ILIAD as a Patient Case Simulator To Teach Medical Problem Solving

Robert Cundick, PhD, Charles W. Turner, PhD, Michael J. Lincoln, MD, James P. Buchanan, MD
Curtis Anderson, MS, Homer R. Warner, Jr., MS, Omar Bouhaddou, PhD

Department of Medical Informatics, University of Utah School of Medicine;
Salt Lake City Veteran's Administration Medical Center, University of Utah Department of Psychology

Abstract

ILIAD is an expert system written in C for the Macintosh computer. The system operates in two modes: as an expert consultant to teach differential diagnosis and as a knowledge-based patient case simulator to teach and test medical problem solving. This paper describes ILIAD's simulation mode. Since relations between a disease and its manifestations are expressed in probabilistic terms within ILIAD's knowledge base, a wide variety of simulated cases can be generated automatically. The student's performance may then be evaluated by comparing his or her problem solving approach to an "optimal" strategy derived from the ILIAD knowledge base. This tool is especially valuable in providing experience with diseases that students otherwise are unlikely to see during their medical training.

Introduction

An essential feature of health care is the process of sequential problem-solving whereby a provider (e.g., physician) works with a patient to formulate a series of decisions for diagnosis and treatment. If appropriately accomplished, these decisions should result in optimal health outcomes for the patient at an acceptable expenditure of resources. Thus, the education of health care providers in effective problem-solving and subsequent decision-making is of high importance. This paper describes a computer-based system that can assist health care workers in improving their medical problem solving skills. The system generates simulated patient management problems as a resource for developing and maintaining clinical judgment; ultimately, this tool may become a valuable supplement to the education of a wide variety of health care providers.

In the past two decades, educational research on the subject of teaching and evaluating clinical problem solving has demonstrated drawbacks with current methods. [1] This research suggests that previous training models frequently place too much emphasis upon the student's recall of isolated facts. Such educational approaches do not reflect the sequential decision processes required in medical decision making.

In sequential decision making, the usefulness of a single fact depends in part upon which previous facts have been acquired. For example, if a student knew the two isolated facts that "localized rales" and "chest x-ray demonstrating consolidation" are each related to pneumonia, these facts are not very useful in and of themselves. The student's problem solving can be improved if educators show how these findings are related to each other. The student first needs to know that localized rales can occur in patients with lung consolidation and lung consolidation occurs in pneumonia. A chest x-ray can provide all of the information to diagnose the presence of lung consolidation, but rales provides only part of the information required. If the student does not know the results of a chest x-ray, then a finding of localized rales can provide additional information about lung consolidation and pneumonia. If the student already has the results of the chest x-ray, then a finding of localized rales provides no additional information about the presence of either lung consolidation or pneumonia. Thus, the student would not need to know about the finding of localized rales if the chest x-ray existed, but the student could benefit from this finding if the chest x-ray did not exist. In short, the value of a piece of information in the sequential decision process depends upon the order in which it is used.

The work of Elstein and colleagues [2] has blended "insights of decision analysis and cognitive psychology" to develop alternative "models of medical inquiry" which facilitate clinical problem solving education. This approach requires students to actively participate in a series of patient management exercises. As part of the procedure, student judgments are consistently exercised; but they also must receive immediate feedback in order to benefit maximally from their successes and failures. First, this feedback should guide students in their data gathering and synthesis. A second, more important goal of feedback is to help the student understand WHY each step in this process is made. Finally, the feedback should provide students with the consequences of their actions in terms of diagnostic accuracy and prognostic implications for the patient.

Students' actual medical experience depends upon the epidemiologic profile of patients seen in the institution where they are trained. This can result in their exposure being limited to relatively common diseases. Such limitations in education can be overcome by generating simulated patient cases to expand the students' range of experience. These practice cases can mimic many of the decision processes made with real patients, and the computer can provide appropriate feedback about the students' decision processes.

