Haug PJ, Warner HR, Clayton PD, et al. A decision-driven system to collect the patient history. *Comput Biomed Res* 1987;20:193-207.

A Decision-Driven System to Collect the Patient History[1]

PETER J. HAUG, HOMER R. WARNER, AND PAUL D. CLAYTON

Department of Medical Informatics, University of Utah, Salt Lake City. Utah 84112

AND

C. DUWAYNE SCHMIDT, JAMES E. PEARL, ROBERT J. FARNEY, AND ROBERT O. CRAPO

Department of Pulmonary Medicine. LDS Hospital. Salt Lake City. Utah 84143

AND

IRENA TOCINO, WILLIAM J. MORRISON, AND PHILIP R. FREDERICK

Department of Radiology. LDS Hospital, Salt Lake City, Utah 84143

Received October 6, 1986

We have developed a computer-administered history designed to directly interview hospitalized patients with pulmonary disease. A frame-based decision system is used to direct the history and to generate a one to five-member differential diagnostic list based only on this history. This system incorporates a cognitive model of question selection and a Bayesian scaring algorithm. Structures to control the choice of questions are embedded in the diagnostic frames and in a QUERY program that makes the final choice of questions. We have compared the behavior of this decision-driven approach with a history taken using a paper questionnaire. The paper-based history presents 182 questions to every patient and captured 75% of 85 pulmonary diseases in its differential lists. The decision-driven system asks 50.7 ± 31.0 (mean ± standard deviation) and captured 74% of 61 pulmonary diseases. Our experience suggests that the use of a computerized diagnostic knowledge base to direct the selection of pertinent questions can substantially reduce the number of questions necessary to collect a diagnostically useful patient history.

INTRODUCTION

We describe a computer system designed to directly interview pulmonary patients, gather a history of present illness (HPI) using a hypothesis-driven model, and provide a plausible set of diagnostic interpretations in the form of a five-member differential diagnostic list. The processes described function in the context of the HELP hospital information system (*1*). They add the historical information collected to the HELP clinical data base.

Hospital information systems (HISs) that capture large amounts of clinical information and communicate this information throughout the hospital are currently available. Some of these systems support computerized decision subsystems capable of providing interpretations and patient-specific suggestions based on the clinical data from the patient record. A notable deficiency in most systems, however, is the lack of a comprehensive, coded patient history. While such a history would be valuable for documentation and communication among those caring for the patient, the advent of decision support systems capable of aiding in the interpretation of

[1] Supported in part by Grant 2-Rl8-HS03810/04, National Center for Health Services Research

such data promises new opportunities for using it to improve patient care.

A variety of alternatives capable of producing a computerized history have been developed. These can largely be divided into paper-based questionnaires (*2-4*), whose results are later computerized, and computer-administered branching questionnaires (*5, 6*). These systems generally have been designed for a screening function. Their histories provide cues for the physician to follow as he takes his own history of the present illness. Its goal is to gather historical information from the patient with enough detail to allow formulation of an initial differential diagnostic list.

This project represents an extension of the work done previously in a system designed to interview patients in the multiphasic screening setting (*7*). Important changes in the structure and underlying philosophy of the current system are discussed below.

The processes described are based upon the tools available in the HELP system. This system is described in detail elsewhere (*1, 8*) and will be discussed only briefly here. In this report we describe the earlier history system mentioned above and contrast it to the current system. In addition, we discuss creation of the diagnostic modules that drive the new system and present the initial results of tests that compare the cognitive approach used in this system to a standard paper questionnaire.

Methods

The HELP System

The core of the HELP system is an online clinical data base that is the primary repository of much of each patient's clinical data. This clinical data base is complemented by a computerized medical decision-support system. It consists of a set of modular decision criteria and a program for processing this logic. The decision tool is used to assist and enhance patient care through computerized interpretation of patient data. The results, in the form of data interpretations and patient-specific suggestions and alerts, are fed back to the physicians to aid in the clinical decision process.

The Original Multiphasic Screening History

The system described here is the descendant of a system in use in the early 1970s to collect a history in the multiphasic screening setting. The earlier system did not use the decision tools but instead was a single-purpose application, driven from a table of statistics. The table contained the sensitivity of each of 320 symptoms for each of 134 diseases. The version of Bayes equation used was

$$P(D_i/S_j) = \frac{P(D_i)P(S_j/D_i)}{\Sigma P(D_k)P(S_j/D_k)} \qquad \text{for } k = 1 \ldots n, \qquad [1]$$

where $P(D_i/S_j)$ is the probability of disease i after symptom j is known, $P(D_i)$ is the probability of disease i before symptom j is known, and $P(S_j/D_k)$ is the sensitivity of symptom j for disease k.

