

INTERNIST-I *Properties*: Representing Common Sense and Good Medical Practice in a Computerized Medical Knowledge Base

FRED E. MASARIE, JR., RANDOLPH A. MILLER, AND JACK D. MYERS

Decision Systems Laboratory, 1360 Scaife Hall, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261

Received March 29, 1985

INTERNIST-I is an experimental diagnostic consultant program for use on clinically challenging cases in internal medicine. *Properties* are a part of the INTERNIST-I knowledge representation scheme which embody essential information about the medical facts contained in the knowledge base. *Properties* provide the INTERNIST-I program with a measure of common sense and encourage the program to follow good medical practice. *Properties* substantially influence the behavior of the INTERNIST-I diagnostic program during case analyses. This paper reviews the implementation and significance of *properties*. © 1985 Academic Press, Inc.

1. INTRODUCTION

Properties are an essential, but previously undescribed, component of INTERNIST-I, an experimental consultant program for the diagnosis of complex cases in general internal medicine. *Properties* embody information about the individual patient findings represented in the INTERNIST-I knowledge base (KB) and also detail interrelationships among findings. *Properties* were incorporated into the INTERNIST-I KB from its inception over a decade ago, but have assumed an increasingly important role in influencing the diagnostic program's behavior in the following ways:

- a. They convey the clinical significance of findings, represent the cost (both in economic and morbidity/mortality terms) of eliciting findings, and group possible test outcomes into logically related sets.
- b. They detail contraindications for obtaining certain findings and indicate confounding situations in which tests may become difficult to interpret.
- c. They can prevent the INTERNIST-I system from seeking redundant or contradictory information because *properties* detail interdependencies among findings.
- d. They provide the INTERNIST-I program with a measure of common sense and encourage the program to follow good medical practice.

This paper discusses the implementation of *properties* in the INTERNIST-I diagnostic system and describes how they substantially affect the program's behavior.

2. REVIEW OF INTERNIST-I KNOWLEDGE BASE

The design and implementation of INTERNIST-I are the result of a decade-long collaboration between two internists (authors J.D.M. and R.A.M.) and Harry Pople, Jr., a computer scientist with expertise in artificial intelligence. We will briefly review the structure of the knowledge base (KB) here. A detailed discussion of the INTERNIST-I consultant system has been published in the *New England Journal of Medicine* (1).

The knowledge base represents over 20 person-years of effort. Built under the supervision of one of the authors (J.D.M.), the KB consists of approximately 530 profiles of diseases in internal medicine. A disease profile is a comprehensive list of the findings (manifestations) which can occur in patients with a given disease. Manifestations include historical information, symptoms, physical signs, and the results of laboratory tests. A profile usually contains between 25 and 200 (average 85) separate findings. Each finding within a disease profile has two numbers associated with it—an evoking strength (EVOK) and a frequency (FREQ).

The evoking strength of a finding in a disease profile is a semiquantitative estimate of the relative specificity of the finding for the given disease. Evoking strengths are integers between 0 and 5 (inclusive). If a finding is nonspecific, it receives an EVOK of 0. That is, observation of the finding by itself should not suggest any specific diseases as being its cause. Examples of nonspecific findings include ANOREXIA, WBC 14000 TO 30000, and FEVER. On the other hand, a finding which is pathognomonic for a given disease is given an EVOK of 5 in the disease profile. For example, the finding ARTERIES SUPERFICIAL CRANIAL TENDERNESS REDNESS SWELLING is pathognomonic for CRANIAL ARTERITIS and is given an EVOK of 5 in the disease profile of CRANIAL ARTERITIS.

The frequency number of a finding in a disease profile represents the incidence of the finding in patients with the disease. Frequencies are integers on a scale from 1 to 5. A FREQ of 1 indicates that the finding seldom occurs in patients with the disease; 2, that the finding occurs in a significant minority of patients; 3, that the finding occurs in about half; 4, in a significant majority; and 5, the finding occurs in essentially all patients with the disease. For example, in the disease profile for acute appendicitis, the finding ABDOMEN PAIN RIGHT LOWER QUADRANT carries a FREQ of 4, meaning that a significant majority (but not all) of the patients with acute appendicitis will have right lower quadrant pain.

The internal representation form of individual findings in the KB is not readily apparent to the user. Within the INTERNIST-I program, each text string expressing a manifestation name is represented simply as an integer. For example,

KIDNEY <S> IVP URETER <S> OBSTRUCTED

is represented internally as "958" and

KIDNEY (S) RETROGRADE PYELOGRAPHY URETER (S)
OBSTRUCTED

is represented as "961."¹ Medically knowledgeable persons reading the two manifestations above realize that they refer to the same anatomical abnormality. However, to the INTERNIST-I system, the findings are "958" and "961," and since the system cannot parse the text strings, the relationship is not apparent. Thus, the "atomic" nature of findings is a major problem with the basic INTERNIST-I knowledge representation scheme.

3. EVOLUTION AND CREATION OF PROPERTIES

Over the past decade, many hundreds of patient cases have been analyzed using INTERNIST-I to evaluate and refine its KB and diagnostic algorithms. Since the system has not been validated for prospective use on humans, the clinical material has come from retrospective case histories. We have used clinico-pathological conference (CPC) cases published in medical journals as well as cases from our own institution. The evaluation process has uncovered deficiencies in the knowledge base (1, 2). These deficits included errors of omission (failure to enter important facts), errors of commission (entry of incorrect facts), and the failure to incorporate important relationships or concepts into the knowledge representation schemes.

The lack of a comprehensive representation of relationships among manifestations was a major problem noted early in the development of the INTERNIST-I diagnostic programs. To improve the diagnostic program's behavior, a significant amount of information (in addition to that in the disease profiles) had to be represented in the KB. Therefore, *properties* were created as KB structures that incorporate additional information about individual manifestations. The INTERNIST-I knowledge representation scheme thus encompasses two separate, intricately related knowledge bases—disease profiles and *property* lists.

Each manifestation has a unique list of associated *properties*. The *properties* of a manifestation are independent of its presence in disease profiles. Figure 1 provides an example of a finding—T3 SERUM DECREASED—with its *property* list. The syntax and function of specific *properties* will be described later. The information in Fig. 1 can be summarized as:

(1) IMPORT 2: If an isolated decreased serum T3 level is found in an otherwise normal patient, it does not necessarily have to be explained by a diagnosis. In the appropriate context, however, it may be important.

