Testimony for NCVHS Hearings

on

Medical Technology and Code Development

Washington, D.C.
May 17-18, 1999

ELMER R. GABRIELI, M.D., F.C.A.P.

Founder & President, Computer-Based Medicine
215 First Street
Cambridge, MA 02142

NCVHS Hearings on Medical Terminology and Code Development

Name: **Elmer R. Gabrieli, M.D.**
Fellow, College of American Pathologists
Organization: **Computer Based Medicine, Inc.**
Founder and President: E. R. Gabrieli, M.D.

This research organization was started 19 years ago with the sole objective to develop a computer-based information system able to analyze narrative medical text. Briefly, the task was to program the computer to:

- **read the narrative;**
- **find the information (clinical facts) in the narrative;**
- **extract the clinical facts, preserving all modifiers;**
- **code the clinical facts;**
- **deposit the clinical facts into an anonymous database for statistical analysis; and**
- **be able to retrieve clinical experience instantly, with statistical inferences, if it is so desired.**

The project of automatic electronic medical record generation has now been completed and is operational.

It is difficult to report two decades of research, a combination of linguistics, Artificial Intelligence and clinical knowledge, in a condensed form. Therefore, a recent writing is submitted for distribution, to share our experiences.

1. Definitions and requirements for Patient Medical Record Information (PMRI):

How would you define or describe PMRI?

Answer: In the ambulatory care and emergency care settings, the PMRI should contain all information that was traditionally recorded in the paper record. The hospital inpatient record is more complex.

The stored electronic patient records should be securely protected. Access to the data in the electronic record must be fully controlled by the patient. An access trail should display to the patient a list of persons who had access to confidential information and their reasons for requesting access.

A. Why is comparability of PMRI required? What functions does it serve?

Answer: In the future, when PMRI will gradually replace paper records nationwide, the data access between care sites, within an enterprise or at remote locations must be a simple task and thereby reducing redundant laboratory tests, eliminating repetition of time-consuming history taking, questioning past treatments and drug sensitivity inquiry. PMRI will allow longitudinal life-long records, particularly important in chronic diseases (diabetes, etc.), in gene disorders often with long incubation time, and for monitoring familial-hereditary diseases.

B. How comparable does the PMRI need to be for these purposes, i.e., how precise, how accurate?

Answer: Comparability of the PMRI depends largely on the quality of intelligence provided by the lexicon(s). Accuracy: transformation of the English narrative into digital data must be high fidelity, i.e., the meaning of each clinical fact must reflect the author's exact intention. The text processor output must be semantically identical to the input from the patient record. This way, the information in the PMRI would be exactly the same as the dictated or written description and comparable with other PMRIs.

What are the consequences if the PMRI is not accurate?

This would negatively affect acceptance of the PMRI, medically and legally.

2. The role your medical terminology plays in representing PMRI:

A. What is the intended purpose of your medical terminology?

Answer: To provide the linguistic and clinical intelligence to the function of our automated medical text analyzer. The attributes of the medical lexicon are particularly critical. Here is our brief list of the major, essential attributes of the lexicons which are to cover both the medical and non-medical words and phrases of the narrative medical text:

1. Comprehensive: the lexicon must be able to handle all medical terms/phrases as well as all non-medical words/phrases that may be present in a medical record. An unrecognized word/term/phrase not present in the lexicons could cause problems.
- Synonyms must be linked to assure retrieval, and near-synonyms must be subcoded to the canonical head term of the synonym (e.g., cancer/carcinoma)
- word families (full morphology, plurals, past tense) must be listed as subterms of the entry term;
- vernaculars (words/terms used by the patient and quoted verbatim, such as "head cold," must be listed and linked to the corresponding canonical term;

2. The lexicons should be computer-based to allow efficient, timely updating, and to maintain forward-compatibility;

3. Medical language oriented, listing collocations (fixed word combinations, e.g., "ambulance chaser"), and characteristic clinical phrases "left the operating room in stable condition;"

4. Meaningful code scheme: codes should keep term kinships together. To each entry term we apply two codes: a permanent random code which shares no semantic attributes with the term, and a meaningful hierarchical code which may change as medical opinion changes. The random code assures forward compatibility.

5. Source sensitive: a statement may originate from the patient, or may come from a diagnostic test report or from the examining physician (e.g.,

"dyspnea" may be a complaint, or a sign observed by the physician). The source of clinical data should be clearly differentiated by the assigned code, to qualify the clinical fact.

