Report 83-32
Stanford -- KSL

Scientific DataLink

Comparison of Techniques of Computer-Assisted Decision Making in Medicine.
John C. Kunz, Edward H. Shortliffe, Bruce G. Buchanan, et al., Apr 1983

card 1 of 1

Heuristic Programming Project
Report No. HPP-83-32

April 1983

Comparison of Techniques of Computer-Assisted Decision Making in Medicine

John C. Kunz
Edward H. Shortliffe
Bruce G. Buchanan
Edward A. Feigenbaum

Heuristic Programming Project
Department of Computer Science
Stanford University

Memo HPP-83-32

Comparison of Techniques of Computer-Assisted Decision Making in Medicine

John C. Kunz*

Edward H. Shortliffe**

Bruce G. Buchanan

Edward A. Feigenbaum

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

Submitted for publication in the Pure and Applied Biostructure, World Press, Singapore (1983), Claudio Niccolini, Editor

April 1983

Reprinted, with revisions, by the authors, from PROCEEDINGS OF THE IEEE, Vol. 67, No. 9, pp. 1202-1224, September 1979 (C) 1979 IEEE.

*Current address: IntelliGenetics
124 University Avenue
Palo Alto, Ca. 94301

**Dr. Shortliffe is a Henry J. Kaiser Family Foundation Faculty Scholar in General Internal Medicine and recipient of research career development award LM00048 from the National Library of Medicine.

Abstract

This article reviews representative examples of five major paradigms of computer-assisted decision making (CADM) in medicine. These paradigms include 1) clinical algorithms, 2) statistical analysis of collections of patient data, 3) mathematical models of physical processes, 4) decision analysis, and 5) symbolic reasoning or artificial intelligence. We note that no one technique is best for all applications, and we review promising work which combines two or more established techniques. We emphasize both the inherent power of symbolic reasoning and the promise of artificial intelligence and the other techniques to complement each other.

1 INTRODUCTION

As early as the 1950's it was recognized that computers could conceivably assist with clinical decision making [40], and both physicians and computer scientists began to analyze medical diagnosis with a view to the potential role of automated decision aids in that domain [39]. A recent article in the medical literature suggests that "computers may never replace physicians, but they can and will assist their decisions" [81]. A variety of techniques have been applied to CADM, accounting for at least 800 references in the clinical and computing literature [68]. This paper revises an earlier review of computer-based medical decision making methodologies [60]. That review discussed important issues that account for both the multiplicity of approaches to the problem and the limited clinical success of most of the systems developed to date. That review also includes an extensive bibliography which is revised in this article. We focus on the representation and utilization of knowledge, termed "knowledge engineering". We discuss some inadequacies of data-intensive techniques which have led to recent interest in the symbolic reasoning approaches developed during the past. In addition, we review some research projects which have combined two or more techniques of computer-based decision making, discuss the way in which different techniques complement each other, and suggest that artificial intelligence may offer an accommodating paradigm under which appropriate techniques can be used for different problems.

1.1 Perspectives for comparing CADM systems

The models on which computer systems base their advice can be compared along the three dimensions of their inputs (i.e., use of data), processing (i.e., use of knowledge), and their outputs (including both form and content). Complexity of CADM system input can range from minimal data from a single patient to large amounts of data from many patients. Processing can be based on small amounts of domain-independent knowledge or relatively large amounts of domain-specific knowledge. CADM system output may range from the quantitative value of a single parameter to presentation of a ranked list of multiple possible diagnoses or therapy plans, with explanations of conclusions.

When based on relatively limited amounts of domain-independent knowledge, powerful decision-making systems typically use rich bases of domain-specific data. For example, statistical methods are often used to analyze the contents of large data banks in order to make descriptions of a class of patients or prognostic inferences about an individual patient. The output of such systems is typically a score or a quantitative value for a statistical probability. In contrast, some knowledge is relatively specific and applicable to a limited domain, such as the heuristic knowledge of an expert diagnostician. When based on large amounts of domain-specific knowledge, powerful decision-making systems typically use very small amounts of data. Artificial intelligence techniques have been used to make inferences based on large amounts of domain-specific knowledge; these systems can sometimes provide diagnoses and explanations of their decisions.

If there is a general chronology to the field over the last 20 years, it is that there has been progressive trend from emphasizing techniques based on general knowledge, such as statistical methods, toward emphasis of techniques based on domain-specific symbolic knowledge, such as diagnostic inference rules. This trend has resulted in an emphasis on systems which interpret and explain the clinical significance of their findings, rather than simply producing another number for the user to interpret.

We include with domain knowledge a 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 utilize the knowledge of experts in decision making programs. Symbolic processing allows explicit representation and manipulation of heuristic judgmental knowledge. Thus, symbolic processing approaches allow exploiting judgmental knowledge, and they support explanation of their judgements to provide credibility for decisions which are correct and insight into the problems with decisions which are incorrect.

Knowledge-based symbolic reasoning systems derive their power from making inferences which are consistent with, but not explicitly described in the knowledge base. Inference is the process of concluding a fact which was not explicitly described. Making inferences is a key issue in knowledge-based programing. Inferences can be made given appropriate knowledge, such as a diagnostic rule. Simple data, however, such as simple records of observations, provide no inherent basis for making inferences. Knowledge and data are closely related since inferences are made using both general knowledge of how to interpret data and specific data.