The power of using simulated patient management problems has long been recognized. Early versions were an outgrowth of the "Tab Test" technique developed by the U.S. Army. The technique was introduced to medicine by the National Board of Medical Examiners, and was developed further at the Univ. of Illinois Office of Medical Education Research by Miller, McGuire and associates.[3,4] These "PMP's" (Patient Management Problems), as they were called, provided one of the earliest means of evaluating clinical problem solving. PMP evaluations supplemented other medical education evaluation techniques that focused primarily upon the assessment of factual recall. Subsequently, simulated problems have been widely applied by the National Board of Medical Examiners and most American Specialty Board examinations.

0195-4210/89/0000/0902$01.00 © 1989 SCAMC, Inc.

Unfortunately, the development and use of these methods has been limited due to the labor intensity and expense required to have medical specialists construct such problems. Consequently, after over a decade of effort, the educational establishment has compiled a relatively small and topic-limited pool of simulated patient problems.

This report presents a unique approach to using an expert system's knowledge base to automatically generate "artificial" patient simulations. The patient case simulator and simulation algorithms are first described, followed by discussion on limitations and future implications of patient simulation models. We are optimistic about the potential of this educational tool and believe that the future development of patient simulations can be even more sophisticated, approaching real life experience.

ILIAD: Patient Case Simulator (PCS)

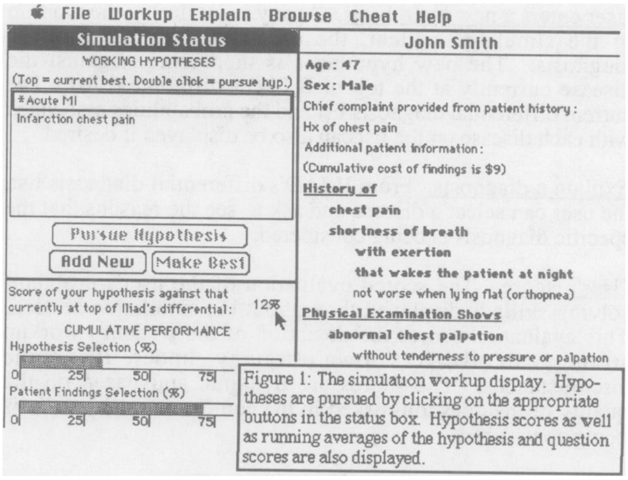
ILIAD is an expert system that can be used in a simulation mode to create hypothetical patient cases for use in teaching medical decision-making skills. The knowledge base used for these simulations is the same one used by ILIAD's consultation mode. Although the two modes share the same knowledge base, they provide very different experiences for the medical student. In the consultation mode, ILIAD accepts data based upon actual patients by prompting the user for relevant patient observations in a decision-driven approach. Interpretations of these observations and the logic behind them are provided.[5]

In the simulation mode, ILIAD generates "synthetic" patient cases from its knowledge base and presents patient findings upon request by the student. As the student continues to gather data, enough information is provided to make a specific diagnostic decision. ILIAD evaluates the diagnostic capabilities of the student by grading each successive decision against the information contained in the knowledge base. Thus, the simulation mode provides a different definition of the roles of the patient case, the medical expert (ILIAD), and the user.

The User Scenario

When a simulation session is initiated, the Patient Case Simulator module (PCS) generates a set of patient observations from a disease frame in ILIAD's knowledge base. The user can request that a simulated case be created from a specific internal medicine discipline such as pulmonary or cardiology, or from the entire knowledge base. Currently, ILIAD can simulate more than 200 diseases based on more than 2000 distinct findings. From the set of candidates (e.g., all pulmonary diseases), the system draws one disease at random in accordance with the disease prevalence. Once a patient case is constructed (as described below), the user is presented with the standard demographic information about the patient (e.g., age, sex) and a chief complaint. The user proceeds to perform a "workup" on the simulated patient by interacting with the "Simulation Status" window (Figure 1):

User enters tentative differential diagnosis list. First, the user indicates the primary hypothesis being pursued. Any plausible competing hypotheses may be entered as well. As part of the patient workup, the user may pose questions either to confirm a primary hypothesis or to rule out a competing hypothesis. At each step, users are asked to indicate which of the two strategies they are pursuing. This information is important in evaluating the quality of each successive decision the students make in their problem solving strategy (described in more detail below). A convenient, "mouse-driven" user interface has been developed so that the user can quickly indicate whether they have switched from one diagnostic hypothesis to another one.



