amer. Med. Assoc. 166, pp. 113-125 (1958).

64. Lusted, L.B. Introduction To Medical Decision Making. Springfield, Ill.: Charles C. Thomas, 1968.

65. Mabry, J.C., Thompson, H.K., Hopwood, M.D., and Baker, W.R. "A prototype data management and analysis system (CLINFO): system description and user experience." In MEDINFO 77, Amsterdam: North-Holland Publishing Co., 1977, pp. 71-75.

66. McDonald, C., Bhargava, B., and Jeris, D. "A clinical information system (CIS) for ambulatory care," Proc. of the 1975 NCC, AFIPS Press, vol. 44 (1975) pp. 749-756

67. McNeil, B.J., Keeler, E., and Adelstein, S.J. "Primer on certain elements of medical decision making." N. Eng. J. Med. 293, pp. 211-215 (1975).

68. McNeil, B.J. and Adelstein, S.J. "Determining the value of diagnostic and screening tests." J. Nucl. Med. 17, pp. 439-448 (1977).

69. Menn, S.J., Barnett, G.O., Schmechel, D., et al. "A computer program to assist in the care of acute respiratory failure." J. Amer. Med. Assoc. 223, pp. 308-312 (1973).

70. Mesel, E., Wirtschafter, D.D., Carpenter, J.T., et al. "Clinical algorithms for cancer chemotherapy - systems for community-based consultant-extenders and oncology centers." Meth. Inform. Med. 15, pp. 168-173 (1976).

71. Nordyke, R.A., Kulikowski, C.A., and Kulikowski, C.W. "A comparison of methods for the automated diagnosis of thyroid dysfunction." Comput. Biomed. Res. 4, pp. 374-389 (1971).

72. Norusis, M.J. and Jacquez, J.A. "Diagnosis. I. Symptom nonindependence in mathematical models for diagnosis." Comput. Biomed. Res. 8, pp. 156-172 (1975).

73. Patrick, E.A. "Pattern Recognition in Medicine," Systems, Man and Cybernetics Review, 6, p. 4 (1977).

74. Pauker, S.G. and Kassirer, J.P. "Therapeutic decision making: a cost-benefit analysis." N. Eng. J. Med. 293, pp. 229-234 (1975).

75. Pauker, S.G., Gorry, G.A., Kassirer, J.P., and Schwartz, W.B. "Towards the simulation of clinical cognition: taking a present illness by computer." Amer. J. Med. 60, pp. 981-996 (1976).

76. Pauker, S.G. "Coronary artery surgery: the use of decision analysis." Anns. Int. Med. 85, pp. 8-18 (1976).

77. Pauker, S.P. and Pauker, S.G. "Prenatal diagnosis: a directive approach to genetic counseling using decision analysis." Yale J. Biol. Med. 50, pp. 275-289 (1977).

78. Peck, C.C., Sheiner, L.B., Martin, C.M., et al. "Computer-assisted digoxin therapy." N. Eng. J. Med. 289, pp. 441-446 (1973).

79. Pipberger, H.V. "Clinical application of a second generation electrocardiography computer program." Amer. J. Electrocardiology 35, pp. 597-608 (1975).

80. Pliskin, J.S. and Beck, C.H. "Decision analysis in individual clinical decision making: a real-world application in treatment of renal disease." Meth. Inform. Med. 15, pp. 43-46 (1976).

81. Pople, H.E., Myers, J.D., and Miller, R.A. "DIALOG: A model of diagnostic logic for internal medicine." Proc. 4th Int. Joint. Conf. on Artif. Intell., MIT, Cambridge, Mass., 1975.

82. Pople, H. "The formation of composite hypotheses in diagnostic problem solving: an exercise in synthetic reasoning." Proc. of 5th Intl. Joint Conf. on Artif. Intelligence, Cambridge, Mass, 1977, pp. 1030-1037.

83. Prutting, J. "Lack of correlation between antemortem and postmortem diagnosis." N.Y. J. Med. 67, pp. 2081-2084 (1967).

84. Raiffa, H. Decision Analysis: Introductory Lectures on Choices Under Uncertainty. Reading, Mass.: Addison Wesley, 1968.

85. Richards, B. and Goh, A.E.S. "Computer assistance in the treatment of patients with acid-base and electrolyte disturbances." MEDINFO 77, Amsterdam: North-Holland Publishing Company, 1977, pp. 407-410.

86. Rodnick, J., and Wiederhold, G. , "Review of automated ambulatory medical record systems: charting services that are of essential benefit to the physician," MEDINFO 77, Amsterdam: North-Holland Publishing Co., 1977, pp. 957-961.