We use the term "knowledge engineering" to refer to the process of developing computer-based symbolic reasoning systems. Knowledge engineering concerns issues such as knowledge representation, acquisition, and explanation [16]. Knowledge engineering uses techniques of artificial intelligence which, for our purpose, can be described as the branch of computer science concerned with symbolic reasoning by computers. Approaches to the issues of inference and knowledge engineering characterize the distinctions between programs based on artificial intelligence and programs using conventional calculations. For example, symbolic processing programs 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 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 research groups to investigate the knowledge engineering approach to the development of CADM systems.

1.2 Overview Of This Paper

An exhaustive review of computer-aided diagnosis will not be attempted in light of the vast size of the field, and we have therefore chosen to review the methodologies by discussing several representative examples of systems that have been described. The principal examples we have selected are not necessarily the best nor the most successful; however, they illustrate the issues we wish to discuss and encompass most of the major methodologies that have been applied to computer-based medical decision making. Any attempt to categorize programs in this way is inherently fraught with problems in that several systems appropriately lay claim to more than one methodology. 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.

2 Clinical Algorithms and Automation

2.1 Overview

Clinical algorithms, or protocols, are structured decision making flowcharts to which a diagnostician or therapist can refer when deciding how to manage a patient with a specific clinical problem [57]. Like recipes, algorithms are simple, short, and relatively easy to use with limited prior training. Clinical algorithms take small amounts of input: a few selected observations of an individual patient. Their processing is simple flow-chart logic, based on highly distilled representation of the clinical knowledge of a domain. Their output is a decision: a suggested action. Algorithms provide no explanations per se, but the sequence of decisions used in arriving at a conclusions allows a user to understand the program logic.

Typically, clinical algorithms have been designed by expert physicians. Goldman recently reported on a large study which used statistical analysis of the contents of a data bank to generate an algorithm for managing patients who complained of chest pain [22]. Algorithms have been developed for use by physicians' assistants, nurse practitioners or other physicians to guide performance of specified routine clinical-care tasks. Grimm has found that algorithms are most readily accepted by novice and expert nurses and physicians, and they are less well-accepted by physician interns and nurse students [29]. He showed, however, that physician performance could improve when protocols were used in certain settings. 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 4-8 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 [42], pharyngitis [29], and headache [28]. Wirtschafter reports improved adherence to a protocol and fewer complications when physicians used an algorithm for chemotherapy treatment of Hodgkin's Disease [77]. Reporting that physician assistant users are more likely to follow the suggestions of a computer-based algorithm than those of a paper algorithm, Cannon suggests that use of paper algorithms may alienate users, while the use of computer-based algorithms computer elicits better compliance [9]. McDonald [43] hypothesizes that physicians have an acute problem with information overload, and that using algorithms can help to reduce the effects of this problem. He reports significant improvement in specific indices of the quality of care when physicians used algorithms. Further, he reports that the improvements tend to disappear quickly when physicians stop using the algorithms, indicating that the improvements are due to use of the algorithm and not due to educational effect.

2.2 Example

The cancer chemotherapy system developed in Alabama by Mesel et al. [45], [77] uses algorithms to allow private practitioners, at a distance from the regional tertiary-care center, to manage the complex chemotherapy for their cancer patients. The algorithms are popular with private physicians since they can effectively care for their patients without routinely referring them to oncologists at a tertiary care center. Mesel 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 been entered 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 an 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 if additional questions arose.

The authors present data suggesting that their system was well-accepted by physicians and patients, and that excellent care was delivered. This is an interesting result in light of Grimm's experience [29]. Perhaps 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 at all. Retrospective review of cases that were treated at the referral center, 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 utilized directly. Thus algorithms may be effective tools for the administration of complex specialized therapy in circumstances such as those described.

2.3 Discussion of the Methodology

Although clinical algorithms are among the most widespread and 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 or is not present), 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 complex 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 logic, and the supervising expert "escape valve", 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 many applications. 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 methodology provides a limited base for extension to the complex decision tasks to be discussed in the following sections.

3 Statistical Analysis of Collections of Patient Data

3.1 Overview

Making decisions based on analysis of patient records has been a major research concern since the earliest days of medical computing. The input to CADM systems based on statistical analysis includes either selective or exhaustive data for an individual patient, and it may include the raw or summarized data from a set of patients from one or more populations to which the individual may belong. In comparison with the other techniques discussed in this paper, statistical-based CADM systems typically use relatively large amounts of input data. The processing of these systems has been based on different analysis methods, including statistical analysis, multi-variable statistical pattern recognition [48], and Bayesian analysis [12]. These processing methods are uniformly domain-independent. The output of these systems is usually numeric: a probability or a statistical index, possibly including a statistical measure of confidence; if these systems contain only knowledge of statistical methods, then they have no ability to explain the credibility of their conclusions either in terms of clinical knowledge of disease or in terms of pathophysiology.

In each of these methodologies, the general approach is to use a domain-independent method to analyze the data of a population of patients. In each case, the investigator selects a group of patients which belong to a common population. The statistical properties of the population are then characterized. In the individual case, the inference is made that if a patient is a member of a population whose properties are known, then the patient has the properties of the population, with some statistical measure of confidence.