(Top = current best. Double click = pursue hyp.)

*Acute MI

Infarction chest pain

Score of your hypothesis against that currently at top of Iliad's differential: 12%

John Smith

Age: 47

Sex: Male

Chief complaint provided from patient history:

Hx of chest pain

Additional patient information:

(Cumulative cost of findings is $9)

History of

- chest pain
- shortness of breath
  - with exertion
  - that wakes the patient at night
  - not worse when lying flat (orthopnea)

Physical Examination Shows

- abnormal chest palpation
  - without tenderness to pressure or palpation

Figure 1: The simulation workup display. Hypotheses are pursued by clicking on the appropriate buttons in the status box. Hypothesis scores as well as running averages of the hypothesis and question scores are also displayed.

At each step in the process, ILIAD evaluates and scores the "accuracy" of the user's working hypothesis. Based upon the Boolean and Bayesian logic contained in ILIAD, the simulated findings known at each step are processed by the decision logic contained in the knowledge base. These frames compute an aposterior probability for the relative likelihood of each of several diagnoses. These diagnoses include the ones selected by the user. In addition, ILIAD evaluates any diagnoses contained in its knowledge base that use any findings known about the simulated patient at that point in the workup. The computed probability of the user's working hypothesis is compared to the hypothesis ILIAD considers to be best, given what it knows about the simulated patient. At each step, users learn how close their working hypothesis is to the "best" hypothesis estimated from ILIAD's knowledge base.

User poses questions to the system about specific findings. Next, the user poses questions about a specific manifestation that may or may not be present in the simulated case. This is done by entering a keyword that is linked to ILIAD's knowledge base. This keyword generates a list of potential findings that may apply to the patient. For example, the key word "cough" would bring up the phrase "History of cough." The student can pursue this finding for the simulated patient. The knowledge base contains a hierarchically structured list of findings that may apply to a patient. Thus, if a "yes" answer is returned to the query about cough, an additional list of questions is presented. These findings can then be pursued to better understand the nature of the cough. The student can learn about the cough's duration, whether it is associated with chest pain, whether it is productive, etc. The keyword system is important because it permits the user to locate relevant findings without having to know the specific phrasing or terminology used by the system to represent the finding.

For potential quantitative laboratory findings (e.g., SMAC or blood gases), ILIAD provides a number that is either inside or outside the normal range of values for that procedure. (The actual method for creating the findings for simulated patients is described in a section below). Users can also ask ILIAD to identify the finding that provides the most useful information about the patient; this finding is selected according to an algorithm described below.[5] All responses are logged on the patient chart.

Explain patient finding. At any point of the workup, the user may ask for an explanation of a specific patient finding. A ranked list of diseases (from most to least probable) is then displayed which accounts for the specific finding in this patient.

<u>User enters a new hypothesis</u>. At any point during the workup of the simulated patient, the user can alter his working diagnosis. The new hypothesis is then scored against the disease currently at the top of ILIAD's differential list. The current differential diagnosis list and the probabilities associated with each disease on the list can also be displayed if desired.

<u>Explain a diagnosis.</u> From ILIAD's differential diagnosis list, the user can select a disease and ask to see the reasons that the specific diagnosis is being considered.

