DECISION-DRIVEN ACQUISITION OF QUALITATIVE DATA

Peter J. Haug, M.D., and Homer R. Warner, M.D.,Ph.D.

LDS Hospital and University of Utah
Department of Medical Biophysics and Computing
Salt Lake City, Utah

During the past 15 years researchers at the LDS Hospital in Salt Lake City have worked to develop a hospital-based medical information system. A primary goal of their efforts has been to establish a patient data base sufficiently complete to support sophisticated computer assisted medical decision making. The system created is called HELP, an acronym for Health Evaluation through Logical Processing (1,2,3).

The HELP system includes in its computer-based medical record quantitative patient data such as white blood count and serum creatinine. At present it acquires data from the clinical laboratory, a multiphasic screening area, the pulmonary laboratory, pharmacy,ECG, the intensive care wards, the cardiac catheterization laboratory, Radiology, and a variety of other sources. A major advantage in gathering this quantitative data is that it can be collected directly from the instruments that generate it. Exceptions such as vital signs can be easily transcribed by trained data-entry technicians.

The consistent collection of the type of qualitative data epitomized by the history and physical exam has been a more elusive goal. The optimal source for this data are the patient and the physician respectively. Branching questionnaires appear to work if limited to a narrow area of medical subspecialty. However the collection of a history and physical exam sufficently detailed for the decision-making needs of general internal medicine patients remains the province of the practicing physician and the paper chart. The set of findings which might be appropriately considered in an internal medicine patient is too extensive to adequately collect with a branching questionnaire. It consists of the possible descriptors for all signs and symptoms related to diseases in this broad specialty.

This paper reports progress in implementing a new data collection tool within the HELP system designed to overcome the limitations in qualitative data gathering that have kept these rich sources of information out of the computer-based medical record. The approach described in this paper takes advantage of the medical decision making capability present in the HELP system. It is called Decision-Driven Data Acquisition.

HELP Decision Function

Medical decisions in the HELP system are represented as modules containing a collection of statements that 1) are created and displayed using a editor/compiler that interacts with the decision author in a natural medical language, 2) allow specification in detail of the data required to make the decision, and 3) support logical, statistical, or mathematical constructs which can be applied to that data in order to confirm or deny the decision under consideration. These modules are called HELP sectors.

HELP sectors are organized into blocks, groupings of sectors dealing with specific subsets of medical decisions. Experts in each of these specialty areas can take responsibility for the creation and updating of the decision logic for these sectors, usually with the assistance of a knowledge engineer.

Within the HELP system the decision making process is triggered by the storage of prespecified data in the patient record. As a sector is created, the decision author (expert) specifies what data and what values of that data should evoke the sector. The author of a decision to give oral penicillin might, for instance, choose to drive consideration of this decision whenever the results of a throat culture positive for Streptococci is stored in the patient record. To do this he places a flag in the sector on the data item specifying group A streptococci in a throat culture. Thereafter, whenever the microbiology laboratory indicates this finding for a patient, the system will process the designated sector. If the logic in that sector is satisfied, a suggestion to treat with penicillin will be generated. This mode of sector selection is known as data-driven **decision making.**

Decision-Driven Data Aqisition

The Decision-Driven Data Acquisition (DDA) process is best described in terms of the two procedures developed to accomodate this human-machine interface. The first is the ASK function, a call embedded in a HELP sector specifying data needed to complete the decision making process that sector represents. The arguments of the ASK function are 1) a reference to the item of data required and 2) an address field representing the expected data source (i.e. nurse, attending physician, patient, etc.). The address field can specify more than one possible data source so that data from the examination of the lungs might be requested from the nurse or physician whichever signed on to the computer first. The ASK process operates only if the referenced data is absent from the patient's file.