In addition, researchers have directed great effort into design, development and operation of computer-based databanks for gathering and maintaining appropriate and accurate patient records. For example, the system of Weed [72] has been instrumental in showing the importance of the concept of the problem-oriented medical record. The ARAMIS system has developed a data-management system for patient data [19] which is now used at several research centers around the country. The databank systems all depend on a complete and accurate medical record system. Accurate computerized medical records depend both on careful computer system design and on meticulous attention to entering accurate and complete records. If 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.

Pattern recognition techniques define the mathematical relationship between measurable features and classifications of objects [13]. 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 [36], 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 a diagnostic pattern, or discriminant function, the pattern matching method requires a "training set" of data from a population whose data are known accurately and for whom the correct classifications are already known. If the form and parameters are not known for the statistical distributions underlying the features, then they must be estimated. After training, then, the pattern can be matched to new, unclassified objects to aid in deciding the category to which the new object belongs. 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. Recent work emphasizes structural relationships among sets of features more than statistical ones [66].

In addition to pattern matching, Bayesian probability analysis is used to aid diagnosis by explicitly considering the prevalence of diseases in making diagnoses. Ideally, some investigators argue, the choice of a diagnosis should consider both the conditional probability of a disease given the patient data and the prevalence of hypothesized diseases in the population, as related by Bayes' theorem [12]. The appeal of Bayes' Theorem 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. For example, deDombal reports that a computer system, using Bayesian statistics, outperformed physicians in diagnosing causes of abdominal pain (e.g., appendicitis, diverticulitis, perforated ulcer, etc.) [12]. The proportion of correct diagnoses was 91% for the computer, and it varied from 42% to 81% for different physician groups. In addition, the fraction of surgeries for nondiseased appendices dropped from 25% to 7%. However, there are also several limitations to the approach which we discuss below. Lusted's classic volume [41] 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. A variety of additional assumptions must be made. For example: (1) the diseases under consideration are assumed to be mutually exclusive and exhaustive (i.e., the patient is assumed to have exactly one of the n diseases), (2) the clinical observations are assumed to be conditionally independent over a given disease(1), and (3) the incidence of the symptoms of a disease is

---

(1) 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 [64].

assumed to be stationary (i.e., the model generally does not allow for changes in disease patterns over time).

3.2 Example

The ARAMIS system of Fries [19] is one of the most successful projects using statistical analysis of patient data. ARAMIS was designed originally for use in an outpatient rheumatology clinic, but then broadened to a general clinical database system (TOD) [76] so that it became transferable to clinics in oncology, metabolic disease, cardiology, endocrinology, and some pediatric subspecialties. All clinic records are kept in a flow-charting format in which a column in a large table indicates a specific clinic visit and the rows indicate the relevant clinical parameters that are being followed over time. These charts are maintained by the physicians seeing patients in clinic, and each new column of data, corresponding to each new visit, 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 [30].

The information in the TOD databank can be utilized 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 [76]. However, it is in the analysis of stored clinical experience that the ARAMIS system has its greatest potential utility [20]. In addition to performing search and statistical functions such as those developed in databank systems for clinical investigation, 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. Based on 2 to 5 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, etc.). 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.

3.3 Discussion of the Methodologies

Statistical analysis systems offer powerful capabilities to individual clinical decision makers. 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 analysis of statistical data, however, which we discuss below.

(1) A continuing issue is recognizing the context in which analysis of data is appropriate. Statistical analyses -- whether based on analysis of confidence limits, Bayesian analysis, multivariate analysis or some other technique -- depend ultimately on classifying an individual patient record as an instance of data drawn from a well characterized population. Diagnosis or prognosis is based on the inference that the properties of the patient are those of the population, with some statistical measure of confidence. However, knowledge of statistical techniques provides no basis for assessing the clinical importance of the differences between a patient record and the summary characteristics of a population. For example, continuing problems with use of pure statistical methods include deciding whether the presence of an arbitrary second disease or the recent use of a drug should change the interpretation that a patient's data belong to some population. A recent editorial discusses an extreme but important case in which the issue of analysis of context is a crucial issue in interpreting the results of a statistical analysis [47]. The issue of context is complicated by the need to change context, for example in making a local decision that some area of an X-ray is inside or outside the heart, and then making a global decision about whether heart volume is abnormal for the particular patient.

(2) 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.

(3) Analysis of data can be complicated by missing values that frequently occur, outlying values, and poor reproducibility of data across time and among physicians. Unfortunately, neither statistical nor data analysis methods provides a basis for acquiring, representing and manipulating the domain-specific knowledge of users about how to handle problems in data.

(4) The decision aids provided tend to emphasize patient management rather than diagnosis. For example, the ARAMIS (TOD) prognostic routines, which are designed for patient management, assume that the patient's rheumatologic diagnosis is already known.

(5) There is no formal correlation between the way expert physicians approach patient management decisions and the way the programs arrive at recommendations. Psychological studies have now shown repeatedly that human clinical decision-making depends upon skilled use of large amounts of symbolic knowledge [15], [14]. Lacking this correlation, statistical analysis systems have a severely limited ability to explain either their significance or the basis for their results.

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. Selected data from a single patient is the typical input to a CADM system based on a mathematical model. Large amounts of raw data are often summarized using mathematical techniques; computing an average is a simple example. The processing is mathematical, using either direct analytical or simulation solution methods. The model itself is usually a mathematical representation of a general law of physics as it may apply in a particular situation.

Computer-based decision aids have been applied in such domains since the issues are generally well-defined. The actual techniques used by such programs tend to reflect the details of the individual applications, as in pharmacokinetics (e.g., digitalis dosing), acid-base/electrolyte disorders, and respiratory care.