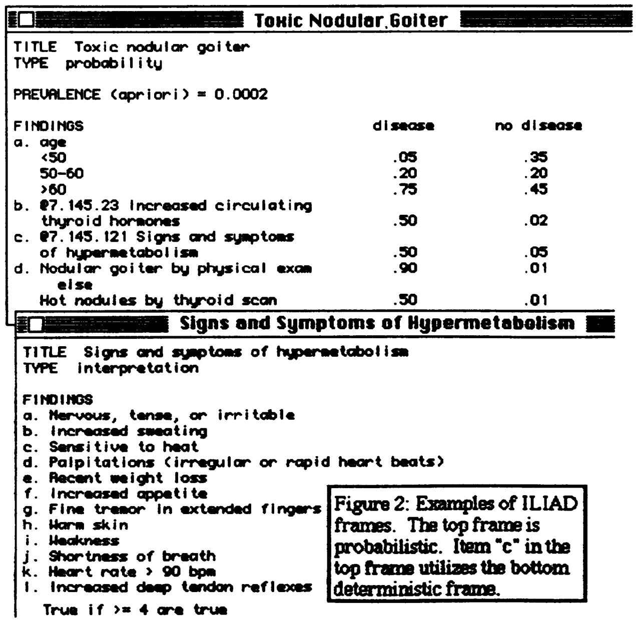
<u>User's score.</u> The scored evaluation of the user's problem-solving skills is displayed throughout the simulation session. This evaluation scores the deviation of the student's workup strategy to ILIAD's own strategy under the same circumstances. The scores are dynamic and based on the quality of the questions asked in the context of the hypothesis being considered.

<u>The Iliad Knowledge Base</u>

ILIAD's knowledge base has been developed and refined with major emphasis on the following considerations:

- representation of a variety of relationships between findings from general to specific (e.g., cough, prolonged cough and cough with blood).
- clustering of related findings with common pathophysiological causes which often are not independent.

The information about a given disease or syndrome is described in a frame format. ILIAD allows for probabilistic as well as deterministic (i.e., If A then B) knowledge representation. Figure 2 shows an example of each. A probabilistic model has a number of features which include estimates of the prevalence of the disease and of the quantitative relationships between the symptoms and the disease (i.e. sensitivities and specificities) according to Bayes Theorem. In deterministic frames, causal knowledge is represented by clustering related findings into nested subdecisions which are then related by a boolean logic statement at the end. The nesting is not so deep as to inhibit the functionality of the system. In the decision frame for "Toxic Nodular Goiter," the manifestation "Signs and Symptoms of Hypermetabolism" is an example of causal information represented as a nested decision (Figure 2).



```
TITLE  Toxic nodular goiter
TYPE  probability

PREVALENCE (apriori) = 0.0002

FINDINGS                                  disease      no disease
a. age
   <50                                      .05           .35
   50-60                                    .20           .20
   >60                                      .75           .45
b. @7.145.23 Increased circulating
   thyroid hormones                         .50           .02
c. @7.145.121 Signs and symptoms
   of hypermetabolism                       .50           .05
d. Nodular goiter by physical exam          .90           .01
     else
   Hot nodules by thyroid scan              .50           .01
```

```
TITLE  Signs and symptoms of hypermetabolism
TYPE  interpretation

FINDINGS
a. Nervous, tense, or irritable
b. Increased sweating
c. Sensitive to heat
d. Palpitations (irregular or rapid heart beats)
e. Recent weight loss
f. Increased appetite
g. Fine tremor in extended fingers
h. Warm skin
i. Weakness
j. Shortness of breath
k. Heart rate > 90 bpm
l. Increased deep tendon reflexes
  True if >= 4 are true
```

Figure 2: Examples of ILIAD frames. The top frame is probabilistic. Item "c" in the top frame utilizes the bottom deterministic frame.

Other structures are available to the knowledge engineers to insure independence of findings that are clearly related.[5] For example, if several findings describe the same information with more or less specificity, the knowledge engineer can use an "else" statement to make sure that only one of the findings is used by the program in calculating its differential diagnosis list. For example, in the decision frame for "Pulmonary Neoplasm," the manifestations: "Cough," "Prolonged daily cough" and "Hemoptysis" are arranged in an "else" statement.

A browse option is also available to the user at any point during a simulation session. This option permits the user to explore the content of the disease frames and their nested clusters.

<u>The Inference Engine (PCS)</u>

The PCS creates "artificial" cases from ILIAD's knowledge base using a random number generator, the observed frequency of an illness, and the observed frequencies associated to each manifestation of the illness. The operations performed by the PCS are as follows:

<u>Selecting a Disease Name</u>

ILIAD maintains a table that maps a finding to the diseases that make use of it. From this table, the PCS retrieves the list of diseases associated to the user-specified medical subspeciality and one disease is selected at random according to the prevalence of each disease. This approach causes the more frequent diseases to be selected more often but ILIAD records the user ID to avoid presenting cases from the same diseases to the same user.