The program had a list of 50 questions which were asked of every patient. Other questions from the set of 320 were chosen dynamically by calculating the most likely diagnoses from the data currently available and using a set of heuristics to select those questions most likely to contribute to the diagnostic process. This approach was used to focus the questioning process by asking those questions which would best elucidate the diseases deemed most likely at any point in the history.

The Revised History System

A revised history gathering system was implemented for this experiment. Rather than relying on a hardcoded, single-purpose application, it uses the standard decision-making tools in the HELP system. The goal

was to develop a modular, frame-based diagnostic system capable of directing the collection of the data required to evaluate the likelihood of a group of diseases.

These diagnostic frames also use a statistical approach. A module was developed for each disease to estimate the probability of a patient having that disease. The patient's answers to questions about relevant symptoms were used as data.

The version of Bayes equation used in this system differs from that described above. It is the standard version supported in the HELP decision system:

$$P(D/S) = \frac{P(D)P(S/D)}{P(D)P(S/D) + P(ND)P(S/ND)}, \qquad [2]$$

where $P(D/S)$ is the probability of a specific disease D in a patient with a specific manifestation S, $P(D)$ is the expected disease probability prior to observation of the manifestation or the a priori probability, $P(S/D)$ is the probability of the manifestation in any patient with the disease, $P(ND)$ is the probability of not having the disease in question (i.e., 1-$P(D)$), and $P(S/ND)$ is the probability of the manifestation in patients who do not have the disease. $P(S/D)$ is the sensitivity of disease D to manifestation S and $P(S/ND)$ is the one's complement of the specificity of manifestation S for disease D (i.e., 1-specificity).

Acute Bronchitis	Histiocytosis X
Asbestosis	Hodgkin's Disease
Aspiration Pneumonia	Influenza
Asthma	Lung Abscess
Bacterial Pneumonia M	etastatic Neoplasm
Bronchiectasis	Non-Hodgkin's Lymphoma
Chronic Bronchitis	Primary Pulmonary Neoplasm
Coal Worker's Pneumoconiosis	Primary Pulmonary Hypertension
Coccidioidomycosis	Pulmonary Embolism
Congestive Heart Failure	Sarcoidosis
Diffuse Idiopathic Fibrosis	Silicosis
Drug Related Pneumonitis	Spontaneous Pneumothorax
Emphysema	Tuberculosis
Goodpasture's Syndrome	Wegner's Granulomatosis
No Pulmonary Disease	

FIG. 1. Diseases for which diagnostic frames were constructed.

Building the New Knowledge Base

To build a knowledge base for this experiment, a group of five physicians was assembled. This group consisted of a specialist in pulmonary medicine, an internist, and three radiologists with special interest in chest radiology. They chose the set of diseases to include in the system and provided a list of the patient data most useful in the diagnosis of these diseases. They also assisted in assigning initial probability estimates for each of the manifestations.

Since a sequential, Bayesian approach to diagnosis was chosen, the knowledge we obtained from the physicians was in the form of probability estimates; an a priori probability for each disease in the inpatient population and a sensitivity and specificity for each manifestation in each disease. These estimates were supplemented by data from the medical literature and by a review of pertinent information derived from the original, hard-coded, Bayesian history system.

A group of 28 diseases was modeled in this way (Fig. 1). An additional diagnostic module was created to explicitly identify patients with no pulmonary disease. Figure 2 is an example of a module for the diagnosis of

pneumonia. These frames formed the knowledge base used in subsequent tests of two approaches to collecting the patient history.

History Collection by Paper Questionnaire

After these diseases had been modeled, two methods of history collection were tested. First, a paper questionnaire was created. This questionnaire was designed specifically to capture the information used in the diagnostic modules. "Yes" or "no" answers reflecting the presence of 182 symptoms were required for the functioning of these modules.