One or two cooperating experts in the field generally assist with the definition of pertinent variables and the mathematical characterization of the relationships among them. Often an interactive program is then developed which requests the relevant data, makes the appropriate computations, and provides a clinical analysis or recommendation for therapy based upon the computational results. Some of the programs have also involved branched-chain logic to guide decisions about what further data are needed for adequate analysis(2).

Programs to assist with digitalis dosing have progressed to the inclusion of broader medical knowledge over the last ten years. The earliest work was Jelliffe's [32] 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% [33]. Later, 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 [56]. More recently, a third group in Boston, noting the insensitivity of the first two approaches to the kinds of nonnumeric observations that experts tend to use in modifying digitalis therapy, augmented the pharmacokinetic model with a patient-specific model of clinical status [26]. 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.

----

(2) "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.

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 [4], it was later expanded to consider electrolyte abnormalities as well [5], [6]. 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. 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.

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 [5]. This significant educational impact contrasts with a result described in Section 3.1 [43], where there was rapid loss in benefits after the using physicians stopped using the system.

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 therapeutic measures. 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 consideration of specific relationships only when Bleich specifically thought to include them 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 [52] who analyzed Bleich's program along with 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 Methodology

The programs mentioned in this section are very different in several respects, and each tends to overlap with other methodologies 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 limitations of all algorithmic approaches, most importantly its highly structured and inflexible logic which is unable to contend with unforseen circumstances not specifically included in the algorithm. The digitalis dosing programs all draw on mathematical techniques from the field of biomedical modeling (not discussed here), 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 captured and utilized in conjunction with mathematical techniques [26].

Thus, a general problem with mathematical approaches is that they depend on appropriate selection of the model to use, on assignment of default values to parameters whose measured values are unknown or unknowable, and ultimately on interpreting the clinical significance of the results of an analysis. Interpretation of results must be done both in terms of properties of the data used in the analysis itself and in terms of the general situation, including factors such as history, current therapies and clinical course of the patient. As in the techniques previously discussed in sections 3 and 4, the techniques of mathematical analysis themselves provide no basis for representing and manipulating the knowledge necessary to set up the model and to interpret its results.

5 Decision Theoretical Approaches

5.1 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 medically-oriented reviews of the field are available in textbooks [73], journal articles (such as an entire issue of the New England Journal of Medicine which was devoted to papers on this methodology [31]), and recently in a series in the journal Medical Decision Making [27]. This series describes cases referred to the new Decision Analysis service at Tufts New England Medical Center.

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 [54]. 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 [21]. 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. Thus, a path corresponds to a set of choices and outcomes. 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 inguinal 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 cost 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(3) separate from their knowledge of the probabilities linked to the associated chance nodes. An alternate approach has been the development of lotteries [50]. 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.

---

(3) also termed "utilities" in some references; hence the term "utility theory" [51].

(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 in [34] and [44].

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 [54]. 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 [24].

5.2 Example

We briefly describe the program of Gorry et al., developed for the management of acute renal failure [24]. Drawing upon Gorry's experience with the sequential Bayesian approach previously mentioned [23], 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 utilized a sequential test selection process based on Bayes' Theorem and omitting more advanced decision theoretical methodology [23]. The conditional probabilities utilized were subjective estimates obtained from an expert nephrologist. Leaper et al [38] discuss the fact that, in comparison with accurate statistics, the subjective probability estimates of experts are frequently unreliable for decision-making (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 nor 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 better to treat the patient immediately or to 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.

5.3 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 [38]. 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 that considerable expertise in both decision analysis and in the domain area are needed to construct and use the decision tree itself. From the perspective of the user, the difficulty is similar to the issue of context, as discussed in Section 4.3. Knowledge of the domain-independent method of decision analysis does not, in itself, provide a basis for representing a decision tree which pertains to a particular issue. In addition, if a generic tree is produced by one investigator, in one institution, then there is inevitable and appropriate question about its use in a different institution which uses somewhat different procedures. 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. An additional 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 [69].

Psychological studies are showing the nature of the clinical reasoning of humans

[15], [14], and attempts to model or be consistent with that decision-making process have provided one motivation for the applications of symbolic reasoning techniques to be discussed in the next section.

6 Symbolic Reasoning Approaches

6.1 Overview

In the early 1970's researchers at several institutions simultaneously began to investigate the potential applications of artificial intelligence (AI) to clinical decision making. The field is well-reviewed in a recent series of books [3], [10], and another recent book reviews the applications of AI in medicine [65]. The term "artificial intelligence" is generally accepted to include those computer applications which perform symbolic reasoning rather than numeric calculation. 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.

The bases for symbolic reasoning programs, and the source of their power, are qualitative judgments - codified as "heuristics", or experiential or "good guess" knowledge. In contrast, the bases for numerical calculation programs are analytical equations or statistical techniques. Heuristics encode knowledge about a problem area, such as medical diagnosis. They are used to 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 directly address the issue of context discussed above in Section 4.3 by both focusing the reasoning process and by deleting irrelevant items from consideration. The result is that these programs pursue a line of reasoning as opposed to following a sequence of steps in a calculation.

A landmark paper by Gorry in 1973 first critically analyzed conventional approaches to computer-based clinical decision making and outlined his motivation for turning to newer symbolic techniques [25]. He used the acute renal failure program discussed in Section 7.2 [24] as an example of the problems arising when decision analysis is used alone. In particular, he analyzed some of the cases on which the renal failure 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 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 utilize 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.