6. Taxonomically sound: classified by a single key. Potential ambiguities must be resolved (e.g., "ovary": a member of the endocrine system and also the reproductive system). The two separate locations apart in the semantic space must be linked with a pointer.

7. Readily updateable: in our experience, there are 4,000 to 6,000 new words/terms/phrases per annum. A specially trained team should collect these new terms.

8. Nationwide compatibility: essential for protecting nationwide interoperability. Perhaps there should be an umbrella lexicon which can include other lexicons such as the SNOMED/Read combination, ICD, CPT, etc. This umbrella nomenclature, for medical and non-medical words/phrases, should be formally and centrally updated, with periodic updating at all user sites. Users may change the classification scheme for their own use, but dissemination of user changes should not be allowed, as that could undermine compatibility by creating dialects.

9. Customer service: users encountering a legitimate word/term/phrase not listed in the umbrella nomenclature should notify the nomenclature maintenance center to add the term to the vocabulary.

10. Dealing with lexical ambiguities: Many medical terms are inherently ambiguous. "Cervical" may refer to the neck or to the uterus. The lexicon must notify the text processor about such ambiguities.

11. Polysemy is a frequent text processing problem in automated text processing. Polysemy refers to lexical entry terms with more than one meaning. For example, "chip" may mean a piece of wood or electronic circuit. Close to a quarter of all non-medical terms in our lexicons are polysemous. The lexicon must list all legitimate uses of every polysemous entry.

12. Medical and non-medical lexicons: In a typical patient record, 25 to 40 percent of the test is factual information; the rest is conversational (non-medical) text, which may also carry important information. Therefore, the text processor must analyze the entire. In our system, the medical and non-medical

lexicons are nearly equal in size, approaching a total of over a million lexical entries.

What is it currently used for?

Answer: We use the medical text analyzer system to process the narrative parts of the patient record.

B. What is the clinical domain, scope, or healthcare setting addressed by your medical terminology?

Answer: It covers all aspects of clinical medicine, at different care sites. Currently we process ambulatory care records, but we have also processed x-ray reports, emergency care records, operation notes, etc. The system is able to process virtually any kind of medical narrative. The system accepts the output from the transcriptionist's word processor.

C. What evidence do you have of market acceptance of your terminology? {Ask users what their perceptions of gaps are}

Answer: Since clinicians do not even notice the post-dictation activity of the PMRI, they do not care how the record is processed, but they like the results: the electronic list of clinical facts in English, and also in digital form.

D. In what areas are you now planning to expand your terminologies?

Answer: I plan to combine the current automated medical text processor with a speech recognizer front end and use an algorithm to disambiguate chronic phonetic difficulties of similarly sounding words by context algorithm..

3. How does your medical terminology relate to other medical terminologies?

Answer: We have cross-referenced ICD, CPT and COSTART

4. How does your medical terminology relate to healthcare message format standards?

Answer: At this moment, we protect our PMRI by not allowing messages to risk patient privacy. As soon as data security arrangements are reliably in place, we will install some data exchange system.

B. Is the use of your medical terminology within these message format standards required, preferred, or optional?

Answer: See above answer. I believe it will be optional.

5. Are there issues related to medical terminologies that deserve government attention or action?

Answer: Yes.

A. What are they?

1. A formal governmental structure should be created for continuous updating of the lexicons maintained to support medical text processing. This team should include physicians, medical text processing experts, and linguists.

2. Perhaps the same team could deal with user requests for adding new terms/phrases or changing term relationships.

3. Support for accelerated development of speech recognizers. The current manual transcription process costs about $91 billion per year and causes delays.

B. What can be done to address these issues in a one- to four-year time frame?, a five- to ten-year time frame?

In a five-year time frame:
1. HCFA should take national leadership in introducing PMRs.

2. Grants should be given specifically for the development of steps to enhance benefits such as:
- succinct electronic guidelines for strategies of clinical diagnostic workups,
- succinct electronic guidelines for management of different clinical conditions, in the form of real-time assistance of care providers,
- automation of electronic drug ordering to reduce errors,
- guidelines for a well-secured national data sharing procedure allowing authorized data access between care sites,

- algorithmic creation of longitudinal care files by linking patient records generated at different sites and different times. This will be most important for handling documentation of gene disorders with long incubation times,
- development of progressive databases for disidentified clinical case histories which could be continuously analyzed using advanced statistical methods to derive the currently "best" clinical care for the lowest cost.