One hundred patients received paper questionnaires designed to collect the historical data required by the diagnostic modules. The patients were selected by reviewing X-ray orders and randomly approaching those patients, not in intensive care units, who were about to have a chest X-ray. Patients willing to, and capable of, filling out the questionnaire received it with instructions to answer the questions as they would have on their day of admission. The data obtained were entered into the patient records.

```
(1)   FRAME 1    =.== PNEUMONIA (HISTORY)

      FINAL EVALUATION:
(2)   A   VAL:  M

      SECTOR LOGIC:
(3)   A   ARITH:      0.014

(4)   B   SEARCH:  A  (A)  HAVE YOU HAD RECENT CHEST PAIN?

      C   SEARCH:      (A)  HAVE YOU HAD A FEVER WITH THIS ILLNESS?

      D   SEARCH:      (A)  HAVE YOU HAD CHILLS WITH THIS ILLNESS?

      E   SEARCH:  A  (A)  HAVE YOU HAD A COUGH WITH THIS ILLNESS?

      F   SEARCH:      (A)  IS YOUR CHEST PAIN INCREASED BY BREATHING DEEPLY?
                         (B) IS YOUR CHEST PAIN INCREASED BY COUGHING?
                 USE ANSWER MAX(A, B)

      G   SEARCH:      (A)  HAVE YOU BEEN SHORT OF BREATH WITH THIS ILLNESS?

      H   SEARCH:      (A)  IS YOUR SPUTUM YELLOW, GREEN OR BROWN?

(5)   I   PROB:      A,   IF ex: C OR D,   USE val: MAX(C, D)
          ANSWER: (N, Y),   TRUE: (0.15, 0.85),   FALSE: (0.7, 0.3)

      J   PROB:      I,   IF ex: E,   USE val: E,   ANSWER: (N, Y)
          TRUE:      (0.1, 0.9),   FALSE: (0.8, 0.2)

      K   PROB:      J,   IF ex: F,   USE val: F,   ANSWER: (N, Y)
          TRUE:      (0.71, 0.29),   FALSE: (0.9, 0.1)

      L   PROB:      K,   IF ex: G,   USE val: G,   ANSWER: (N, Y)
          TRUE:      (0.56, 0.44),   FALSE: (0.87, 0.13)

      M   PROB:      L,   IF ex: H,   USE val: H,   ANSWER: (N, Y)
          TRUE:      (0.35, 0.65),   FALSE: (0.95, 0.05)

(6)   N   ARITH:    IF  M  LT  A  THEN GOTO FINAL EVALUATION

(7)   O   EXIST:     ASK((PATIENT QUESTIONS)C,  D,  E,  F,  G,  H)
```

FIG. 2. Parts of a diagnostic frame for the computer-directed history: (1) frame label; (2) final evaluation slot; (3) a priori probability for this disease; (4) data specification, indicates the questions required to calculate disease likelihood; (5) specifications of statistics (sensitivity and specificity) associated with yes and no answer to referenced question; (6) control logic for ASK function, (7) specification of questions to ask patient.

Each patient chart was then examined to determine the illnesses present at the time of the patient's admission. This determination was based upon the ICD-9 (*9*) codes which are assigned independently to each

patient by the medical records department. This assignment occurs after discharge and is based on the diagnoses listed by the patient's attending physician. If none of the 28 diseases for which we had developed decision logic were present, the patient was designated as having "no pulmonary disease."

After the assignment of one or more diagnoses to each patient, the diagnostic modules were run and the likelihood of each of the 28 diseases was calculated. The likelihoods were ranked and the top 5 diseases were designated as the differential diagnostic list (Fig. 3). The differential diagnostic list was limited to diseases whose likelihood exceeded 1%. The module for "no pulmonary disease" was treated as a 29th possible diagnosis in creating this list.

The diagnostic lists were then compared to the ICD-9-based discharge diagnoses to determine whether the patients pulmonary disease or diseases had been included in the computer-generated, differential diagnostic list. A list was considered correct if it contained the pulmonary disease assigned to the patient by the medical records department. When a patient had more than one pulmonary disease, the accuracy of the computer's diagnostic list was assessed for each disease separately.

The patient history suggests:

.82 Asthma
.56 Pneumonia
.24 Acute Bacterial Bronchitis
.06 Chronic Bronchitis

FIG. 3. A typical computer-generated differential diagnostic list. Differential diagnostic lists consisting of I to 5 diagnoses were generated from the history collected from each patient.

Using the Knowledge Base to Direct the Patient History

The second approach to collecting this information uses the computer's knowledge base to focus and streamline the process. Data are acquired by the computer directly from the patient. The goal is to ask the questions best suited to elaborating the most likely diagnoses.

The process operates in a cyclic fashion. First, a select group of questions is submitted to the patient. When these questions are answered they are used to evaluate the diagnostic frames in the knowledge base. In the process, a new set of questions is selected and again submitted to the patient. These questions represent the systems assessment of the most useful information to request.