The ASK function is embedded in the standard logical constructs available in the HELP decision logic. Since its use depends on sector logic the asking process is effectively under the control of coded medical knowledge. A simple example of three statements that use medical knowledge to ask an appropriate question is shown below.

```
A. SEARCH: Have you had black or tarry stools?,
B. SEARCH: Hematocrit, LAST, FROM 2 days ago,
           TO now
C. ARITH:  If B LE 35 then ASK((Patient
           questions),A)
```

These statements might appear in a sector designed to decide the likelihood of gastrointestinal bleeding in an anemic patient. Item B could be flagged to data-drive this decision. The hematocrit would automatically be forwarded to the patient data base from the clinical laboratory. In the case where the value of the hematocrit was less than or equal to 35 and the answer to the question in A was not found in the patient's data base, the question (A) would be asked the patient.

The ASK construct localizes the control of data collection within the sector and provides a truly hypothesis (decision)- driven questioning tool. The expert who generates the decision module can model the data collection process as well as the decision algorithm as he builds a HELP sector.

When activated the ASK function places the necessary codes in a file. At that point the process of collecting the requested data becomes the province of the second part of DDA, the QUERY DRIVER.

The QUERY DRIVER is the program responsible for formatting the data requests stored by the ASK function and for presenting them to be answered. If the addressee is logged on at a terminal as sectors generating questions for him are run, the data requests will be presented directly. The result is a question /answer session between the designated source for the data and the decision system. If the addressee is absent at the time of question generation the requests are held until he/she signs on. This essentially establishes an electronic mail service between the computer and the human data source. As the data collected is stored in the patient data base the QUERY DRIVER reruns the sector that requested the data. In this way the decision process can be completed.

In combination with the data-driven mode of decision initiation described above, this method of data collection provides an effective way to reduce the number of questions asked in collecting a history or physical exam. A small set of screening questions are asked the patient and the answers are stored in the patient record. Hypotheses (possible diagnoses) are generated through the data-driver and the ASK functions in these sectors send requests for additional data back to the patient through the QUERY DRIVER. The patient's answers to these questions from the querying program may generate additional hypotheses in which case the cycle is repeated. The questioning process is complete when adequate data is present in the patient record to resolve all hypotheses under consideration.

This method of controlling the history gathering process reduces question numbers in two ways. First, questions are asked only to explore hypotheses, the potential viability of which have been established by the storage of suggestive data in the patient record. Second, sectors may be written so that internal sector logic halts the data gathering process when sufficient information is present to either confirm or deny the decision in question.

Since the history tends to focus the set of diagnostic hypotheses, the list of physical exam findings requested from the physician will, in most cases, be relatively short. To enhance understanding of the data requests, the questions asked the physician or nurse are accompanied by a statement of the hypothesis under consideration.

Hierarchical question process

Another level of control of the questioning process has been introduced into the QUERY DRIVER by taking advantage of the hierarchical nature of the data structure itself. In interviewing patients it is common to proceed from general questions to specific questions. For instance, the question "Do you have chest pain" preceeds the questions "Is your chest pain brought on by exertion?" and "Do you have chest pain that is worse with a deep breath?" as well as a variety of others. A "no" answer to the first question implies a negative response to the others and they need not be asked, but a "yes" answer to the first question generates a lists of hypotheses for which answers to the more specific, second level questions may be necessary.

In developing our symptom data base, we have embedded this hierarchical relationship in the data dictionary structure. Thus a group of questions are designated as contingent on a parent question and cannot be asked until the QUERY DRIVER asks and stores the parent.

As an example consider a sector designed to diagnose pulmonary embolism. To determine the presence or absence of pleuritic chest pain the following sequence is created:

A. SEARCH: Do you have chest pain which is sharp or stabbing?
B. SEARCH: Do you have chest pain made worse by a deep breath?
C. EXIST: ASK((patient questions)A,B)

The data codes associated with these questions indicate that they are both contingent on the answer to "Do you have chest pain?". The parent question is asked first. A "yes" answer leads the QUERY DRIVER to ask the two specific questions. If, however, the patient denies chest pain, the query driver recognizes from the code structure that neither specific pain manifestation will be present. Without further questioning, a "no" is stored for each of the three questions.