¹ '<S>' is a separate word in the INTERNIST vocabulary which is used to express plural forms of various words. Thus KIDNEY <S> means one or two kidneys.

T3 SERUM DECREASED

- (1) IMPORT 2
 - (2) TYPE LAB1
 - (3) GROUP THYROID FUNCTION TEST <S>
 - (4) FIFT T4 TOTAL SERUM INCREASED
 - (5) FIFT T3 SERUM INCREASED
 - (5) YFRN T3 SERUM INCREASED
 - (6) PREREQ T4 TOTAL SERUM DECREASED
 - (7) UNREQ ASPIRIN ADMINISTRATION IN LARGE
AMOUNT RECENT HX
 - (7) UNREQ PHENYTOIN ADMINISTRATION HX
-

FIG. 1. Properties for T3 SERUM DECREASED.

(2) TYPE LAB1: The measurement of serum T3 is a laboratory test which is mildly expensive, mildly invasive, or is not routinely available at all hospitals.

(3) GROUP THYROID FUNCTION TEST (S): The test is one of the ways to assess thyroid function.

(4) FIFT T4 TOTAL SERUM INCREASED: It would not be appropriate for a physician to ask whether the T3 is decreased in a patient if it were already known that the T4 was increased. This *property* embodies the fact that T4 and T3 levels usually parallel one another in the blood.

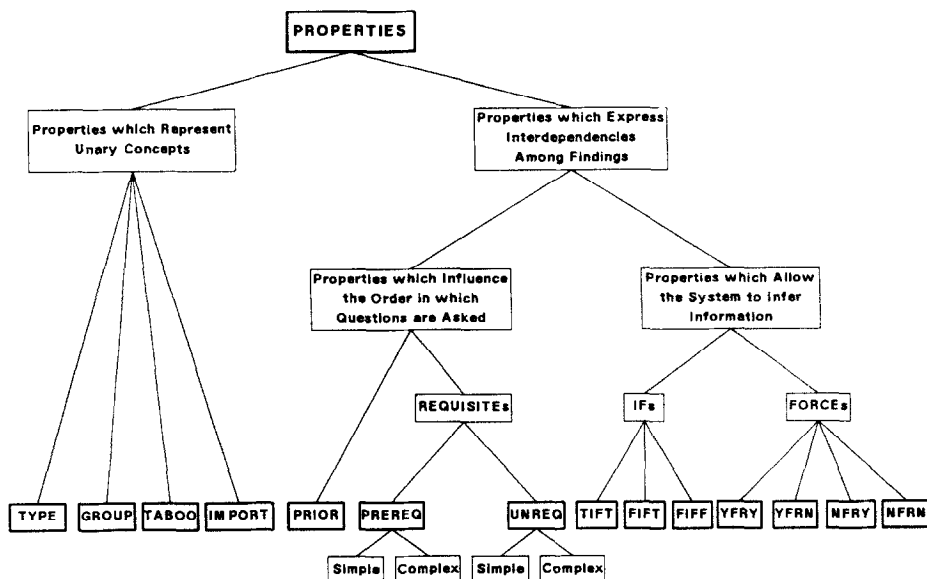
(5) FIFT T3 SERUM INCREASED and YFRN T3 SERUM INCREASED: If one knows that the T3 is increased in the blood, then one can assume that the T3 is not decreased.

(6) PREREQ T4 TOTAL SERUM DECREASED: One should first ask if the blood T4 is decreased, before asking if the blood T3 is decreased. The reason for this *property* is that T4 is technically easier to measure and that T3 is metabolized from T4.

(7) UNREQ ASPIRIN ADMINISTRATION IN LARGE AMOUNT RECENT HX and UNREQ PHENYTOIN ADMINISTRATION HX: T3 levels are difficult to interpret in patients who have been taking large doses of aspirin or phenytoin. Both of these substances are capable of displacing T3 and T4 from plasma proteins both *in vivo* and *in vitro*.

As disease profiles are built, it becomes necessary at times to create new manifestations. *Properties* are added to the knowledge base whenever new findings are created. However, it is nearly impossible for persons building the KB to derive all appropriate relationships between a new finding and the nearly 4000 existing findings. Literature review or human memory are imperfect tools for accomplishing this task. We have found that the experimental method is a useful adjunct in this regard. Many *properties* have been added to the KB when unexpected or inappropriate behavior is encountered during case analyses.

Properties can be separated into two main categories: those which represent unary concepts (i.e., facts about a given finding); and those which express interdependencies among two or more findings. Figure 2 outlines a hierarchical organization of the 14 different classes of *properties*. These classes can be

FIG. 2. A hierarchical organization of *properties*.

conceptually organized according to how they influence the behavior of the system.

4. PROPERTIES WHICH REPRESENT UNARY CONCEPTS

Properties representing unary concepts express specific attributes of individual findings. These *properties* function to assist the INTERNIST-I diagnostic program in filtering the information obtained from a patient case and help the program to organize its diagnostic evaluation.

4.1. TABOOS

Some diseases do not occur in certain patient populations. TABOOS associated with a finding detail those diseases which should not be considered if the finding is present in a patient. TABOOS are employed in two situations: where it is a physiological impossibility for the disease to occur if the finding is present, such as the disease ENDOMETRIOSIS given the finding SEX MALE; or where there is an exceedingly small probability that a disease will occur if the finding is present, such as the disease SICKLE CELL ANEMIA given the finding RACE WHITE. (N.B. Sickle cell disease has been reported in Caucasians in Middle Eastern countries but is exceedingly rare in the U.S.)

4.2. IMPORTS

A diagnostician must determine the clinical significance of a finding to form clinically relevant diagnostic hypotheses. In the INTERNIST-I knowledge base, the IMPORT of a manifestation conveys the necessity of explaining (by

TABLE 1

IMPORTS

Value	Explanation	Examples
1	Manifestation is usually unimportant, occurs commonly in normal persons, and is easily disregarded	ABDOMEN XRAY COLON DISTENSION WITH GAS CONSTIPATION IMPOTENCE
2	Manifestation may be of importance, but can often be ignored; context is important	DIARRHEA ACUTE PLATELETS GTR THAN 450000 CHEST AP DIAMETER INCREASED
3	Manifestation is of medium importance, but may be an unreliable indicator of any specific disease	DIARRHEA CHRONIC OBSTIPATION PLATELETS 50000 TO 150000
4	Manifestation is of high importance and can only rarely be disregarded as, for example, a false-positive result	PLATELETS LESS THAN 50000 FEVER LIVER ENLARGED MODERATE
5	Manifestation absolutely must be explained by one of the final diagnoses	ABDOMEN ENTERIC CUTANEOUS FISTULA ABDOMEN MASS EPIGASTRIUM ABDOMEN MASS FLANK (S) BILATERAL

making a diagnosis) the finding's presence in a given patient. IMPORTs are represented in the INTERNIST-I KB independently of the disease profiles. Each finding has an IMPORT value (an integer between 1 and 5). Table 1 provides an explanation for each IMPORT value.