<u>Assigning a Status to the Manifestations of the Selected Disease</u>

ILIAD attributes, at random, a status (e.g., present or absent) to each manifestation of the selected disease; the probability of a positive status is based on the estimated frequency (i.e., sensitivity) of the manifestation. For example, Figure 2 shows the disease frame for "Toxic Nodular Goiter" where the observed frequency of the manifestation "Signs and Symptoms of Hypermetabolism" is 0.50. If the random number drawn from a uniform distribution between 0 and 1 is below 0.50 then the status of "Signs and Symptoms of Hypermetabolism" is set to present, otherwise the status is set to absent. The age and sex attributes are treated differently. In the example described in Figure 2, the age distribution of a group of patients with "Toxic Nodular Goiter" is as follows:

<u>bin value</u>	<u>bin frequency</u>
age less than 50	.05
between 50-60	.20
over 60	.75

A random number is generated from a distribution adjusted to the above empirical distribution. This approach is used for any multiple bin variable (e.g., Haptoglobin values in Hemolytic Anemia patients are modeled with different sensitivities for haptoglobin values less than 20, between 20-27, between 27-250 and over 250.)

Another approach is used when two or more manifestations have been identified to be related within the disease process.

The syntax of the knowledge representation model for these related manifestations is the "else" structure. For example, in the decision frame for "Pulmonary Neoplasm" the manifestations: "Cough," "Prolonged daily cough" and "Hemoptysis" are arranged in an "else" statement since they are related terms. This structure is efficiently used by the PCS for consistency among patient observations. When a knowledge frame is "compiled," the elements of an "else" structure are ordered from least specific to most specific and the PCS will observe that hierarchy. In the example above, the PCS will first attribute a status to Cough based on the incidence of Cough in patients with pulmonary neoplasm. Then, if the status of Cough is set to be present, the PCS assigns a status to the next element of the hierarchy (i.e., Prolonged daily cough). In other situations, these constructs are used with similar items using different measurement methods (e.g., x-ray, ultrasound, biopsy).

The PCS handles clusters in a recursive way. Given the sensitivity of the cluster in the disease frame, a status is assigned to the cluster (e.g., true or false). Then a status is assigned to each element of the clusters based on the role of these different elements. The weighting scheme is derived from the conclusion or truth statement of the cluster using the following principles:

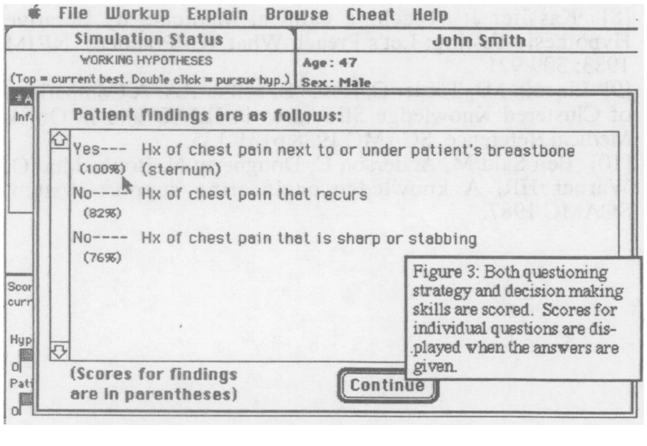
Truth statement: If a AND b then weight (a) = weight (b) = 1/2.

Truth statement: If a OR b then weight (a) = weight (b) = 1.

If the asked manifestation is not in the disease frame, the PCS generates a value based on the incidence of that manifestation in the general population. These general frequencies are stored in ILIAD's dictionary with each manifestation. If the finding is a laboratory test not used by the simulated disease frame, the system chooses a value from the normal range of the test.

Evaluating the User Performance

In these first experiments with the PCS, the assessment of the user clinical competence focuses simply on measuring the efficiency of the user from a physician's performance. Other components of the user competence are under study such as a proficiency component, measured from the results of the user performance noted in decisions made about the patient.[2] The efficiency of a student is evaluated by comparing his strategy to the expert's (ILIAD) strategy when given the same patient observations. Two scores reflect the student's problem-solving performance: a first score evaluates the questioning skills and a second score grades the decision making accuracy (Figure 3).