Subsequent hypotheses may also request data similarly related to "Do you have chest pain?". The sector for angina pectoris may, for example, need the answer to "Do you have chest pain brought on by exertion?". The QUERY DRIVER can test the patient record for an answer to the parent question and if that symptom has been denied can automatically provide the "no" answer required to process the new hypothesis, without presenting any question to the patient.

This data structure reduces the questioning process and also eliminates the confusion caused when the system requests detailed information about a general symptom that the patient does not have. The sector author's burden is eased by the fact that he/she does not have to foresee this questioning sequence while designing the sector.

The ability to specify a hierarchical structure for families of questions abets the effort to collect the maximum useful information during the minimum interaction with a patient or physician. Combined with the data-driven mode and DDA it provides another level of system "intelligence for the task of gathering qualitative medical data.

Ranking

One additional method can be used to control the questioning process. In some cases the QUERY DRIVER processes questions from sectors using a scoring algorithm, such as Bayesian probabilities, to rank decisions. The scores associated with these decisions may be used to order the questions that the QUERY DRIVER asks.

For example, in a group of sectors designed to seek the patient history necessary to assess the likelihood of each of a collection of pulmonary diagnoses, a sequential Bayesian rule is applied to the gathered symptoms. A probability is estimated for each disease as the sector is executed. In selecting from the list of questions generated the QUERY DRIVER chooses those needed to complete the evaluation of the currently most likely sector. These questions are asked and all sectors using the answers are rerun. The new ranking determines the information to be sought during the next cycle. This technique hastens the collection of the data most relevent to the patient's primary problems.

Current status

The processes described above have been implemented in a version of the HELP decision system which runs on a microcomputer in a dedicated data collecting mode. This machine communicates with the central computer facility allowing transfer of data to and from the central patient data base as needed. We chose this approach to avoid the competition for compute power inherent in a heavily used, time-shared medical information system. Interactive questioning is most effective when response times are minimal. This system runs fast enough to provide the short response times required in interactions with patients and physicians.

A collection of 70 diagnostic sectors designed for use on this system have been created and have undergone initial testing (3). They are primarily concerned with pulmonary illness and use a Bayesian scoring algorithm. Findings from both the history and physical exam are heavily represented in their logic. These sectors will be used in direct, computerized patient interviews as a part of comparative testing of various methods of collecting inpatient data. They currently run through a mobile terminal wheeled to the patient bedside. Communication is over phone lines to the microcomputer.

A group of 65 sectors representing diagnostic and therapeutic decisions in obstetrics and using primarily boolean logic has also been created. These will collect data from both pregnant patients and their attending physicians. The accuracy of the logic is currently being tested with questionnaire-input data. The sectors will be run in the DDA mode in an outpatient obstetrics clinic.

A group of 130 sectors was developed twelve years ago to gather historical data for a multiphasic screening setting (4). They ran on an early version of DDA and collected histories from approximately 35,000 patients. A rewritten version of these sectors will be used to test DDA in the screening setting on patients being admitted for elective surgery.

In summary, DDA is a system for the collection of qualitative medical data based in an active medical decision making system. It depends on ongoing computerized decision activity to generate questions appropriate to the individual needs of the patient. Data is collected from the patient, the physician, and other health care professionals. Testing of this addition to HELP is in progress.

References

1) Computer-assisted Medical Decision-Making; Warner HR; Academic Press, 1979

2) The HELP System; Pryor TA, Gardner RM, Clayton PD, Warner HR; Journal of Medical Systems, Vol. 7, No. 2, 1983

3) Medical Data Aquisition using an Intelligent machine; Warner HR and Haug P; Proceeds of Medinfo 83 Seminar, Amsterdam, Holland, July 1983

4) A Computer System for Diagnosing Pulmonary Disease; Gerard MJ, Haug PJ, Morrison WJ, Tocino I, Fredrick PR, Crapo ROm, Harada SK, Clayton PD; (Submitted Annual Mtg. Am. Assoc. Med. Inform., San Francisco, 1984)

5) A Sequential Bayesian Approach to History Taking and Diagnosis; Warner HR, Rutherford BD, Houtchens B; Computers and Biomedical Research, Vol. 5:256-262, 1972