The INTERNIST-I consultant program utilizes IMPORTs during three critical phases of its diagnostic reasoning: scoring of diagnostic hypotheses, construction of "problem areas," and in determining when to terminate diagnostic inquiries. Given the initial positive findings and pertinent negative findings in a patient case, the system constructs a comprehensive differential diagnosis list. This list includes all diseases which could potentially explain any of the patient's important (*import* of 3 or greater) findings. The diagnostic system then uses a scoring algorithm to rank the hypotheses on its differential diagnosis list. Part of the scoring process utilizes IMPORTs. A penalty score, based on the combined IMPORTs of the unexplained findings, is assessed against a hypothesis if it does not explain findings present in a patient case. Thus, a disease hypothesis is penalized for being "irrelevant" if it does not explain important findings present in the patient.

The master differential diagnosis list created by the program usually contains diseases involving multiple organ systems. It would be unusual for a physician, given a diverse list of possible patient diagnoses, to make a statement of the form, "Hmmm . . . Is the diagnosis this heart disease or that kidney disease or

this other lung disease?'. It is more appropriate to compare a heart disease to other heart diseases, and a renal disease to other renal diseases. Thus, an important task for the diagnostic program is to formulate a problem domain based on the most attractive diagnostic hypothesis—the one with the highest score. This process is analogous to problem area formulation by physicians. The purpose is to determine which diseases to consider as competitor diagnoses to the leading (topmost) diagnosis. Two diseases are considered as competitors if the important findings *not* explained by one are a subset of the important findings *not* explained by another. Two heart disease hypotheses will both leave as unexplained the findings due to renal or lung disease, whereas a heart disease and a renal disease will leave different sets of findings unexplained. IMPORTs play a key role in problem area formation in that findings of low or questionable significance are ignored. This prevents the diagnostic system from focusing on less important data and eliminates a tendency to create a plethora of inappropriate problem domains. The diseases not included as competitors to the topmost diagnosis are used to form subsequent problem areas. The system can make multiple diagnoses through the serial creation of separate problem domains.

Finally, it is important for a diagnostic program capable of making multiple diagnoses in a given patient case to "know when to stop." The INTERNIST-I program continues to make diagnoses until all observed findings with IMPORT values of 3 or greater have been explained. Therefore, IMPORTs provide INTERNIST-I with criteria for determining which findings are clinically important and should be explained by a diagnosis.

4.3. TYPES

Diagnoses should be pursued in a cost-efficient manner. TYPEs were developed to categorize manifestations according to the "cost" of eliciting each finding in a patient. The "cost" includes both the monetary expense and the morbidity and/or mortality risk associated with obtaining the information. The TYPE of a manifestation can be historical fact (HISTORY), patient symptom (SYMPTOM), physical finding (SIGN), or laboratory result. Laboratory findings are divided into three classes: LAB0, LAB1, and LAB2. LAB0 findings include simple laboratory tests such as EKG or hematocrit, which are routinely available, even at night, in most hospitals. LAB1 findings are mildly expensive or invasive tests such as an upper gastrointestinal barium series, or inexpensive tests not done at all hospitals such as certain radioimmunoassay tests (for example, angiotensin converting enzyme levels). LAB2 findings are expensive or invasive tests, such as aortography or liver biopsy. Table 2 shows the distribution of TYPEs within the INTERNIST-I KB and gives an example of a finding from each category.

TYPEs help the INTERNIST-I diagnostic program direct questioning in a cost-effective manner. In general, the program attempts to exhaust less "costly" discriminating questions before asking for more expensive or invasive results. The INTERNIST-I program usually asks further questions about the

TABLE 2

TYPES

Type	Number in the INTERNIST-I KB	Example
History	434	WEIGHT LOSS GTR THAN 10 PERCENT
Symptom	215	ABDOMEN PAIN COLICKY
Sign	1043	ABDOMEN TENDERNESS LEFT LOWER QUADRANT
LAB0	281	WBC 14000 TO 30000
LAB1	1145	COLON BARIUM ENEMA PSEUDOPOLYPOSIS
LAB2	766	AORTOGRAPHY AORTIC REGURGITATION

history and physical examination after being presented with a complex patient case. We have observed that the system tends to achieve greater yield from historical items, symptoms, and signs than many recently trained clinicians. These physicians tend to order laboratory tests earlier in a workup and depend more on the test results to make a diagnosis.

4.4. GROUPS

Many times one can determine several details about a patient's illness by performing a single test. This concept is incorporated into INTERNIST-I through GROUPS which organize findings into related sets. A GROUP contains findings that can be obtained by a common method. Examples of these GROUPS include BARIUM ENEMA, PALPATION ABDOMEN, and KIDNEY BIOPSY. Other GROUPS are based on a higher level conceptual organization such as THYROID FUNCTION TEST (S) or OCCUPATION HX. Figure 3 lists the findings comprising the GROUP THYROID FUNCTION TEST (S).

GROUPS help expedite user interactions during INTERNIST-I case analyses. Before the program asks the status of a specific finding, it will first prompt a user with the finding's GROUP name. This provides the user with the option to enter all pertinent positive and negative findings from the GROUP. Otherwise, the user would be asked each individual question, and would have to answer using 'YES,' 'NO,' or 'N/A' (information not available). The user can enter OMIT, if none of the items in the GROUP are known. All the findings in that GROUP are then recorded as being unavailable in the patient case.