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

> (1) Concept Formation. 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 utilized. So-called "common sense" was often clearly lacking when the programs failed, and this was often what most alienated potential physician users.
>
> (2) Language Development. 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 [49].

Due to the limitations of the older data-intensive techniques, it was perhaps inevitable that some medical researchers would turn to the AI field for new methodologies. 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 complete issue of the journal Artificial Intelligence was devoted entirely to applications of AI to biology, medicine and chemistry [62](4).

Among the programs using symbolic reasoning techniques are several systems that have been particularly novel and successful. Pople and Myers have developed a system called INTERNIST that assists with test selection for the diagnosis of all diseases in internal medicine. A recent article summarizes the results of using INTERNIST-1 to analyze 19 trial cases reported in New England Journal of Medicine clinicopatholgical exercises [46]. The program performance appeared to be qualitatively similar to discussions of these cases

---

(4) Many of the AI-based systems described in this this article were developed on the SUMEX-AIM computing resource, a nationally shared system devoted entirely to applications of AI to the biomedical sciences. SUMEX-AIM computers are physically located at Stanford and Rutgers Universities, but they are used by researchers nationwide via connections to the TYMNET. The resource is funded by the Division of Research Resources, Biotechnology Branch, National Institutes of Health.

by hospital physicians. Quantitatively, of 43 total possible diagnoses when the correct diagnosis was known by subsequent pathological study, the expert physicians missed 21; the hospital physicians missed 28; and the program missed 29. The program utilizes 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. The system designers characterize it as a research project, not yet ready for routine clinical use. For the designers, the most troublesome system limitations are its inability to reason anatomically or temporally and its limited representation and use of pathophysiological knowledge. In addition, the system does not yet have the well developed human engineering features which are necessary for routine use. In an accompanying editorial, Barnett suggests that the contribution of INTERNIST-1 should not be measured simply by its limitations or failures [2]. Rather, experiments such as INTERNIST provide insight into both the human diagnostic process and the way that computers can help the practitioner. Barnett continues: "The issue is whether such artificial intelligence models can reach conclusions similar to those of a competent clinician and can then justify those conclusions in a rational and clinically acceptable fashion."

Most AI programs have been developed on relatively large mainframe computers. In recent years, several medical diagnosis programs based on AI techniques have been made to work on small and relatively inexpensive computers. For example, an expert system designed for operation on a chip has been reported [75], and the PUFF program runs routinely on a personal computer in a hospital laboratory [1]. PUFF interprets pulmonary function test results. Taking measured data directly from a laboratory computer, PUFF identifies the presence and severity of one or more of obstructive or restrictive diseases and diffusion defect. The program produces a report for each patient which explains the reasons for its diagnosis, considers the measured and potential effect of drug therapy, and discusses the potential of further testing when appropriate to confirm a diagnosis. The PUFF system report is reviewed by a physician, edited if necessary (about 90% of the reports need no editing), signed and entered into the patient record. The program has interpreted over 4000 cases since it first went into routine use in 1978 at Presbyterian Hospital, Pacific Medical Center, San Francisco.

6.2 Example

The symbolic reasoning program selected for discussion is the MYCIN System at Stanford University [59]. The researchers cited a variety of design considerations which motivated the selection of AI methodologies for the consultation system they were developing. 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 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 [59](5). A

(5) The production rule is a representation of knowledge which is frequently employed in AI research [11], and it has been effectively applied to other scientific problem domains [8].

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:

(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;

(2) the rules are used by the computer program, the using physicians, and the designing experts. Experts' rules can be translated from stylized English to internal representation, and rules can be translated from internal representation to stylized English for display to users;

(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 in this complex task domain. They allow the judgmental knowledge of experts to be captured in rule form and then utilized in a consistent fashion.

The MYCIN System has been evaluated regarding its performance at therapy selection for patients with either septicemia [80] or meningitis [79]. The program performs comparably with experts in these two task domains. Since it has no rules regarding the other infectious disease problem areas, however, questions regarding its acceptability to physicians cannot yet be assessed.

6.3 Discussion of the Methodology

Symbolic reasoning techniques differ from the other methodologies mentioned in this article in that the computer techniques themselves are as yet experimental and rapidly changing. Whereas the computations involved in Bayes' Theorem, for example, involve straightforward application of computing techniques already well-developed, basic researchers in computer science continue to develop new methodologies for knowledge representation, language understanding, heuristic search, and the other symbolic reasoning problems we have mentioned. Thus the AI programs tend to be developed in highly experimental environments where short term practical results are often unlikely to be found. Prototype programs typically require large amounts of space and tend to be slow, particularly in time-sharing environments. As has been true for most of the methodologies discussed, AI researchers have still not developed adequate methods for handling concurrent diseases, assessing the time course of disease, nor acquiring adequate structured knowledge from experts. Furthermore, inexact reasoning techniques tend to be developed and justified largely on intuitive grounds.

Despite these significant limitations, the techniques of artificial intelligence do provide a way to respond to many of Gorry's observations regarding the inadequacies of prior methodologies as described above [25]. There are now several programs responsive to his criticisms. The INTERNIST system has received close attention within both the computer science and the medical communities because of its ability to make multiple and complex diagnoses within a very broad problem domain. The psychological research into the nature of diagnosis shows the complexity of the process followed by humans and the richness of the knowledge used in following that process. The INTERNIST system is the best documented example of the power of the symbolic processing methods to make diagnostic decisions at a high level of competence and to justify its conclusions to human decision-makers.