87. Rosati, R.A., Wallace, A.G., and Stead, E.A. "The way of the future." Arch. Intern. Med. 131, pp. 285-287 (1973).

88. Rosati, R.D., McNeer, J.F., Starmer, C.F., et al. "A new information system for medical practice." Arch. Intern. Med. 135, pp. 1017-1024 (1975).

89. Rosenblatt, M.B., Teng, P.K., and Kerpe, S. "Diagnostic accuracy in cancer as determined by post-mortem examination." Prog. Clin. Cancer 5, pp. 71-80 (1973).

90. Rubin, A.D. and Risley, J.F. "The PROPHET system: an experiment in providing a computer resource to scientists." MEDINFO 77, Amsterdam: North-Holland Publishing Co., 1977, pp. 77-81.

91. Safran, C., Tsichlis, P.N., Bluming, A.Z., and Desforges, J.F. "Diagnostic planning using computer-assisted decision making for patients with Hodgkins' disease." Cancer 39, pp. 2426-2434 (1977).

92. Schoolman, H. and Bernstein, L. "Computer use in diagnosis, prognosis, and therapy." Science 200, pp. 926-931 (1978).

93. Schwartz, W.B. "Medicine and the computer: the promise and problems of change." N. Eng. J. Med. 283, pp. 1257-1264 (1970).

94. Schwartz, W.B., Gorry, G.A., Kassirer, J.P., and Essig, A. "Decision analysis and clinical judgment." Amer. J. Med 55, pp. 459-472 (1973).

95. Scott, A.C., Clancey, W., Davis, R., and Shortliffe, E.H. "Explanation capabilities of knowledge-based production systems." Amer. J. Computational Linguistics, Microfiche 62, 1977.

96. Sheiner, L.B., Halkin, H., Peck, C., et al. "Improved computer-assisted digoxin therapy." Anns. Int. Med. 82, pp. 619-627 (1975).

97. Sherman, H., Reiffen, B., and Komaroff, A.L. "Ambulatory care systems." In Problem-Directed and Medical Information Systems (M.F. Driggs, ed.), New York: Intercontinental Medical Book Corporation, 1973, pp. 143-171.

98. Shimura, M. "Learning procedures in pattern classifiers - introduction and survey." Proc. Intl. Joint Conf. on Pattern Recognition, Kyoto, 1978, pp. 125-138.

99. Shortliffe, E.H., Axline, S.G., Buchanan, B.G., and Cohen, S.N. "Design considerations for a program to provide consultations in clinical therapeutics." Proc. 13th San Diego Biomedical Symposium, pp. 311-319, San Diego, Calif., February 1974.

100. Shortliffe, E.H. and Davis, R. "Some considerations for the implementation of knowledge-based expert systems." SIGART Newsletter, No. 55, 9-12, December 1975.

101. Shortliffe, E.H., and Buchanan, B.G. "A model of inexact reasoning in medicine." Math. Biosci. 23, pp. 351-379 (1975).

102. Shortliffe, E.H. Computer-Based Medical Consultations: MYCIN, New York: Elsevier/North Holland, 1976.

103. Slamecka, V., Camp, H.N., Badre, A.N., and Hall, W.D. "MARIS: a knowledge system for internal medicine." Inform. Process. & Manage. 13, pp. 273-276 (1977).

104. Sox, H.C., Sox, C.H., and Tompkins, R.K. "The training of physicians' assistants: the use of a clinical algorithm system." N. Eng. J. Med. 288, pp. 818-824 (1973).

105. Sridharan, N.S. Guest editorial. Artificial Intelligence 11, pp. 1-4 (1978).

106. Startsman, T.S. and Robinson, R.E. "The attitudes of medical and paramedical personnel towards computers." Comput. Biomed. Res. 5, pp. 218-227 (1972).

107. Stead, W.W., Brame, R.G., Hammond, W.E., et al. "A computerized obstetric medical record." Obstet. & Gyn. 49, pp. 502-509 (1977).

108. Szolovits, P. and Pauker, S.G. "Categorical and probabilistic reasoning in medical diagnosis." Artificial Intelligence 11, pp. 115-144 (1978).

109. Taylor, T.R. "Clinical decision analysis." Meth. Inform. Med. 15, pp. 216-224 (1976).

110. Vickery, D.M. "Computer support of paramedical personnel: the question of quality control." In MEDINFO 74, Amsterdam: North-Holland Publishing Company, 1974, pp. 281-287.