Patient findings are as follows:

Yes--- Hx of chest pain next to or under patient's breastbone (100%) (sternum)

No---- Hx of chest pain that recurs (82%)

No---- Hx of chest pain that is sharp or stabbing (76%)

(Scores for findings are in parentheses)

Figure 3: Both questioning strategy and decision making skills are scored. Scores for individual questions are displayed when the answers are given.

Each score represents a deviation value between the student's action and ILIAD's action under the same circumstances. When determining the most informative item to collect next about the patient, ILIAD ranks all potential questions based on a utility score.[5] Each time the student asks a question the utility of the question asked is compared to the highest utility to provide a score between 0 and 100:

$$100 \ * \ \frac{\text{utility(question asked by the user)}}{\text{utility(maximum)}}$$

Similarly, the decision making score is evaluated on the basis of the following comparison (deviation):

$$100 \ * \ \frac{\text{probability(user's hypothesis)}}{\text{probability(ILIAD's most likely hypothesis)}}$$

The two scores are made available to the user in the context of the available patient observations (Figure 3).

Evaluating the Simulator

Quality of Simulated Cases

Valuation of knowledge-based simulated cases has been compared to the "Turing test" [6] for decision modeling in an expert system. We recognize that a formal evaluation of the quality of the simulated patient cases is essential, but equal efforts should be allocated to elaborate methods for evaluating a student using the system. The evaluation of the quality of the simulation is based upon an assessment of how well the knowledge base can be used to generate "synthetic" cases. One test of this approach is whether the simulator can generate cases which are similar to real cases. This assessment is being done in two steps. First, a large sample of "artificial" cases from each decision frame are evaluated by a human expert. These individual cases are reviewed for internal consistency and for absurdities (e.g., 68 years old male with complicated pregnancy). A second level of evaluation is performed at an epidemiological validity level: Does the simulator create combinations of findings that commonly occur together in real patients?

Assessment of Medical Student Knowledge

An important assumption of the present approach is that the simulator can improve medical student problem solving skills. One implication of this assumption is that more experienced clinicians should perform better than less experienced clinicians when they use the simulator (without concurrent feedback).

A preliminary test of these assumptions has been completed. The problem solving skills of the medical students were tested at the beginning and the ending of their six week medicine clerkship. Standardized cases were created using the patient simulator. All third year medical students ($\underline{n}$ = 100) at the University of Utah received the same cases presented by the patient simulator (without concurrent feedback). These students attempted to identify the diagnoses of these simulated cases. During the test phase, the students posed questions to the simulator about these cases and received answers about the simulated findings. The student's task was to continue selecting findings until they could make a diagnosis with enough confidence that they would be willing to begin treatment of the patient.

Each finding entered by the student was scored for its information value. In addition, each question was classified as either history, physical exam, or laboratory findings. Finally, each question was coded as reflecting either information to "rule in" a primary diagnosis or to "rule out" a competing diagnosis. The results of this preliminary test indicate that students perform better at the end of their clerkship experience than at the beginning of the rotation. In particular, students were able to pursue history findings that have greater information value at the end of the rotation than they could at the beginning of the rotation. Thus, the findings provide evidence that more knowledgeable individuals achieve higher scores from the patient simulator. Although these results provide tentative support for the reasoning that guided the development of the Patient Simulator, further research is needed to demonstrate the utility of the system as a teaching tool.

Discussion

The ILIAD simulation mode was developed to provide a useful educational tool for medical students and practitioners. It provides a valuable complement to ILIAD's consultant mode. The consultant mode was designed to improve students' problem solving skills with their actual patients. However, the potential for learning from this mode is restricted by the epidemiological mix of patients that students happen to encounter on their wards. A study of actual patient mix on student wards demonstrated that most students experienced a very narrow range of patient diagnoses. From a knowledge base validation point of view, a patient case simulator based on an expert system's knowledge base provides feedback to the knowledge engineers as to the structure and adequacy of the decision models.