Szolovits and Pauker 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 [64]. They identify several serious 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 some of the reasoning styles observed in experts [15],[14], 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.

7 Integration of multiple techniques

Sections 3-7 of this paper discussed the approach, the strengths and the limitations of five basic methods of computer-based decision making. No one technique is best for all applications, and some techniques have strengths in the same areas as other techniques have limitations. Thus, it is not surprising that some investigators have attempted to develop systems which combine multiple techniques.

These integrated systems show power and promise. Bleich's system, for example, combines techniques of algorithms and mathematical modeling, as discussed in Section 5.2. The Digitalis program was designed to recommend Digitalis dose rates based on therapeutic goals, clinical description of the patient state, and measured laboratory results []GORRY78}. The program uses heuristics to develop a description, called a "patient specific model" of an individual patient. An associated mathematical model uses the results of heuristic analysis as input, and it produces quantitative results which are added to the patient-specific model. Another example, the AI/MM system, represents and manipulates a physiological model. The program includes symbolic knowledge of anatomy and physiological function, and it uses simple mathematical descriptions of first principles of physics and physiology [37]. AI/MM integrates the use of AI and simple mathematical models. Using the current AI/MM knowledge base, the program analyzes the physiological model to infer the behavior of the renal system. It explains the basis for behavior in terms either of first principles or of high-level heuristics. This symbolic analysis is used to suggest the use of, to set up, and to evaluate the results of quantitative analyses; quantitative results in turn are considered in symbolic analysis of physiological behavior. Blum has done work in discovery of causal relations in a clinical data base [7]. Blum's program (RX) integrates techniques of artificial intelligence and statistics. Though not a CADM system, a CADM system could be built using the insight of RX that AI and statistics can be combined. The principal objective of RX is to automate the process of generating and testing hypotheses about causal relations in large data banks. RX uses knowledge about clinical medicine for generating and testing hypotheses about new causal relations. It uses AI techniques to identify relevant knowledge and to choose how to employ it in particular situations, and it uses statistical techniques to test the validity of hypothesized relations.

An additional example is discussed below in greater detail.

7.1 Example

Warner's group in Salt Lake City has developed a computer-based system for implementing clinical algorithms [70]. The HELP system is widely used within the laboratories and patient-care areas of LDS hospital in Salt Lake City. It has been integrated with a hospital information system, so data from the hospital laboratories, intensive care units, and general patient care facilities can be used as input for its decision-making capability.

The HELP system includes many simple clinical algorithms, a large data bank, some mathematical models, and a program which uses decision-analysis methods to identify expected values of decisions and perform sensitivity analyses. In addition to supporting decisions about individual patient care, the contents of the data bank can be analyzed statistically to improve the clinical algorithms. Thus, uncertain decisions can be analyzed with Bayes' theorem using probabilities derived from the results of treating the local hospital population. The system provides editing and display facilities both for users and for designers of new HELP modules, or "sectors".

The heart of the system is a set of simple data interpretation and decision-making algorithms encoded in HELP sectors. A representative HELP sector describes an algorithm which says that if there is Grade I or II aortic regurgitation in the angiographic data, and aortic valve area is less than a specified value, then the patient has severe aortic stenosis.

The HELP system is remarkable for its integration of computer-assisted decision-making into the hospital information processing system. More than any other system, the HELP system integrates the process of collecting, reporting and interpreting the clinical significance of all the data in a large hospital. One of the clear features of the system is its explicit focus on decision-making. Unlike a "hospital information system", which focuses on information rather than decisions, the HELP system always has a clear focus on the purpose for which information is collected. In addition, it integrates different techniques of decision-making: clinical algorithms, statistical analysis and decision analysis. Another feature of the HELP system is that is designed for efficient, economical operation now, with available minicomputer technology.

7.2 Discussion

The systems discussed in this section suggest some of the potential benefits of integrating different decision making techniques into a unified system. This integration appears to have important benefits from the perspective of the user. The HELP system is an integrated system. The system provides current patient data, data about similar kinds of patients who have been seen in the institution; it suggests the common implications of the patient data, as conceived by thoughtful experts. It uses algorithms and mathematical, statistical, and decision-analysis techniques for identifying the implications of the data of an individual patient. Thus, from an integrated system, the user can obtain the data and the analysis of the data by several specialized scientific disciplines.

In addition, the integrated system offers great power to the expert decision-maker. The HELP system provides a collection of tools which the decision-maker can use in a CADM system for assisting problem-oriented patient management. To the extent that problems can be inferred from data, the HELP system itself can assist in identifying clinically important problems. The system's algorithms can assist the user to validate the accuracy of any given problem or to determine the effectiveness of therapy. The system provides the mechanism to use the techniques of mathematical, statistical or decision analysis as appropriate.