111. Vishnevskiy, A.A., Artobolevskiy, I.I., and Bykovskiy, M.L. Machine Diagnosis And Information Retrieval In Medicine In The USSR, DHEW Publication No. (NIH) 73-424, 1973.

112. Wagner, G., Tautu, P., and Wolber, U. "Problems of medical diagnosis: a bibliography." Meth. Info. Med. 17, pp. 55-74 (1978).

113. Walsh, B.T., Bookhein, W.W., Johnson, R.C., et al. "Recognition of streptococcal pharyngitis in adults." Arch. Int. Med. 135, pp. 1493-1497 (1975).

114. Wardle, A. and Wardle, L. "Computer-aided diagnosis: a review of research." Meth. Info. Med. 17, pp. 15-28 (1978).

115. Warner, H.R., Toronto, A.F., and Veasy, L.G. "Experience with Bayes' Theorem for computer diagnosis of congenital heart disease." Anns. N.Y. Acad. Sci. 115, pp. 558-567 (1964).

116. Warner, H.R. "Experiences with computer-based patient monitoring." Anes. & Analgesia Current Researchers 47, pp. 453-461 (1968).

117. Warner, H.R., Olmsted, C.M., and Rutherford, B.D. "HELP - a program for medical decision-making." Comput. Biomed. Res. 5, pp. 65-74 (1972).

118. Warner, H.R., Rutherford, B.D., and Houtchens, B. "A sequential approach to history taking and diagnosis." Comput. Biomed. Res. 5, pp. 256-262 (1972).

119. Warner, H.R., Morgan, J.D., Pryor, T.A., et al. "HELP - a self-improving system for medical decision making." In MEDINFO 74, Amsterdam: North-Holland Publishing Company, 1974.

120. Warner, H.R. "Knowledge sectors for logical processing of patient data in the HELP system." Proc. of 2nd Ann. Symp. on Computer Applications in Medical Care, IEEE, Wash. D.C., 1978, pp. 401-404.

121. Watson, R.J. "Medical staff response to a medical information system with direct physician-computer interface." In MEDINFO 74, Amsterdam: North-Holland Publishing Company, 1974, pp. 299-302.

122. Wechsler, H. "A fuzzy approach to medical diagnosis." Int. J. Biomed. Comput. 7, pp. 191-203 (1976).

123. Weed, L.L. "Medical records that guide and teach." N. Eng. J. Med. 278, pp. 593-599,652-657 (1968).

124. Weed, L.L. "Problem-oriented medical records." In Problem-Directed and Medical Information Systems (M.F. Driggs, ed.), New York: Intercontinental Medical Book Corporation, 1973.

125. Weiss, S.M., Kulikowski, C.A., Amarel, S. and Safir, A. "A model-based method for computer-aided medical decision-making." Artificial Intelligence 11, pp. 145-172 (1978).

126. Weyl, S., Fries, J., Wiederhold, G., and Germano, F. "A modular self-describing clinical databank system." Comput. Biomed. Res. 8, pp. 279-293 (1975).

127. Wiederhold, G., Fries, J.F., and Weyl, S. "Structured organization of clinical databases," Proc. of the 1975 NCC, AFIPS Press, Vol. 44, 1975, pp. 479-485.

128. Winston, P.H. Artificial Intelligence, Reading, Mass.: Addison-Wesley, 1977.

129. Wirschafter, D., Carpenter, J.T., and Mesel, E. "A consultant-extender system for breast cancer adjuvant chemotherapy." To appear in Anns. Int. Med. (1979).

130. Wortman, P.M. "Medical diagnosis: an information processing approach." Comput. Biomed. Res. 5, pp. 315-328 (1972).

131. Yu, V.L., Fagan, L.M., Wraith, S.M., et al. "Computer-based consultation in antimicrobial selection - a comparative evaluation by experts." Stanford University School of Medicine. Submitted for publication, December 1978.

132. Yu, V.L., Buchanan, B.G., Shortliffe, E.H., et al. "An evaluation of the performance of a computer-based consultant." Comput. Prog. Biomed. 9, pp. 95-102 (1979).

133. Zadeh, L.A. "Fuzzy sets." Information and Control 8, pp. 338-353 (1965).

134. Zoltie, N., Horrocks, J.C., and deDombal, F.T. "Computer-assisted diagnosis of dyspepsia - report on transferability of a system, with emphasis on early diagnosis of gastric cancer." Meth. Inform. Med. 16, pp. 89-92 (1977).