6. Are there issues related to the comparability of PMRI?

Answer: Data centers producing PMRI could use the encoded clinical facts to provide mandated reimbursement codes, e.g., ICD, CPT, etc.\

A. If so, what should be the role of the government?

Answer: To select the official U.S. code scheme and make it available to all users. This should be the best available medical lexicon to support PMRI production.

B. Is there a need for increased coordination among terminology developers? If so what type.

Answer: Yes, but most current terminologies are proprietary, expected to be marketed for profit. This may be in conflict with national interests. Tactful, sensitive leadership is required.

D. In the short run (1-4 years)/long run (5-10 years)?

Answer: The privacy, confidentiality, and security of personal health information are very important to us but, because NCVHS has held several hearings in the past two years on these topics, we have not made them the focus for this testimony**.**

BENEFITS OF ELECTRONIC MEDICAL RECORDS

- Electronic medical records can be fed into a clinical database after all patient and care provider identifications are stripped, and with the use of statistical techniques, these records may be used to derive **the best clinical care for the lowest cost.**

They also:

- allow algorithmic monitoring of the **QUALITY** of care rendered;
- greatly facilitate **OUTCOME** studies;
- allow **DATA ACCESS** between care sites within an enterprise or at remote locations, thereby reducing redundant laboratory tests;
- could automatically optimize **DRUG PRESCRIPTION**, preventing errors in drug ordering;
- could detect **DRUG SENSITIVITY** in the case history and remind the care provider before an error is made;
- could compare the dictated clinical record with prestored **DECISION SUPPORT** guidelines, and thus assist the care providers in decision making in real time;
- can greatly enhance the efficiency of **CLINICAL TRIALS** and assist in postmarketing surveillance;
- allow automated linking of patient records generated at different sites and different times, creating **LONGITUDINAL RECORDS** , crucially important in handling gene disorders with long incubation time, and for monitoring familial-hereditary illnesses;
- In many chronic illnesses such as cancer, diabetes or certain cardiac diseases, with **LONGITUDINAL CARE RECORDS**, various quantitative parameters could algorithmically compared with the current findings, to quantify the progression of the disease, and to quantify the therapeutic value in terms of affecting the rate of progression of the illness;
- Electronic medical records will be **DEMANDED** by insurers, regulatory agencies and malpractice insurers, as well as the millions of patients who will know that the electronic health records offer **SAFER, BETTER CLINICAL CARE FOR THE LOWEST COST.**

II. Other Benefits of the Electronic Medical Record

Although the primary task of patient records should be to serve our clinical system, their financial and administrative benefits are also significant.

The following **administrative benefits** are immediately apparent:

1. **Security and confidentiality of medical records:** The recently passed HIPAA law mandates that medical information must be protected. It is often stated that the traditional paper chart is neither secure nor confidential. It is a fallacy that storage at the care site is automatically secure. All too often, patient charts are found unattended in the hospital ward. Anyone who has the right appearance and attitude to look like he/she belongs can access almost any chart. Over the years, the list of authorized access to paper charts has also been alarmingly increased. It started with lawyers involved in malpractice suits demanding all care documentation. This was followed by the investigating police, Health Care Financing Administration and other insurers. Copying of charts has become a major task of the Record Room staff. And once a chart is copied and released, there is often little control, if any, to know who has had access to the chart.

Well designed electronic medical record systems can be far better secured from unauthorized access. A big difference with electronic records is the possibility of limiting access to specific sensitive areas within the patient record, such as sexual behavior in the preadmission history, the gynecological or surgical note about the operation, the psychiatric notes, etc. This predesigned security can also maintain a data trail, allow an audit trail, can identify the individual who has received access to a part of the record, and by comparing it with the original text, and can detect any illegal modifications of the contents.

In the U.S., by 2001, proven improved security will be mandatory by law, with the risk of governmental fines and other penalties. We can expect that the industry will gradually become convinced that an acceptable security can be achieved only by means of electronic medical records.

2. **Reduction of the clerical staff:** Large amount of personnel reduction will be accomplished simply by eliminating the current cost of inefficient manual handling of the paper records. Currently, the constantly growing expense of paper charts include (a) extensive checking of each chart, for completeness of the record, after discharge from the

hospital, including all diagnostic studies, consultations, etc., then identifying and filing the seemingly complete chart. The next step is retrieval, when needed, such as readmission or when the patient is moved to the ambulatory system. HMOs have an armada of technicians for clerical tasks such as to pull all the records for the patient scheduled for the next visit date, and then refile the same records, after the visit, to the proper position. In a small/medium HMO with a visit rate of one million visits per year, this represents 2 million transactions, an enormous clerical task. Misfiling is easily done, and finding a missing record may require extensive search.