GROUPS also provide a convenient way for users of INTERNIST-I to locate the appropriate INTERNIST-I terminology (manifestation names) for a patient's findings. The word order of the GROUP names was made to be different than the related INTERNIST-I manifestation, and usually is closer to the "normal syntax." This technique provides an alternative to the user who is attempting to find the wording of a desired concept. Suppose a patient had a filling defect in the left colon on barium enema. If the INTERNIST-I user does not know the wording for that finding in the KB, he could refer to the list of

THYROID FUNCTION TEST <S>

TSH STIMULATION TEST UNRESPONSIVE
 TSH STIMULATION TEST RESPONSIVE
 TSH SERUM INCREASED
 TSH SERUM DECREASED
 THYROID RADIOIODINE UPTAKE INCREASED
 THYROID RADIOIODINE UPTAKE DECREASED
 THYROID RADIOIODINE SCAN IRREGULAR UPTAKE
 THYROID PERCHLORATE DISCHARGE TEST ABNORMAL
 T4 TOTAL SERUM INCREASED
 T4 TOTAL SERUM DECREASED
 T4 FREE SERUM INCREASED
 T4 FREE SERUM DECREASED
 T4 BINDING GLOBULIN INCREASED
 T3 SUPPRESSION TEST NON SUPPRESSION
 T3 RESIN UPTAKE INCREASED
 T3 RESIN UPTAKE DECREASED
 T3 SERUM INCREASED
 T3 SERUM DECREASED
 CALCITONIN PLASMA INCREASED

FIG. 3. An example of the group THYROID FUNCTION TEST (S).

approximately 300 GROUP names. He would then quickly find the GROUP BARIUM ENEMA, and note that the appropriate finding name is COLON BARIUM ENEMA FILLING DEFECT LEFT COLON.

5. PROPERTIES WHICH EXPRESS INTERDEPENDENCIES AMONG FINDINGS

The INTERNIST-I KB contains nearly 4000 individual manifestations, many of which are interrelated. The second major class of *properties* is the largest and most complex (Fig. 2). These *properties* affect the behavior of the system by:

- a. influencing the order in which questions are asked by the diagnostic system;
- b. suppressing certain questions from being asked; and
- c. allowing information to be inferred about one finding from the presence or absence of another.

Table 3 shows the distribution of these *properties* in the KB and Table 4 provides an explanation and example of each class of these *properties*.

5.1. Properties Which Affect the Order of Questioning

The diagnostic system selects which questions to ask based on the ability of the question to discriminate among competing hypotheses. The basic order in which questions are asked depends on the TYPE of a finding as mentioned earlier. PRIORS, PREREQS, and UNREQs are *properties* which also influence the order of questioning. These are *properties* of a manifestation (finding 'A') which require the system to determine the status of another manifestation (finding 'B') before asking about finding 'A' (Fig. 2). When a PRIOR relates two findings, finding 'A' will be asked regardless of the presence or absence of

TABLE 3
DISTRIBUTION OF PROPERTIES
WHICH SHOW
INTERDEPENDENCIES IN THE
INTERNIST-I KB

Property	Total number in the INTERNIST-I KB
PRIOR	673
PREREQ	1271
UNREQ	486
TIFT	1248
FIFT	7219
FIFF	194
YFRY	344
YFRN	667
NFRY	154
NFRN	69

TABLE 4
PROPERTIES WHICH SHOW INTERDEPENDENCIES BETWEEN TWO GIVEN FINDINGS

Property syntax	Interpretation	Example
Finding 'A'	Finding 'B' must be asked or already known before asking about Finding 'A'	FERRITIN SERUM LESS THAN 10 NG PER ML PRIOR IRON SERUM LESS THAN 45
PRIOR Finding 'B'		
Finding 'A'	Finding 'B' must be asked or already known <i>and</i> must be present before asking Finding 'A'	ABORTION RECENT HX PREREQ SEX FEMALE
PREREQ Finding 'B'		
Finding 'A'	Finding 'B' must be asked or already known <i>and</i> must be absent before asking Finding 'A'	LIVER BIOPSY PIECEMEAL NECROSIS UNREQ PLATELETS LESS THAN 50000
UNREQ Finding 'B'		
Finding 'A'	Blocks asking about Finding 'A' if Finding 'B' is present	EKG VENTRICULAR ARRHYTHMA PRESENT TIFT EKG VENTRICULAR FIBRILLATION
TIFT Finding 'B'		
Finding 'A'	Blocks asking about Finding 'A' if Finding 'B' is present	HEMATOCRIT BLOOD LESS THAN 35 FIFT HEMATOCRIT BLOOD GTR THAN 50
FIFT Finding 'B'		
Finding 'A'	Blocks asking about Finding 'A' if Finding 'B' is absent	ABDOMEN XRAY MEGACOLON FIFF COLON BARIUM ENEMA DILATATION
FIFF Finding 'B'		
Finding 'A'	Blocks asking about Finding 'B' and assumes it to be present if Finding 'A' is present	BREAST (S) GYNECOMASTIA YFRY SEX MALE
YFRY Finding 'B'		
Finding 'A'	Blocks asking about Finding 'B' and assumes it to be absent if Finding 'A' is present	EMACIATION YFRN TRUNCAL OBESITY
YFRN Finding 'B'		
Finding 'A'	Blocks asking about Finding 'B' and assumes it to be present if Finding 'A' is absent	CSF XANTHOCHROMIA NFRY CSF OBTAINED
NFRY Finding 'B'		
Finding 'A'	Blocks asking about Finding 'B' and assumes it to be absent if Finding 'A' is absent	CHEST XRAY LUNG FIELD (S) ABNORMAL NFRN CHEST XRAY PNEUMATOCOELE (S)
NFRN Finding 'B'		

finding 'B.' In contrast, PREREQs and UNREQs (*requisites*) are conditional; the presence or absence of finding 'B' determines whether finding 'A' is ever asked.

5.1.1. PRIORS

(1) PRIORs help the system direct questioning about laboratory tests in a clinically logical and cost-effective manner. As mentioned earlier, TYPEs serve this purpose as well. However, PRIORs can impart an order of precedence to lab tests *within* a given TYPE. For example, both an abdominal plain film and an oral cholecystogram are of TYPE LAB1. However, in the workup of suspected cholelithiasis, a clinician would first want to know whether a plain film of the abdomen had demonstrated gallstones before ordering an oral cholecystogram. Thus the following *property* exists in the KB:

GALLBLADDER ORAL CHOLECYSTOGRAPHY ROUND FILLING
DEFECT (S)
PRIOR ABDOMEN XRAY GALLSTONE (S)

(2) Because INTERNIST-I findings are nonparsed text strings, many separate findings had to be created which represent a spectrum of deviations from normal. For example,

SGOT BLOOD 40 TO 119
SGOT BLOOD 120 TO 400
SGOT BLOOD GTR THAN 400

PRIORs can be used to force the system to ask about the smaller deviation before larger ones (when appropriate), e.g.,

PRESSURE ARTERIAL DIASTOLIC GTR THAN 125
PRIOR PRESSURE ARTERIAL DIASTOLIC 95 TO 125.