Two tools are built into the HELP system to manage this cycle. The first is a process designed to trigger the evaluation of a diagnostic module whenever prespecified historical data are captured and stored. When, for instance, the patient responds "yes" to the question, "Do you have a cough?" the diagnostic module for pneumonia will automatically be processed. A flag, assigned by the creator of the module, indicates which data used for the decision will evoke (data-drive) that decision.

The second tool is a function designed to direct the collection of missing data. It is called the ASK function and takes as arguments (1) a reference to the item of data required and (2) an address field representing the expected source of the data requested (i.e., nurse, attending physician, patient, etc.). This function may be embedded in each of the modules. In this way, the creator of the module can control the data collection process as well as the diagnostic logic. The ASK function is "intelligent" in that it will only request a given piece of information after checking the patient record to be sure the data are, indeed, absent. In this experiment the address field always specified the patient as the data source.

As an example the set of statements

A.^Search: Have you had a cough with this illness?
B. Search: Have you had a fever?
C. Search: Does your cough produce yellow, green, or brown sputum?
D. Exist: ASK((Patient Questions),B,C)

might appear in a module designed to diagnose pneumonia. Statement A contains a datadriven flag ("^") and statement D is a call to the ASK function. If the patient indicates that he has a cough, the module containing these statements is activated by the data-driven method. The answer to "Have you had a fever?" is searched for in the patients clinical data base and if it is absent, statement D sends the question to a buffer to be asked of the patient when he is available at a terminal. The same process is applied to the question from statement C.

An answer to one question may be a data-driver for several modules, in which case a number of additional questions may be sent to the buffer. These questions represent the combined information requirements of the triggered modules. A program called the QUERY DRIVER uses prespecified control logic to restrict and order the asking of these questions. As answers are collected, the QUERY DRIVER sends these answers to the data base for storage and causes reevaluation of the modules that sent the requests.

In practice, the process is cyclic, with the computer asking a group of questions and then using the answers to evaluate the diagnostic modules from which they came, as well as to trigger processing of new modules for which these answers are data-drivers. The new modules contribute questions of their own which may be asked by the QUERY DRIVER during the next questioning cycle. We call this process decision-driven data acquisition.

In some instances, answers to questions can be inferred by the QUERY DRIVER from previous answers. The questions are structured hierarchically in a data dictionary with the upper level a general question branching to lower more specific follow-up questions. A "no" answer to an upper level question implies "no" answers to those below it.

If, for instance, a module requires an answer to "Do you have chest pain made worse by coughing?' the frame's author can set a flag instructing the QUERY DRIVER to attempt to infer the answer to this question. The program would then begin by searching the patient record for the answer to "Do you have chest pain?", a higher level question. If a "no" answer were found, a "no" answer would be inferred to the initial question. A "yes" answer would mean that the initial question still must he asked. If an answer to "Do you have chest pain?" were absent, the QUERY DRIVER would begin by asking it rather than the more specific question.

The control functions described above are all specified through logic in the diagnostic frames. They are therefore controlled by the author of the frame. He additionally has access to a variety of algebraic and Boolean tools to restrict questioning when appropriate. A typical use of these tools occurs when the probability of a disease has dropped below a predefined threshold. Under these circumstances, a disease module will cease addressing further questions to the patient. This is accomplished through a simple statement within the module (item N in Fig. 2).

Two further control strategies are built into the QUERY DRIVER to focus and constrain the questioning process. First, the QUERY DRIVER contains a question selection algorithm to choose which of the questions sent to it to ask. This algorithm is based on the assumption that the most satisfactory history will be collected by attempting to match the data requirements of the most Likely diseases. To perform the selection, the QUERY program sums for each question in the question buffer the likelihoods of the disease frames that sent that question to the question buffer. Thus, questions whose answers would contribute to more than one diagnosis tend to score higher than questions that are used by a single diagnostic hypothesis. The process then compares the totals for each question and selects the top five questions to present to the patient. The answers to these questions may evoke additional hypotheses.

Another algorithm in the QUERY DRIVER is designed to terminate questioning when it appears unlikely that further questioning will significantly alter the probabilities of any of the hypothesis under consideration. This stopping algorithm is invoked after the first 30 questions have been asked and, subsequently, after each

group of 5 questions. The algorithm sums the probabilities of the unexplored diseases that have questions remaining in the question buffer. If the sum of the probabilities of these incompletely explored hypotheses (diagnoses) is less than 0.05, the QUERY program terminates the questioning process.