Three or four decades ago, a typical 500-bed hospital's record room staff consisted of five to six record administrators and one or two clerks. This number has swollen today often to more than a hundred. Both in the hospital and at the HMOs, the installation of electronic patient records will eliminate this armada of clerical staff.

Another area of enormously expensive clerical overdevelopment is the process of billing for services. Unfortunate experience with not infrequent overcoding/overbilling has elicited a kind of paranoia in insurance circles, overly suspecting incorrect billing. Checking the validity of a bill issued may require laborious investigation, often in need of reviewing the original chart. With electronic records, the same process should be reducible to focused algorithms, kept reusable, requiring minimal effort. It has been estimated that the current clerical handling of the patient chart is more than $90 billion per year. The Institute of Medicine has correctly stated that electronic patient records represent a fiscal imperative for the health care industry.

3. **Efficiency of clinical practice:** Physicians can access current electronic patient records, the last one, or all previous records, with only a few key strokes, whereas unavailable paper charts or laboratory reports have grown up recently, creating frustration and unnecessary laboratory testing. Computer-based pathology reports should be instantly available. The patient's past medical, social and family history should be easily retrievable from the electronic medical record, without going through a list of previously raised questions, several minutes of the precious encounter time, which is often limited to 30 minutes. Similarly, automation of drug prescriptions would save time and reduce the risks of prescription errors, currently at a frequency of three to four percent.

4. **Control of expensive procedures:** During the last two decades, the insurance industry has discovered that expensive procedures such as hysterectomy, magnetic resonance imaging, computerized tomography or cardiac catheterization, to name but a few, are sometimes overused and could be reducible by external control. This approach to questioning the necessity of a procedure at the insurer's site has often transferred clinical decisions from the physician to a nurse or clerk, at a remote site. Such interference with medical decisions making has elicited deep resentment in medical circles, leading to irritation and resistance. All this would be prevented by the electronic records with electronic guidelines reminding the doctor the exact indication for each expensive procedure.

5. **Cost saving:** due to algorithmic control of expensive items such as special imaging studies, consultations or hospitalizations can be determined accurately only after the installation of the electronic record system, but a crude estimation is a savings of 18-22% of the total operating cost of a hospital or HMO. Pathways and guidelines which are accepted by the medical staff, and assisted by proper algorithms should be the way to satisfy physicians, insurers and patient alike. Guiding rules should take over replacing individual intuitive judgements, eliminating disagreements.

6. **Malpractice** insurance, as recent actual cases indicate, has been reduced by 10-15 percent when electronic records of some sort have been installed.

7. **Improved revenues:** The overall cost reduction due to the improvements listed above, with automated reimbursement systems and better enterprise-wide cost accounting with clear definition of costs via transparent clinical activities, should result in a new, constructive self-control guided by informative actuarial accounting system, eliminating controversial payment rejections which lead to investigations, and patient, doctor and insurer dissatisfaction. All this, in toto, enterprise-wide, an estimated 20-25% total cost reduction should be realized.

AUTOMATED ANALYSIS OF FREE NARRATIVE CLINICAL TEXT:

THE ROAD TO ELECTRONIC MEDICAL RECORDS

A Brief Summary

Prepared for presentation to
the NCVHS Hearings on
Medical Technology and Code Development
May 17-18, 1999
before the U.S. Congress

Washington, D.C.

COMPUTER-BASED MEDICINE, INC.

215 First Street
Cambridge, MA 02142
Phone: (617) 494-0909

USEFULNESS OF THE ELECTRONIC MEDICAL RECORD

By:
Elmer R. Gabrieli, M.D.

Prepared for presentation to
the NCVHS Hearings on
Medical Technology and Code Development
May 17-18, 1999
before the U.S. Congress

Washington, D.C.

1. INTRODUCTION

In 1991, the Institute of Medicine was succinct in its milestone report: ***Introduction of electronic patient records is a medical imperative, and an absolute fiscal necessity*.**

Unfortunately, in 1991 the contemporary technology was not ready to implement the excellent visionary recommendation of the Institute of Medicine, and at the time, the available arguments supporting the report's progressive conclusion were more speculative than factual. For example, the report argued that switching to automated data handling had saved 33-35% of the operating cost in various areas of U.S. industry; therefore, the same degree of saving should be true in health care. But the road to reach this postulated goal, viz. to produce electronic medical records, was not recommended, and even the most visible "leaders" of the field had only meager suggestions about the expected benefits, such as automated reminders to the physician that an influenza vaccine or a mammogram is due.