5.1.2. Requisites

Requisites are divided into two groups: *prerequisites* (PREREQs) and *unrequisites* (UNREQs). Most *requisites* are simple; that is, a single condition must be satisfied before asking about a given finding. However, some *requisites* are logically complex. Both PREREQs and UNREQs can be complex *requisites*. By "complex," we mean that a series of conditions, related by the propositional logic "OR" and "AND" operators, have to be met before the diagnostic program is allowed to ask about a given finding. Consider a group of findings: 'B,' 'C,' 'D,' and 'E' which are related to finding 'A' in the following way: either 'B,' 'C,' or 'D' and in addition 'E' must be present in the patient before one would want to determine the status of 'A.' In displaying complex *requisites*, the symbols '(' and ')' are used to enclose those findings related by the logical 'OR' operator. By default, *requisites* listed as separate entries on a *property* list are assumed to be tied together by the logical 'AND' operator. Thus, the example given above would be listed as

FINDING 'A'
 <PREREQ FINDING 'B'
 PREREQ FINDING 'C'
 >PREREQ FINDING 'D'
 PREREQ FINDING 'E'

and translates into: Before asking about finding 'A,' the program must establish the presence of ('B,' 'C,' OR 'D') AND 'E.'

5.1.2.1. *PREREQS*. PREREQs are *properties* of a manifestation (finding 'A') which force the system to ascertain that condition 'B' is present in the patient before asking about finding 'A.' If 'B' is not present, 'A' will not be asked. As noted earlier, in simple PREREQs 'B' is a finding, but in complex PREREQs 'B' is a series of logically related findings.

(1) Some findings are refinements of a more general concept. Therefore, aggregate findings were created which summarize the multiple specific findings comprising the general concept. These "super" findings may subsume as many as 60 separate findings, as in the case of the finding CHEST XRAY LUNG FIELD (S) ABNORMAL. The system must ascertain the presence or absence of the general concept before asking about the more specific findings. If an aggregate finding were absent, the system, by using a PREREQ *property*, could be inhibited from asking specific (and redundant) questions. For example, if a patient does not have abdominal pain, then INTERNIST-I will not ask if there is right lower quadrant pain:

ABDOMEN PAIN RIGHT LOWER QUADRANT
 PREREQ ABDOMEN PAIN PRESENT.

(2) PREREQs help represent common sense knowledge about patients. There are certain physiologically impossible combinations of findings. Many of these are obvious even to a layperson, but not to a computer program. One should only ask if the patient has had a recent abortion if the patient is a female:

ABORTION RECENT HX
 PREREQ SEX FEMALE.

(3) PREREQs can inform the program about the practice of good medicine. Usually lymph nodes should be enlarged before being considered for biopsy:

LYMPH NODE BIOPSY EXTRAMEDULLARY HEMATOPOIESIS
 PREREQ LYMPH NODE (S) ENLARGED.

(4) Anatomical relationships are not directly represented in the INTERNIST-I KB. PREREQs can be used to make the system's behavior appropriate when important anatomical relationships exist between findings. For example, only pains from certain abdominal sites are known to radiate to the scapular region:

ABDOMEN PAIN RADIATING TO SCAPULAR AREA
 <PREREQ ABDOMEN PAIN RIGHT UPPER QUADRANT
 PREREQ ABDOMEN PAIN LEFT UPPER QUADRANT
 >PREREQ ABDOMEN PAIN EPIGASTRIUM.

(5) PREREQs can force the system to first ask if a procedure was performed, before asking about findings which are obtained by the procedure:

ASCITIC FLUID AMYLASE INCREASED
 PREREQ ASCITIC FLUID OBTAINED BY PARACENTESIS.

5.1.2.2. *UNREQS.* UNREQs are *properties* of a manifestation (finding 'A') which force the system to ascertain that condition 'B' is absent in the patient before asking about finding 'A.' These are useful in the following situations.

(1) UNREQs make the system responsive to conditions which may interfere with the interpretation of specific laboratory results. For example, if a patient is severely azotemic, then one cannot meaningfully interpret tests which rely on urinary excretion:

CALCIUM URINE DECREASED
 UNREQ UREA NITROGEN BLOOD GTR THAN 100
 UNREQ UREA NITROGEN BLOOD 60 TO 100.

Or, for example, after a splenectomy most patients will have a thrombocytosis, because the spleen no longer sequesters platelets. It would therefore be difficult to interpret thrombocytosis in a patient who has had a splenectomy:

PLATELETS GTR THAN 450000
 UNREQ SPLENECTOMY HX.

(2) UNREQs also provide the system with common sense information. It would be embarrassing to ask about asphasia if the patient is comatose:

APHASIA
 UNREQ COMA.

(3) UNREQs, like PREREQs, can help the system to behave in a manner consistent with the practice of good medicine. UNREQs can prevent the system from asking about findings when there are medical contraindications for eliciting them. For example, a liver biopsy is dangerous to perform in patients with severe thrombocytopenia or a significant bleeding disorder:

LIVER BIOPSY PIECEMEAL NECROSIS
 UNREQ PLATELETS LESS THAN 50000
 UNREQ PROTHROMBIN TIME INCREASED.

5.2. *Properties Which Permit the System to Make Inferences*

The last two major categories—IFs and FORCEs—allow the system to infer the status of a finding based on knowledge about other findings (Fig. 2). The major difference between IFs and FORCEs lies in how the system uses the inferred information. IFs block the diagnostic program from asking about a

finding based on the presence or absence of another finding. The blocked finding is tentatively marked as either present or absent, but does not participate in the scoring of diagnostic hypotheses. On the other hand, through a *FORCE property*, the system records the status of a finding based on the presence or absence of another finding and scores the 'forced' finding as if it were independently entered by the user.

5.2.1. IFS

There are three different classes of IFs—TIFTs, FIFTs, and FIFFs. There are no "complex" IFs or FORCEs as there are for *requisites*.

5.2.1.1. *TIFT—True if true*. Finding 'A' TIFT finding 'B' indicates that the system should not ask about the presence of finding 'A' if finding 'B' is present. If so, finding 'A' becomes unavailable for scoring but is marked internally as 'tentatively positive' for use with respect to other *properties*.

(1) Aggregate findings are assumed to be present (and thus not the subject of future queries) if their more specific instances are known to be present. If a patient is known to have right lower quadrant pain, the system assumes that abdominal pain is present:

ABDOMEN PAIN PRESENT

TIFT ABDOMEN PAIN RIGHT LOWER QUADRANT.