Technical limitations to the HELP system may partially account for the failure of the research community to build on its concepts. The HELP system lacks the reasoning ability associated with AI techniques. Thus, it has limited ability to address the issue of context as described in the discussions of Sections 3-5. Because of the way its knowledge is represented explicitly as procedures, it can only answer questions which its designers planned on its answering. As Bleich's program (discussed in Section 5), the HELP system is limited by its dependence on simple algorithms. In contrast, Szolovits remarks, "the fundamental insight of the MYCIN investigators was that the complex behavior of a program which might require a flowchart of hundreds of pages to implement as a clinical algorithm could be reproduced by a few hundred concise rules and a simple recursive algorithm (described in a one-page flowchart) to apply each rule just when it promised to yield information needed by another rule" [65].

8 Conclusions

This review has shown that there are two recurring issues to confront in considering the field of computer-based clinical decision making:

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

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

We shall summarize by reviewing these points separately.

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 lack of higher level domain knowledge limits the effectiveness of the Bayesian approach in more complex patient management or diagnostic environments. A mathematical approach may support decision making in certain well-described fields in which observations are typically quantified, and related by functional expressions. These examples, and others, demonstrate the the need for thoughtful consideration of the technique most appropriate for managing a clinical problem. In general the simplest effective methodology is to be preferred, but acceptability issues must also be considered as discussed below.

Recent work shows the potential of combining two or more techniques for processing in a CADM system. The systems of Bleich and Warner (see Sections 5.1 and 8.1) each show that there is more power in a combination of two or more processing techniques than there is in any of the techniques individually used in these systems. In addition, there seems to be a special opportunity for combining AI with the quantitative techniques of statistical and mathematical analysis and, potentially, with decision analysis. AI offers a formalism for representing and manipulating the heuristic knowledge of expert human problem solvers. This heuristic knowledge can be used to address the issue of recognizing or establishing an appropriate context for using some processing technique, and heuristic knowledge can be used to assess the qualitative clinical significance of computed results. Thus, for example, Yamomoto suggests that mathematical models might be incorporated into AI-based systems [78]. AI/MM uses AI and simple mathematical models; it uses quantitative and qualitative problem solving to support each other, and it shows the epistemological basis of inference steps, distinguishing between inferences based on physical law and on heuristic associations. RX uses AI and statistics; it generates and interprets statistical results in a more powerful way than would be possible with either technique alone.

It is always appropriate to ask whether computer-based approaches are needed at all for a given decision making task. Some clinical algorithm developers, for example, have discarded the machine [57], [67], and Schwartz et al. pointed out that a useful decision analysis can often be accomplished in a qualitative manner using paper and pencil [54]. Similarly, it is appropriate to ask whether complex computer processing is needed for a task. Sometimes, as in the work of Mesel, the needs of users can be met effectively by a simple computer-based approach which can be developed relatively quickly and run economically.

Finally, however, it is 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 the degree of judgmental or structural knowledge (as opposed to data) that is contained in the program. Analyses of human clinical decision making [15], [14] 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. It is likely that medical computing researchers will similarly have to become "knowledge engineers" in the sense that they will look for effective ways to match the knowledge structures that they use to the complexity of the tasks they are undertaking.

Acceptability Issues

A recurring observation as one reviews the literature of computer-based medical decision making is that essentially no system has been effectively utilized outside of the environment in which it was developed, even when its performance has been shown to be excellent! This suggests that it may be an error to concentrate our research effort primarily on improving the decision making performance of computers when there is evidently much more required before these systems will have clinical impact. An editorial critique suggests that issues of terminal design and interface between user and computer must be addressed effectively before computers will have a substantial impact in clinical medicine [18]. It is tempting to conclude that the biases of medical personnel against computers are so strong that systems will inevitably be rejected, regardless of performance, and in fact there are some data to support this view [63]. 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 [71].

In spite of their goals and achievements, the systems described in this paper have not been used outside of the institutions in which they were developed. Commercial neglect of CADM systems probably stems from at least three issues: lack of demand from the hospital community for computer-assisted decision-making; unwillingness of vendors to develop a product for which there is limited apparent demand; and limited convincing validation that computer decision-support provides more effective or more economical care.

Perhaps one of the most revealing lessons about acceptability is an observation regarding the system of Mesel et al. that we described in Section 3.2 [45]. 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 a tertiary care center where all complex chemotherapy was administered. The introduction of the protocols permitted these physicians to undertake tasks that they had previously been unable to do, and 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 are using the machine have a clear advantage in their practice [53].

A heightened awareness of "human engineering" issues among medical computing researchers is also apt to help improve acceptance of computers by physicians. Fox has recently reviewed this field in detail [17]. The issues range from the mechanics of interaction at a computer terminal to program characteristics designed to make the system appear as a tool for the physician rather than a dogmatic advice-giving machine.

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 [63].

Finally it must be noted that acceptability issues should generally be considered from the outset in a system's design because they may dictate the choice of methodology as much as the task domain itself does. The role of formal knowledge structures to facilitate explanation capabilities, for example, may argue in favor of using symbolic reasoning techniques even when a somewhat less complex methodology might have been adequate for the decision task.

In summary, the trend towards increased use of knowledge engineering techniques for clinical decision programs has been in response to desires for both improved performance and improved acceptance of such systems. As greater experience is gained with these techniques and 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 methodologies we have discussed. One lesson to be drawn lies in the recognition that there is basic computer science research to be done in medical computing, and that the field is more than the application of established computing techniques in medical domains.

References

1. Aikins, J.S., Kunz, J.C., Shortliffe, E.H., Fallat, R.J., "PUFF: an expert system for interpretation of pulmonary function data" Stanford report STAN-CS-82-931 (1982).