2. WHAT IS AN ELECTRONIC PATIENT RECORD?

Quite correctly, in the 1980s, the electronic patient record of the future was expected to be equivalent to the customary paper record, to contain everything that was dictated by the care provider. That is, it should contain the same information as the paper chart, but in a retrievable mode. This original goal was somewhat forgotten, and major compromises have taken place on the market. During the last decade, four discrete categories of electronic patient record systems have been marketed, each with a set of different goals, and with very different information content.

(2a) The first of these categories is a simple move: the entire patient record, from cover to cover, is deposited in a computer as images of the paper record (c.f. Grady Memorial Hospital in Atlanta). This way, the entire electronic patient record is permanently stored. Perhaps this is some improvement over the traditional microfilming, but the core problem is not resolved. The retrieved record still must be read and interpreted by a physician; thus the manual task remains the same. Comparison of many records for useful statistical inferences is not feasible. This approach offers only record retrieval, not information retrieval.

(2b) A second type of marketed electronic patient record system serves as a repository of certain clinical data which are making the captured data retrievable, readily encodable, such as orders, laboratory results, administrative data, demographics and other easily standardizable data. These systems meet a modest but practical objective, making the captured data retrievable readily, but they avoid the primary challenge, which is to analyze the free narrative text, the true patient record where 70-80% of the important clinical information is stored.

(2c) The third category, fixed vocabulary and structured patient record systems, may result in uniform patient records, but this was met with little enthusiasm on the part of the physicians. Patients with similar conditions often differ in details. Physicians are trained to describe their own detailed observations, along with the collected patient information, to "customize" the record with great accuracy. A fixed vocabulary with a structured record serves more the technology than the practitioner's freedom to choose his own wording.

(2d) The fourth category is the electronic medical record, an automated, computer-generated complete electronic record, currently offered only by Computer-Based Medicine, Inc. Here, an "intelligent" computer "knows" the medical language. It can "read" the narrative text sentence by sentence, and locate all the grammatical structures that carry clinical facts. It extracts these clinical fact-carrying structures and then encodes each clinical fact with specific codes, thereby converting each piece of clinical information into numeric data. These digitized facts are then suited for further analysis by the computer.

This is the first such "intelligent" system generating electronic records which then consist of individual clinical facts. The system was developed by our team after nearly two decades of intensive research and testing.

The benefit of such a complete electronic medical record is the listing of all the coded clinical facts present in the physician-dictated record. This comprehensive extraction of all the clinical information content allows flexible retrieval of one office visit record, all records of a given patient accumulated over years ("longitudinal record"), or specific data present in many records, in any combination, for deriving an instant statistical inference ("clinical experience retrieval").

The "currency" of clinical information, the valuable components of a clinical record, are the clinical facts. A typical ambulatory care record contains 60-120 such discrete clinical facts. The complete electronic record provides a list of all these clinical facts that were scattered and buried in the narrative source text. This results in a "currency of the smallest denomination": each clinical fact is molecular in characteristics. Each fact is like a molecule in chemistry, the smallest granule of information which cannot be made smaller without altering the intended meaning (see further pages 22-39).

3. USEFULNESS OF THE COMPLETE ELECTRONIC RECORD

We argue that only a large nation-wide complete electronic patient record system would make clinical experience derivable from millions of patient records. This capability should change the current, increasingly controversial, intuitive bedside decisions of medicine over to scientifically defensible, statistically optimized clinical care. Such a drastic change from memory-based clinical decisions to fact-driven experience-based decisions would fully exploit the potential of modern medicine at a parsimonious cost, because nation-wide experience retrieval, per se, would be as simple as a telephone call, easy, rapid and inexpensive.

This miraculous new opportunity to upgrade clinical medicine is intellectually compelling, but it needs further specifications and perhaps even proof of the concept for the uninformed.

In our first major demonstration project of our complete electronic record system, we tabulated data such as the physicians' prescription habits, variations in drug ordering among physicians, and disease statistics, to illustrate the easy retrievability of various individual clinical facts. Physicians participating in the project liked these reports, but then the obvious

question was raised: How much investment should an HMO make to justify changing from paper records to electronic records? Our original answer was that the physicians and the administration should dictate the next step. This was naïve thinking on our part. Neither the physician of the HMO nor the administration were prepared to develop a proposal.