Thus, two levels of control are incorporated into the computer-directed history system. A local level of control is specified by the frame author who sets the flags specifying which data will trigger each frame and who specifies the logic that can interrupt frame processing prior to the ASK function. A global control structure manipulates question flow and direction after the frames send questions to the question buffer.

Testing of this approach was undertaken in each of two rooms in a medical ward where a terminal was placed at the bedside. The subjects were the patients of three pulmonary medicine specialists who agreed to admit to these rooms. Since it proved impossible to admit all of their patients to these rooms, a mobile terminal was built and used to gather histories from patients in other parts of the hospital. Communication with the central computing system was through the hospital phone system.

Forty-six patients had histories directed by the computer system. In most cases the patient interacted directly with the terminal, but three patients required additional aid. In the case of one blind patient, a technician read the questions posed by the computer and entered the yes and no answers. A second blind patient was assisted by a visiting sister. And a debilitated patient, just released from the intensive care area, required the assistance of his wife to answer the questions.

Accuracy for the computer-generated list of differential diagnoses was again assessed by comparing these lists to the ICD-9 codes representing the patients discharge diagnoses.

TABLE 1

RESULTS OF THE QUESTIONNAIRE-BASED HISTORY

	Patients without pulmonary disease	Patients with pulmonary disease
Number of patients	41	59
Number of pulmonary diseases	0	85
Diseases[a] captured in DD lists	28 (68%)	64 (75%)
Number of questions asked	182	182

[a] Accuracy determined by the presence of the correct diagnosis in the differential diagnostic list (see text). In the case of the patients without pulmonary disease, “no pulmonary disease” was required in the list

RESULTS

Table 1 summarizes the results of the first part of this experiment. Eighty-five pulmonary diseases were identified in the discharge summaries of 100 patients. Of these, 28 were the primary discharge diagnosis and 53 were secondary or complicating diagnoses. No patients had diseases of the lung or pleura for which we did not have diagnostic frames. Forty-one patients were classified as having “no pulmonary disease.”

The diagnostic modules were run against the histories collected by questionnaire from the 100 patients. Sixty-four (75%) of the pulmonary diseases found in this group of patients were present in the diagnostic lists generated by the computer. The mean probability of the captured pulmonary diseases was 0.51 ± 0.18 (mean ± SD) and their mean rank in the differential diagnostic lists was 2.12 ± 1.27. In addition, 28 of 41 (68%) of the patients without pulmonary disease were correctly identified as indicated by the appearance of ”no pulmonary disease” in their diagnostic list. Seventy of 100 (70%) of the diagnostic lists were completely accurate (i.e., contained all pulmonary diagnoses (or “no pulmonary disease” if appropriate) recorded for that patient).

Table 2 shows the results of the diagnostic logic when the computer administered the history. A greater frequency of pulmonary disease was noted in these patients. There were 61 pulmonary diseases recorded

TABLE 2

RESULTS OF COMPUTER ADMINISTERED HISTORY

Number of patients	9	37
Number of pulmonary diseases	0	61
Diseases[a] captured in DD lists	7 (78%)	45 (74%)
Number of questions asked	24.8 ± 16.5	57.0 ± 30.5

[a] Accuracy determined by the presence of the correct diagnosis in the differential diagnostic list (see text). In the case of the patients without pulmonary disease, "no pulmonary disease" was required in the list.

TABLE 3

DETAILED RESULTS OF COMPUTER-ADMINISTERED HISTORY

	Patients without pulmonary disease	Patients with one pulmonary disease	Patients with > one pulmonary disease	Patients with primary pulmonary disease
Number of patients	9	21	16	27
Number of diseases	0	21	40	27[b]
Diseases[a] captured in DD lists	7 (78%)	17 (81%)	28 (70%)	23 (85%)
Number of questions asked	24.8 ± 16.5	52.6 ± 33.2	62.8 ± 26.4	61.1 ± 32.1
Number of questions answered "yes"	8.1 ± 7.9	24.9 ± 20.6	27.5 ± 15.7	28.9 ± 19.6
Number of questions answered "no"	16.7 ± 9.6	27.7 ± 16.1	35.3 ± 16.5	32.3 ± 17.9
Number of "no" answers inferred	14.9 ± 4.9	14.7 ± 7.6	16.8 ± 7.7	14.8 ± 7.5

[a] Accuracy determined by the presence of the correct diagnosis in the differential diagnostic list (see text). In the case of the patients without pulmonary disease. "no pulmonary disease'' was required in the list.
[b] Number reflects only the diseases that were primary discharge diagnoses, i.e., one disease per person.

among 46 patients. Only 9 patients warranted the diagnosis of "no pulmonary disease." One patient had a disease for which we did not have a module. A 31-year-old patient with cystic fibrosis and staphylococcal bronchopneumonia received the history. Although the computer recognized the pneumonia (probability = 0.86), it attempted to model his cystic fibrosis as a combination of asthma and bronchiectasis.