Through TIFT *properties*, INTERNIST-I can tentatively record the aggregate findings as present, when specific instances are observed. This satisfies any PREREQ condition which requires the system to ask about an aggregate finding before asking about those findings which it subsumes. In the case of the above example, the system could now ask other questions about abdominal pain.

(2) TIFTs can prevent INTERNIST-I from for obtaining redundant information about findings physiologically related to observed findings. For example, if it is known that the conjugated bilirubin is elevated in the blood, then the system tentatively assumes that bilirubinuria exists and will not ask about it:

BILIRUBIN URINE PRESENT

TIFT BILIRUBIN BLOOD CONJUGATED INCREASED.

(3) Because finding are represented as integers internally in the INTERNIST-I program, the system is not capable of 'knowing' that two manifestations represent the same abnormality obtained by separate methods. Therefore, TIFTs are useful in suppressing redundant questions of this class:

KIDNEY <S> IVP URETER <S> OBSTRUCTED

TIFT KIDNEY <S> RETROGRADE PYELOGRAPHY URETER <S>
OBSTRUCTED.

(4) TIFTs can be used to indicate that some methods are more reliable than others for obtaining a given finding. For example, a small kidney may be seen

by ultrasonography or by intravenous pyelography (IVP). However, an ultrasound is more accurate than an IVP in demonstrating renal size. Therefore, if a small kidney was seen on an ultrasound, there would be no need to determine the same finding via an IVP:

KIDNEY <S> IVP SMALL UNILATERAL

TIFT KIDNEY <S> ULTRASONOGRAPHY SMALL UNILATERAL.

5.2.1.2. *FIFT—False if true.* 'A' FIFT 'B' indicates to the INTERNIST-I program to block the asking of manifestation 'A,' and mark it as tentatively absent, if manifestation 'B' is present (again, the 'absence' of finding 'A' is not used in scoring hypotheses). This is useful in several instances.

(1) FIFTs can prevent the system from asking about findings which cannot be simultaneously present in a patient. However, some findings which cannot be present at one point in time may individually occur in a patient over a longer time interval. By only marking a finding as tentatively absent via FIFTs, the program leaves room for the user to enter information that would otherwise seem contradictory. For example, a patient's white blood cell count (WBC) can vary considerably over time:

WBC LESS THAN 4000

FIFT WBC 4000 TO 13500 PERCENTAGE NEUTROPHIL <S>
INCREASED.

(2) FIFTs can be used in combination with GROUPs to streamline the question-asking process. When prompted with 'Please Enter Findings of . . .', followed with a GROUP name, the user is expected to enter all the pertinent positive and negative findings within the GROUP. The system can be made to assume, through FIFT *properties*, that the findings in that GROUP that were not entered are absent. For example, when prompted with 'Please Enter Findings of BLOOD CULTURE,' the user is expected to enter all positive blood culture results. If the user enters that the blood culture was positive for *Escherichia coli*, INTERNIST-I will not ask later if the blood culture was also positive for *Staphylococcus aureus*:

BLOOD CULTURE STAPH AUREUS

FIFT BLOOD CULTURE E COLI.

5.2.1.3. *FIFF—False if false.* Finding 'A' FIFF finding 'B' means that INTERNIST-I should block asking about finding 'A,' marking it tentatively negative, if finding 'B' is negative.

(1) FIFFs, similar to TIFTs, can express dependencies between physiologically related findings. For example, the central venous pressure closely parallels the right atrial pressure. Thus:

PRESSURE VENOUS CENTRAL GTR THAN 10

FIFF HEART CATHETERIZATION RIGHT ATRIUM MEAN
PRESSURE INCREASED.

Similarly, it is unusual to have a thrill without a murmur. Therefore, if a patient does not have a heart murmur in the left second interspace, then the system assumes that there is not a thrill in that position:

HEART THRILL SYSTOLIC SECOND LEFT INTERSPACE
 FIFF HEART MURMUR SYSTOLIC EJECTION SECOND LEFT
 INTERSPACE.

(2) FIFFs inhibit the INTERNIST-I program from asking about subordinate findings when a closely related general finding is known to be absent. For example,

HEART MURMUR HOLOSYSTOLIC APICAL TRANSMITTED TO
 AORTIC AREA (S)
 FIFF HEART MURMUR HOLOSYSTOLIC APICAL.

5.2.2. *FORCES*

As has been seen, IFs allow the diagnostic program to infer the status of a finding from the presence or absence of other findings. The system can use this information to satisfy PRIOR or REQUISITE preconditions. However, as mentioned earlier, IFs only tentatively record the information, which can be overridden at a later point in the case analysis. FORCES also allow the system to infer the presence or absence of a finding; however, the finding is actually scored it as if it were entered by the user. There are four classes of FORCES—YFRYs, YFRNs, NFRYs, and NFRNs.

5.2.2.1. *YFRY—yes FORCES yes*. Finding 'A' YFRY finding 'B' means that if 'A' is present, then 'B' should also be registered as present just as if 'B' were entered by the user. This *property* is useful in the following circumstances.

(1) Implied preconditions can be represented using YFRY. For example, if the program learns that a patient has amenorrhea, it concludes that the patient is a female. Similarly, if it is told that a CSF culture was positive for M. Tuberculosis, it concludes that CSF was obtained. Thus:

AMENORRHEA
 YFRY SEX FEMALE
 CSF CULTURE MYCOBACTERIUM TUBERCULOSIS
 YFRY CSF OBTAINED.

(2) YFRY's can be used to make the system conclude the presence of an aggregate finding if one of its subordinate findings is present:

METHOTREXATE THERAPY RECENT HX
 YFRY CYTOTOXIC DRUG ADMINISTRATION RECENT HX.

(3) YFRYs provide another way to represent important anatomical relationships. If a patient is stated to have a cervical sensory level, then it should be concluded that sensation is decreased in both arms and legs:

SENSORY LEVEL CERVICAL

YFRY ARM ⟨S⟩ SENSATION DECREASED BILATERAL

YFRY LEG ⟨S⟩ SENSATION DECREASED BILATERAL.

5.2.2.2. *YFRN—Yes FORCES no.* ‘A’ YFRN ‘B’ indicates that if manifestation ‘A’ is positive, then manifestation ‘B’ is negative. The *property* is useful in assigning appropriate values to mutually exclusive findings. For example, if a patient is known to be amenorrheic, then the system knows that she does not have menorrhagia:

AMENORRHEA

YFRN MENORRHAGIA.