Developmental Limitations of Patient Simulations

The overall utility of this simulation approach is limited primarily by the breadth and depth of information in the knowledge base. Another limitation resides in Iliad's present inability to include disease staging, urgency, and time sequencing.

Many clinical patient simulations are limited by the lack of a comprehensive database (including patient history and physical examination data) and by an inadequate case-mix of patients. Combining databases, or transferring modules from different expert systems would help to overcome these limitations, but it will depend upon the ability of such systems to recognize and use current medical terminology. More representative data, such as that collected by the National Center For Health Statistics in their Household Interview Survey, Physical Examination Survey and Ambulatory Medical Care Survey may one day be used directly by our model.

The current medical knowledge is only partly explained in terms of causality and natural history of disease. In addition, use of clusters of knowledge is in an early stage of development at the present time, as applied to expert systems. Mathematical models to derive clusters from primary data findings need to be further developed.[9]

New and efficient knowledge engineering methods and support systems are becoming available to help capture knowledge from multiple sources [7,10] but the process is still largely an empirical one with minimum underlying formalization. An even more sophisticated medical language processing front end to ILIAD will improve the user's ability to state a query in a natural but unambiguous way. At present, ILIAD does not make use of audio or visual signals which might enhance student learning skills. We are currently developing the resources to include these.

Finally, patient simulators must have the capability of reproducing the same health problems that are seen in different stages of their natural history. For example, a patient in an asymptomatic prodromal stage presents a very different problem from one seen at an advanced stage of disease development.

Future Implications of the Patient Simulation Model

In spite of the limitations of our current PCS, we believe that the ILIAD model will facilitate the rapid development of a wide variety of simulated patient management problems for use at the undergraduate, graduate and continuing education levels of professional advancement. These processes build upon the most advanced educational theory related to the learning of clinical problem solving as elucidated by such specialists as Elstein [2] or Kassirer [8]. In its current form, ILIAD can provide a virtually unlimited number and variety of artificial patient cases to assist in medical education.

The quality and usefulness of ILIAD's patient case simulator is currently being evaluated as part of the University of Utah's "Clerkship" project. The "Clerkship" project, supported by the National Library of Medicine (Research Grant: #5 R01 LM-04604-03, "Access to Knowledge Through Models"), is intended to test the hypothesis that new pathways to medical information (such as ILIAD consultation and simulation) can enhance the student's ability to solve problems in an optimized way.

References

[1] Whitman N, Schwenk TL. Problem Solving in Medical Education: Can it be Taught? Current Surgery 1986; 43:453-9.
[2] Elstein AS, Shulman LS, Sprafka S: Medical Problem Solving. Cambridge, Mass: Harvard University Press, 1978.
[3] Williamson, JW; Assessing Clinical Judgment. J. Med.Ed. 1965; 40:180-7.
[4] McGuire C, Babbott D. Simulation Technique in the Measurement of Problem-Solving Skills. J.Educ Meas 1967; 4:1-10.
[5] Warner HR, Haug P, Bouhaddou, O, Lincoln M, Warner HR Jr., Sorenson D, Williamson JW, Fan, C. IlLIAD As An Expert Consultant to Teach Differential Diagnosis. SCAMC 1988, p371-376.
[6] Parker RC, Miller RA. Using Causal Knowledge to Create Simulated Patient Cases: The CPCS Project as an Extension of Internist-1. SCAMC 1987, p473-480.
[7] Sorenson D, Bouhaddou O, Wang W, Ben Said M, Canfield G, Fu LS, Warner HR. Generation and Maintenance of IIAD Medical Knowledge in a HyperCard Environment. SCAMC 1988 p366-370.
[8] Kassirer J: Teaching Clinical Medicine by Iterative Hypothesis Testing; Let's Preach What We Practice. NEJM 1983; 309:921.
[9] Lincoln MJ, Turner C, Hesse B, Miller RA. A Comparison of Clustered Knowledge Structures in ILIAD and in Quick Medical Reference. SCAMC 1988 p131-135.
[10] Ben Said M, Anderson C, Dougherty N, Bouhaddou O, Warner HR. A knowledge engineering support system. SCAMC 1987.