Forty-five of the 61 pulmonary diseases (74%) in this group were present in the diagnostic lists. The mean probability of the captured pulmonary diseases was 0.58 ± 0.40 and their mean rank in the differential diagnostic lists was 2.09 ± 1.37. Seven of nine (78%) of the patients without pulmonary disease had the correct

diagnosis in their computer-generated lists. Thirty-two of 46 (70%) of the lists contained all of the patients' pulmonary diseases.

To better understand the behavior of the questioning process, more detailed analysis was done in this group. Examination of the distribution of pulmonary diseases (Table 3) revealed that the 9 patients without pulmonary illnesses were asked the minimum number of questions by the computer, an average of 24.8 ± 16.5. The QUERY DRIVER used the hierarchical structure of the data to infer 14.9 ± 4.9 of the "no" answers for these patients. These are questions that the patient was not asked.

In 21 patients with a single pulmonary disease, 17 (81%) had that disease represented in their differential lists. They answered 52.6 ± 33.2 of the possible 182 questions. The QUERY DRIVER was able to infer 14.7 ± 7.6 of the "no" answers for these patients.

The 16 patients with more than one pulmonary disease had 28 of their total of 40 diseases (70%) Listed in the differential diagnoses generated by the computer. They answered an average of 62.8 ± 26.4 questions. The QUERY DRIVER inferred 16.8 ± 7.7 of their "no" answers.

And finally, of the 27 patients who had a pulmonary disease as their primary discharge diagnosis, 23 (85%) had this diagnosis identified by the computer. This group responded to 61.1 ± 32.1 questions while the QUERY DRIVER inferred 14.8 ± 7.5 of their "no" answers.

Overall, the number of questions asked by the QUERY DRIVER ranged from 2 to 139 with a mean of 50.7 ± 31.0. The QUERY DRIVER inferred an average of 15.5 ± 7.1 "no" answers for the group.

Discussion

We have used the ability of the computer to generate an accurate differential diagnostic list as a measure of the adequacy of the two history-gathering procedures described above. Because of differences in our methods for selecting participants, the two patient groups are not entirely comparable. Nonetheless, some observations can be made.

The system described markedly reduced the number of questions asked of the patients in each of the patient subgroups; 182 questions were directed to each patient who filled out the questionnaire, while the users of the DDA mode saw an average of 50.7 questions.

This reduction is due, in part, to the inference mechanism described above. In its behavior this mechanism resembles a branching questionnaire. The choice of which lower level questions to ask is dependent on answers to upper level questions. However, of the overall reduction of an average of 129.3 questions, a mean of 15.5 questions were eliminated by this mechanism. The bulk of the reduction in questions asked comes from the process of decision-driven data acquisition. The use of a set of diagnostic hypotheses to condition the questioning process appears effective in capturing relevant history and in reducing the burden to the patient.

Although the computer-administered history asked fewer questions of each patient than did the questionnaire, the diagnostic lists generated from it were as accurate for patients with, and without, pulmonary disease. In addition, when patients were subdivided by the number of pulmonary diseases listed as discharge diagnoses, it was seen that the number of questions asked by the computer was related to the number of pulmonary diseases present. The system collected more information as it elucidated more complex pulmonary presentations. A marked reduction in questions was apparent for patients without pulmonary disease.

An important motivation for testing this new approach to interactive history collection was the availability of powerful modeling tools within the standard HELP decision system. These tools provide several advantages. First, maintenance of the system is enhanced by adopting a modular disease representation. A disease frame can be added to, or removed from, the system without requiring changes to other components of the knowledge base or to the program. A new disease frame, once tested, can be activated simply by indicating which data should trigger consideration of the disease by setting data-driven flags in the frame.

Second, the availability of standard constructs besides the Bayesian algorithms allows more flexibility in modeling diseases. Boolean and algebraic constructs can be used to implement special diagnostic and question

selection strategies. As indicated above, they were used regularly in this experiment to give the frame author local control over the consideration of diseases and the choice of questions to present via the QUERY DRIVER. These same constructs can be used to avoid the inaccuracies that might result if two symptoms are used that are not independent of each other within a disease. In the example, in Fig. 2, the symptoms "chills" and "fever," which are obviously dependent, are combined as a single trait to avoid this problem.