IF and FORCE *properties* both handle findings which are mutually exclusive. Careful attention to the relationship of findings over time is required when deciding which *property* to use. IFs can and should be used to block asking about two findings which are mutually exclusive at a given point in time, which might both occur in the same patient at different times. For example, the findings WBC LESS THAN 40000 and WBC 4000 TO 13500 PERCENTAGE NEUTROPHIL ⟨S⟩ INCREASED cannot occur simultaneously in the same patient, but may both occur at different times over a period of days. In contrast, FORCES are used to represent mutually exclusive relationships that are more permanent. The findings AMENORRHEA and MENORRHAGIA have a YFRN *property* between them since it is unlikely that the two would coexist during a time interval such as the duration of a hospital admission.

5.2.2.3. *NFRY—No FORCES yes.* ‘A’ NFRY ‘B’ means that if finding ‘A’ is negative, then finding ‘B’ is positive. All instances of NFRY involve findings in which there is an implied precondition. If the system is told that the amylase in the ascitic fluid was not increased, it can conclude that a paracentesis was performed:

ASCITIC FLUID AMYLASE INCREASED

NFRY ASCITIC FLUID OBTAINED BY PARACENTESIS.

Similar to TIFTs and YFRYs, the NFRY *property* is useful in situations where the PREREQ condition requires the system to ask about an aggregate finding before asking about the findings which it subsumes.

5.2.2.4. *NFRN—No FORCES no.* ‘A’ NFRN ‘B’ forces finding ‘B’ to be negative when finding ‘A’ is negative. NFRNs can allow the system to score subordinate findings negatively if their aggregate finding is absent. For example, if the system knows that the lung fields are normal on CXR, then all the chest X-ray findings which involve the lung fields are scored as being absent:

CHEST XRAY LUNG FIELD (S) ABNORMAL
NFRN CHEST XRAY CAVITY (IES)
NFRN CHEST XRAY COLLAPSED LOBE

⋮

NFRN CHEST XRAY UPPER LUNG FIELD (S) DENSITY
(IES).

6. INTERNIST-I'S PROPERTIES AND OTHER EXPERT SYSTEMS

Many of the problems addressed by INTERNIST-I's *properties* are unique to its knowledge representation scheme. Two aspects of INTERNIST-I are responsible for its need for *properties*:

- a. representation of findings as nonparsed text strings, and
- b. the comprehensive but undifferentiated nature of the disease profiles.

Other medical expert systems incorporate much of the information encoded in INTERNIST-I *properties* in their knowledge bases through alternative representation schemes. We review here several systems—PIP, CASNET/Glaucoma, EXPERT, and MYCIN—focusing on how they handle many of the concepts embodied in INTERNIST-I *properties*.

PIP (Present Illness Program) was designed by Pauker *et al.* to take a present illness from a patient with renal disease (3). PIP uses “frames” to represent knowledge about diseases, clinical, and physiologic states, and individual findings. Because PIP findings are represented as frames, they embody more information than INTERNIST-I findings. Unlike INTERNIST-I findings, which can only have a single value (present, absent, or unknown), PIP finding frames allow findings to be multivalued. Finding frames have slots for items which qualify the finding. For example, the PIP finding frame for ‘edema’ has two slots—‘time’ and ‘status.’ The slot for ‘time’ can have multiple values for a given patient case whereas the ‘status’ slot is only allowed one value. This representation scheme allows the PIP program to handle mutually exclusive states. Thus, PIP “knows” that *edema with time past and status present* is compatible with *edema with time now and status present*. However, *edema with time now and status present* is not compatible with *edema with time now and status absent*. PIP’s frame representation allows the system to capture many different aspects of the finding in one place, thereby reducing the total number of findings and the need for links (*properties*) between findings (4).

The format of the disease frames in PIP also eliminates the need for certain INTERNIST-I *properties*. The PIP frames contain only classical patient findings. In contrast, the INTERNIST-I disease profiles are comprehensive in nature and include many findings that are relatively nonspecific for a disease. For example, the PIP frame for the nephrotic syndrome includes only seven findings all of which are characteristic for that syndrome. The INTERNIST-I profile for the NEPHROTIC SYNDROME has 74 manifestations. Table 5 compares the PIP findings for the nephrotic syndrome with their most closely

TABLE 5

COMPARISON OF PIP AND INTERNIST-I FINDINGS FOR NEPHROTIC SYNDROME

PIP finding	Corresponding INTERNIST-I finding	Evoking strength	Frequency	Import
Low serum albumin	ALBUMIN SERUM DECREASED	2	5	3
Heavy proteinuria	PROTEINURIA	2	5	3
>5 g/24 hr proteinuria	PROTEINURIA GTR THAN 3 GRAM (S) PER DAY	3	5	5
Massive symmetrical edema	LEG (S) EDEMA BILATERAL MASSIVE	2	2	5
Either facial or peri-orbital	EYE (S) EDEMA PERIORBITAL	2	3	3
and symmetrical edema	LEG (S) EDEMA BILATERAL SLIGHT OR MODERATE	2	5	3
High serum cholesterol	CHOLESTEROL SERUM INCREASED	1	4	2
Urine lipids present	URINE SEDIMENT FAT BODIES	3	5	3

corresponding INTERNIST-I manifestations (with evoking strengths, frequencies and IMPORT values). Generally, the findings which were chosen to be included in the PIP nephrotic syndrome frame are those which have high evoking strengths,² frequencies, and imports in the INTERNIST-I KB. Thus, PIP does not have to deal with as many findings with limited discriminatory power and low clinical significance as does INTERNIST-I.

CASNET/Glaucoma is an expert system based on a "causal association network" whose domain is the diagnosis and treatment of glaucoma. Several of the concepts incorporated as INTERNIST-I *properties* are included in the CASNET knowledge base. CASNET observations or findings are logically related by observational constraints. For example, the CASNET program "knows" that the "cap to disc ratio of the optic nerve head can be determined only if an ophthalmoscopic exam was performed" (5). CASNET incorporates the reliability of different methods for obtaining the same information through confidence measures. EXPERT, an AI system which evolved from experience with CASNET, has been applied to several medical domains including thyroid diseases, rheumatology, and neuro-ophthalmology (6). EXPERT represents a significant portion of its knowledge through production rules which take three forms. One form, the "F-F rules," allow the system to infer information about one finding if another finding is known. These rules are useful after the initial data collection occurs. They propagate information in a purely deterministic logical manner to other findings. For instance, if the patient's gender is known, then through F-F rules, the system infers that all of the findings associated with the opposite sex are FALSE.