2. Barnett, G.O. "The computer and clinical judgement." (editorial) N. Eng. J. Med. 307, pp. 493-494 (1982).

3. Barr, A., Feigenbaum, E.A. Handbook of Artificial Intelligence Vol. 1-2, Los Altos, CA: William Kaufman Inc., (1982).

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

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

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

7. Blum, R.L. "Discovery, confirmation, and incorporation of causal relationships from a large time-oriented clinical database: The RX project." Comp. Biomed. Res. (1982).

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

9. Cannon, S.R., Gardner, R.M. "Experience with a computerized interactive protocol system using HELP." Comp. Biomed. Res. 13, PP. 399-409 (1980).

10. Cohen, P.R., Feigenbaum, E.A. Handbook of Artificial Intelligence Vol. 3, Los Altos, CA: William Kaufman Inc., (1982).

11. 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).

12. 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).

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

14. Eddy, D.M., Clanton, C.H. "The Art of Diagnosis." New Eng. J. Med., 306, pp. 1263-1268 (1982).

15. Elstein, A.S., Shulman, L.S., and Sprafka, S.A. Medical Problem Solving: An Analysis of Clinical Reasoning. Cambridge, Ma.: Harvard Univ. Press, (1978).

16. 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.

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

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

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

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

21. 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).

22. Goldman, L., Weinberg, M. et al. "A computer-derived protocol to aid in the diagnosis of emergency room patients with acute chest pain." New Eng. J. Med. 307, pp. 588-596 (1982).

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

24. 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).

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

26. 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).

27. Gottle[illegible] J.E., Pauker, S.G. "Whether or not to administer amphotericin to an immunosuppressed patient with hematologic malignancy and undiagnosed fever." Medical Decision Making pp.75-93 (1981).

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

29. 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).

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

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

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

33. 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).

34. Komaroff, A.L. "The variability and inaccuracy of medical data." Proc IEEE 67, pp. 1196-1207 (1979).

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

36. Knaus, W.A., Draper, E.A., et al. "Evaluating outcomes from intensive care: a preliminary multihospital comparison." Critical Care Med. 10, pp. 491-496 (1982).

37. Kunz, John C. "Use of artificial intelligence and simple mathematics to analyze a pysiological model", Ph.D. thesis, Stanford University (1983).

38. 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).

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

40. 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).

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

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

43. McDonald, C.J., Wilson, G.A., McCabe, G.P. "Physician response to computer reminders" J. Am. Med. Assoc. 244, pp. 1579-1581 (1980).

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

45. 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).

46. Miller, R.A., Pople, H.E., and Myers, J.D. "INTERNIST-1, an experimental computer-based diagnostic consultant for general internal medicine." N. Eng. J. Med. 307, pp. 468-476 (1982).

47. Multiple Risk Factor Intervention Trial Research Group "Multiple Risk Factor Intervention Trial" JAMA 248, pp. 1465-1477 (1982).

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

49. 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. 60:981-996 (1976).

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

51. Raiffa, H. Decision Analysis: Introductory Lectures on Choices Under Uncertainty. Reading, Ma.: Addison Wesley, (1968).

52. 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.

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

54. 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).

55. 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).

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

57. Sherman, H., Reiffen, B., and Komoroff, 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.

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

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

60. Shortliffe, E.H., Buchanan, B.G., Feigenbaum, E.A. "Knowledge Engineering for Medical Decision Making: A Review of Computer-Based Clinical Decision Aids." Proc. of IEEE 67, pp. 1207-1224 (1979).

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

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

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

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

65. Szolovits, P., Artificial Intelligence in Medicine. AAAS Selected Symposium 51, Westview Press (1982).

66. van den Akker, T.J., Ros, H.H., et al. "An on-line method for reliable detection of waveforms and subsequent estimation of events in physiological signals." Comp. Biomed. Res. 15, pp 405-417 (1982).

67. Vickery, D.M. "Computer support of paramedical personnel: the question of quality control." Medinfo 74, pp. 281-287, North-Holland: Amsterdam, 1974.

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

69. 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.

70. Warner, H.R. Computer-Assisted Medical Decision Making. Academic Press (1980).

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

72. 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).

73. Weinstein, M.C., Fineberg, H.V., et al. Clinical decision analysis Philadelphia, (1980).

74. 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).

75. Weiss, S.M. and Kulikowski, C.A. "Developing microprocessor-based expert models for instrument interpretation." Proceedings IJCAI-81,pp. 853-855, Vancouver, B.C., (1981).

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

77. Wirtschafter, D.D., Scalise, M., Henke, C., Gams, R.A. "Do information systems improve the quality of clinical research? Results of a randomized trial in a cooperative multi-institutional cancer group." Comput. Biomed. Res. 14, PP. 78-90 (1981).

78. Yamamoto, W.S., Walton, E.S. "On the evolution of the physiological model" in Annual review of biophysics and bioengineering 4,p. 81-102 (1975).

79. 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, November 1978.

80. Yu, V.L., Fagan, L.M., et al. "Antimicrobial selection for meningitis by a computerized consultant: a blinded evaluation by infectious disease experts." J. Am. Med. Assoc. 241(12), pp. 1279-1282 (1979).

81. Ziporyn, T., "Computer-assisted medical decision-making: interest growing." J. Am. Med. Assoc. 248, pp. 913-918 (1982).

82. 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).

FILMED FROM BEST AVAILABLE COPY