The original system used a version of Bayes' equation assuming the diseases considered to be mutually exclusive. Inherent in its application is the supposition that the patient can have only one of the diseases under consideration. The frame-based system uses a version of Bayes' equation that functions local to the module. It evaluates the likelihood of an illness independent of the other diseases under consideration. In our patient population, and in most inpatient settings, multiple diseases are common.

Physicians appear to gather a patient's history through a process which relies strongly on the early generation of a list of possible diagnoses and on a cyclic process of hypothesis-directed questioning and answer-related hypothesis reevaluation (*10*). Indeed, research in the field of cognitive psychology suggests that physicians often generate their initial list of hypotheses based on the minimal information represented in the patient's age, sex, and chief complaint (*11*). The history-gathering program described here is based on a cognitive model. This is combined with a Bayesian scoring algorithm used to discriminate among the hypotheses being pursued. Although other scoring algorithms could certainly be used, a Bayesian model was chosen because it allows one to improve the system's performance over time, as experience provides better estimates of symptom sensitivities and specificities. We are currently evaluating a data base consisting of patients for whom all the pertinent historical information is stored. Initial results show that a significant improvement in diagnostic accuracy is possible (*12*).

Bayes' equation has been used in experimental systems for medical diagnosis since it was initially suggested in 1959 (*13*) and tested in 1960 (*14*). During this time, a variety of objections have been raised to its application in this context. Nonetheless, some surprisingly effective systems have been developed with it. An example is a Bayesian system described by DeDombal and associates in 1975 that was more accurate than physicians in diagnosing the cause of abdominal pain (*15*). Had the computer's predictions about the occurrence of appendicitis been used, not only would fewer unnecessary appendectomies have been done, but also fewer necessary appendectomies would have been postponed. The Bayesian approach has been used with varying degrees of success in other applications including the classification of stroke (*16*), blood dyscrasias (*17*), and diagnosis of the solitary pulmonary nodule (*18*). Experience at our institution includes applications in the diagnosis of congenital heart disease (*19*), as part of an "intelligent" radiology reporting system (*20*), and in the multiphasic screening setting described above.

Several factors in this project worked to reduce the accuracy of the diagnostic lists generated. The first is an artifact of the use of the discharge diagnoses entered by the medical records department as the standard for comparison. While the history-gathering process was designed to identify active disease, the personnel in medical records tend to include any mentioned pulmonary disease in the listings entered there. As a result, the differential diagnostic lists generated by the computer are being compared to a list of active and inactive illnesses leading to some apparent misses by the computer. To test the accuracy of the system when restricted to active illnesses, the 27 patients with a pulmonary disease recorded as their primary discharge diagnosis were reviewed (Table 3). Of these, 23 (85%) had this disease included as one of the five diseases in their differential diagnostic list.

A more important factor was the accuracy of the sensitivity and specificity estimates used in our disease modules. These diagnostic modules are based largely on estimates of the probabilities relevant in a Bayesian model of diagnosis. As mentioned, our expert panel contributed a large segment of these values since appropriate quantitative information could not be found in the medical literature. Sensitivities are present for many symptom/illness combinations, but appropriate specificities were missing. It has been observed that humans are poor at estimating probabilities (*21*). Use of statistics derived from clinical data bases should improve the system's performance on future cases.

Conclusion

We have described a system that collects a pulmonary history at the patient bedside. It uses an approach based on a cognitive model of the patient interview. A frame-based, Bayesian scoring algorithm is used to rank the diseases under consideration. It appears effective in yielding the multiple diagnoses common in hospitalized pulmonary patients.

We have shown that, when compared to a paper questionnaire, these decision-driven data acquisition tools markedly reduce the number of questions presented to each patient. The reduction in numbers of questions asked does not affect the diagnostic accuracy. This property will become increasingly important as the group of diagnostic modules grows, and with it the number of possible questions for each patient.

An aspect of this project touched upon little in this report is the integration of this system into the HELP hospital information system. Once captured, the history can be made available for use throughout the system and can be analyzed, with other data, to make further medical decisions.

Finally, the techniques involved in decision-driven data acquisition may be useful for collecting other data. Questions concerning the physical exam, for instance, can be generated from diagnostic modules in much the same way as is the computer-directed history. We are adapting the processes described above to direct a small, select group of questions to the physician, or nurse, caring for a patient. In this context, the tools will be used to reduce the respondent's input chores while gathering that information necessary for discriminating among likely diagnoses. Decision-driven data acquisition is a part of a larger effort to explore interfaces between medical decision systems and the people who use them.