MYCIN is a rule-based expert system for diagnosis and treatment of bacteremia and meningitis (7, 8). MYCIN's production rules are IF . . . THEN statements which include clinical parameters both in the premise (IF) portion and in the action (THEN) portion. There were approximately 65 clinical parameters in the original MYCIN database (7). Each clinical parameter definition has an associated set of "properties" used to evaluate the parameter in specific context. These properties include information about:

² An evoking strength is considered relatively specific if its value is 2 or greater.

- a. how the system should prompt the user for the status of the clinical parameter;
- b. what range of values to expect, and if a numerical value is expected, what the default units are;
- c. production rules requiring the clinical parameter;
- d. production rules from which the clinical parameter can be inferred; and,
- e. situations in which certain questions should not be asked.

This information, along with the order of premises in rules, determines the sequence in which MYCIN will pursue a diagnostic workup. MYCIN thus infers the values of clinical parameters in a manner similar to the IF/FORCE *properties* when it evaluates intermediate nodes as it backward chains through its rule base.

INTERNIST-I relies heavily on *properties* to represent important information in its KB. The need for many of the *properties* is unique to INTERNIST-I's knowledge representation scheme. Other medical expert systems more directly incorporate into their representation schemes many of the fundamental concepts embodied in INTERNIST-I *properties*. However, through their modular nature, *properties* have allowed us to incorporate many subtleties of clinical medicine into the INTERNIST-I KB.

7. CONCLUSION

The portion of the INTERNIST-I knowledge base devoted to *properties* has grown disproportionately large over the past decade compared to the number of diseases profiled. Before the extensive use of *properties*, there were problems with the behavior of the INTERNIST-I diagnostic program. The system tended to exasperate the user with many irrelevant or redundant questions. It often would not conclude the correct diagnosis in straightforward cases. *Properties* have increased the efficiency and the accuracy of the INTERNIST-I diagnostic system in the following ways:

(1) *Properties* provide the system with at least some degree of medical common sense. The program's behavior reflects the underlying representation of 'good medical practice.'

(2) Through TYPEs, the system can direct its workup in a cost effective manner, weighing the cost of obtaining certain facts against an understanding of the discriminating power of that information.

(3) FORCEs permit the system to check for logical consistency within a given patient case. If a user enters findings which are mutually exclusive, the system can recognize them as incompatible. It will then prompt the user for confirmation or correction.

(4) *Properties* significantly reduce the number of questions that the system asks the user:

- a. by inferring the presence or absence of certain findings from information already known;

b. by facilitating the creation of aggregate findings which streamline the question-asking process;

c. by assisting in the development of relevant problem domains which efficiently focus question-asking.

(5) *Properties* influence the scoring algorithms in several ways:

a. by allowing the system to automatically score findings whose presence or absence was inferred from already known information;

b. by refining the scoring through the creation of aggregate findings;

c. by preventing redundant scoring for interdependent findings through the reduction of the number of interdependent findings that the program elicits;

d. by directly contributing to the penalties for 'irrelevant hypotheses.'

(6) Through IMPORTs, INTERNIST-I can determine which findings present in the patient's case should be explained and which findings must be considered during the formulation of problem areas.

Expert systems require accurate representation of the available relevant knowledge in order to simulate clinical decision-making more effectively. *Properties* were added to the KB because the original INTERNIST-I disease profiles lacked fundamental information about the interrelationships among findings. Manifestations are not independent entities. They are interconnected outward expressions of disease in an organism of unfathomable complexity.

The *properties* portion of the knowledge base captures many subtleties of clinical medicine. As such, it could become a valuable resource for medical education if it is incorporated into QUICK, a user-friendly access system which transforms the INTERNIST-I disease profiles into an electronic textbook of medicine (9).

Properties are not a panacea for INTERNIST-I. The system as evaluated in 1982 (1) included a full implementation of *properties* as described herein. However, *properties* have provided us with a means of substantially improving the program's diagnostic behavior within the framework of the existing INTERNIST-I algorithms.

ACKNOWLEDGMENTS

Support for the INTERNIST-I project has come from the NIH through grants from the Division of Research Resources (R24 RR 01101) and the National Library of Medicine (R01 LM 03710 and R23 LM03589).

REFERENCES

1. MILLER, R. A., POPLER, H. E., AND MYERS, J. D. INTERNIST-I. An experimental computer-based diagnostic consultant for general internal medicine. *N. Engl. J. Med.* **307**, 468 (1982).
2. POPLER, H. E. Heuristic methods for imposing structure on ill-structured problems: The structuring of medical diagnostics. In "Artificial Intelligence in Medicine" (P. Szolovits, Ed.), pp. 119-190. Westview Press, Boulder, Colo., 1982.
3. PAUKER, S. G., GORRY, G. A., SCHWARTZ, W. B., AND KASSIRER, J. P. Towards the simulation of clinical cognition. *Amer. J. Med.* **60**, 981 (1976).
4. SHERMAN, H. B. "A Comparative Study of Computer-Aided Clinical Diagnosis of Birth Defects." MIT Laboratory for Computer Science TR-283 (1981).

5. WEISS, S. M., KULIKOWSKI, S. A., AMAREL, S., AND SAFIR, A. A model-based method for computer-aided medical decision making. In "Readings in Medical Artificial Intelligence: The First Decade" (W. J. Clancey and E. H. Shortliffe, Eds.), p. 167. Addison-Wesley, Reading, Mass., 1984.
6. KULIKOWSKI, S. A., AND WEISS, S. M. Representation of expert knowledge for consultation: The casnet and expert projects. In "Artificial Intelligence in Medicine" (P. Szolovits, Ed.), pp. 119-190. Westview Press, Boulder, Colo., 1982.
7. SHORTLIFFE, E. H. "Computer-Based Medical Consultations: MYCIN" Artificial Intelligence Series, Elsevier Computer Science Library, New York, 1976.
8. YU, V. L., FAGAN, L. M., WRAITH, S. M., *et al.* Antimicrobial Selection by Computer. *JAMA* **242**, 1279 (1979).
9. FIRST, M. B., SOFFER, L. J., AND MILLER, R. A. QUICK (QUick Index to Caduceus Knowledge): Using the INTERNIST-1/Caduceus Knowledge Base as an Electronic Textbook of Medicine. *Compt. Biomed. Res.* **18**, 137 (1985).