References

1. Pryor, T. A., Gardner, R. M., Clayton, P. D., and Warner, H. R. The HELP system. *J. Med. Syst.* **7,** 87-102 (1983).
2. Brodman, K., Erdman, A. J., Jr., Lorge, I., and Wolff, H. G. Cornell medical index: Adjunct to the medical interview. *J. Amer. Med. Assoc.* **140,** 530-534 (1949).
3. Collen, M. F., Cutler, J. L., Siegelaub, A. B., and Cella, R. L. Reliability of a self-administered medical questionaire. *Arch. Intern. Med.* **123,** 664-681 (1969).
4. Rockart, J. F., McLean, E. R., Hershberg, P. I., and Bell, G. O. An automated medical history system. *Arch. Intern. Med.* **l32,** 348-358 (1973).
5. Slack. W. V., Hicks, G. P., Reed. E. R., and Van Cura, L. J. A computer-based medical history system. *N. Engl. J. Med.* **247,** 194-198 (1966).
6. Haessler, H. A., Holland, T., and Elshtain, E. L. Evolution of an Automated Database History. *Arch. Intern. Med.* **l34,** 586-591 (1974).
7. Warner. H. R., Rutherford, B. D., and Houtchens, B. A sequential Bayesian approach to history taking and diagnosis. *Comput. Biomed. Res.* **5,** 256-262 (1972).
8. Warner, H. R. "Computer-Assisted Medical Decision-Making.'' Academic Press, New York, 1979.
9. "International Classification of Disease," 9th rev., Clinical Modification, 2nd ed., U.S. Department of Health and Human Services. Washington, D.C. 1980.
10. Elstein, A. S., Shulman, L. S., and Sprafka, S. A. "Medical Problem Solving: An Analysis of Clinical Reasoning." Harvard Univ. Press, Cambridge, MA, 1978.
11. Kassirer. J. P., and Gorry, G. A. Clinical problem solving: A behavioral analysis. *Ann. Intern. Med.* **89,** 245-255 (1978).
12. Haug, P. J., Clayton, P. D., Schmidt, C. D., Pearl, J. E., Farney, R. J., Tocino, I., Morrison. W. I., Fredrick, P. R., and Warner, H. R. "A Pulmonary Diagnostic System for Early Hospitalization." Soc. for Medical Decision Making, Chicago, 1986.
13. Ledley, R. S., and Lusted. L. B. Reasoning foundations of medical diagnosis. *Science* **130,** 9-21 (1959).

14. GORRY, G. A., AND BARNETT, G. O. Experience with a model of sequential diagnosis. *Comput. Biomed. Res.* **1,** 490-507 (1968).
15. DEDOMBAL, F. T., LEAPER, D. J., HORROCKS, J. C., STANILAND, J. R., AND MCCANN, A. P. Human and computer-aided diagnosis of abdominal pain: Further report with emphasis on performance of clinicians. *Brit. Med. J.* **1,** 376-380 (1974).
16. ZAGORIA, R. J., REGGIA, J. A., PRICE, T. R., AND BANKO, M. Bayesian classification in medicine: The transferability question." *In* "Proceedings, Fifth Annu. Symp. on Comput. Appl. Med.," Inst. Elec. Electron. Eng., New York, 1981.
17. ENGLE, R.L., FLEHINGER, B. I., ALLEN, S., FRIEDMAN. R., LIPKIN, M., DAVIS, B. J., AND LEVERIDGE, L. L. Heme: A computer aid to diagnosis of hematological disease. *Bull. N. Y. Acad. Med.* **52,** 584-600 (1976).
18. TEMPLETON, A. W., JANSEN, C., LEHR, J. L., AND HUFFT, R. Solitary pulmonary lesions. *Radiology* **89,** 605-613 (1967).
19. WARNER, H. R., TORONTO, A. F., AND VEASEY, L. G. Experience with Bayes' theorem far computer diagnosis of congenital heart disease. *Ann. N.Y. Acad. Sci.* **115,** 558-567 (1964).
20. CLAYTON, P. D., OSTLER. D. B., GENNARO, J. L., BEATTY, S. S., FREDERICK. P. R. A radiology reponing system based on most likely diagnosis. *Comput. Biomed. Res.* **13,** 258-270 (1980).
21. TVERSKY, A., AND KAHNEMAN, D. Judgement under uncertainty: Heuristics and biases. *Science* **185,** 1124-1131 (1974).