At that point, we had to recognize that we must go back to the drawing board to offer some clinical applications that could be viewed as irresistible benefits to the HMO's decision makers. This paper is a brief summary of our current view of some obvious initial clinical applications.

4. DATA-DRIVEN COST ANALYSES

A complete electronic medical record system can provide data for many different types of continuous cost analyses. The electronic record makes clinical care transparent, and the record, in whole or in part, is readily retrievable. Therefore, authorized members of the hospital or HMO management could easily carry out numerous direct cost studies, via a keyboard and with simple actuarial computations to identify various known or suspected points, where wasteful practices could be identified. These cost studies could focus on certain specific data. For example, it has been found that changing the entire current drug therapy of a patient with hypertension to a different drug combination is expensive, and it is often based on hasty and impulsive clinical decisions. If that physician would know the cost involved, he could change one or two drugs, not the entire list. The electronic patient record system could also help the HMO administration and hospital staff to propose fair policies for specialist referral and could provide statistical summaries of the indication type and frequency of such referrals. In addition, a list of expensive tests or studies could be catalogued, and statistics for indications could be found and reported by using the electronic records. In this way, enterprise-wide policies could emerge specifying, for example, the indications for

- CAT scan,
- cardiac perfusion,
- cardiac catheterization,
- referral to various specialists,

- organ transplantation,
- home care,
- MRI, etc.

Such cost studies are possible today by manual methods, but it would require a large group of competent staff for collecting and analyzing the original patient records. This could only be done at a high operating cost compared to electronic record. By using a few strokes of the keyboard, recalling the relevant clinical facts could be recalled from electronic records at negligible cost.

Since the electronic patient record contains all the recorded clinical facts in a readily retrievable mode, the use of cost analysis may suggest the review of any combination of data for cost comparison. This could be used to provide detailed analyses of various operational expenses. For instance, the cost of managing certain conditions, such as diabetes mellitus or ischemic heart disease, could be compared across the enterprise. This would include the cost of office visits, drug requirements, emergency room visits, hospitalizations and the like, the case management.

An HMO could store a brief summary record of all its patients in a special database. In case of an emergency or when a patient is traveling and is away from home, the care provider could access the summary record by a simple phone call, reducing repetition of diagnostic tests and establishing diagnoses and drug sensitivities. Confidentiality can be protected by well-tested caller identifier protocol and/or patient data encryption.

5. QUALITY OF CARE

5-1. Defining and Measuring Quality of Care

During the "golden age" of medicine that lasted until about the 1950s, physicians enjoyed full public trust, and it was an unchallenged belief that all doctors knew how to prescribe the best treatments. This great freedom of medical decision making was questioned soon after the dramatic changes in the financing of healthcare services in the U.S. On July 7, 1966, the fee-for-service system was changed and the working man's health insurance was greatly extended by insurers and by Medicare legislation. Initially, the quality of care was

still largely left to doctors. The original congressional Medicare legislation merely placed the responsibility on hospitals to guard against inappropriate care or unnecessary surgeries. Tonsillectomy was the first target. Evidence showed that the many cases of unbridled tonsillectomy—one quarter of hospital admissions—were quite unnecessary.

The reports of John E. Wennberg changed the national attitude toward quality of care by collecting data instead of theorizing. Beginning in 1969 Wennberg simply compared the frequency of certain clinical interventions in different geographic regions. This approach revealed remarkable variations, clearly showing that surgeries like appendectomies, hernia repairs, prostatectomies and hysterectomies were performed excessively. There was a 15 percent chance for a man by age 70 to undergo prostatectomy in one town, and a 60 percent chance in another. Based on his work, a brave new idea emerged: *if medicine is a science, there should be much less variability in therapy decisions.*[1]

Donabedian pioneered in the field of searching for the definition of quality of care.[2] The need for measuring quality of care slowly gained public acceptance.

The latest milestone in this search for defining quality of care was a series of six articles initiated by David Blumenthal at the Massachusetts General Hospital.[3] In this series, Berwick[4] argued that the hasty development of treatment guidelines, critical paths, and customer satisfaction surveys all coincided with the commercialization of the medical marketplace, and pressure for cost saving that has exceeded the importance of quality of care.

Operating cost and quality of care are often viewed as conflicting forces molding our HMOs. Controlling cost is critical for staying in the black, but quality is at least equally important in order to remain competitive and acceptable by patients.

The cardinal issues in this problem are a) to define firmly and explicitly what quality of care is, and to inspire the practitioner to adopt that definition of quality of care, and b) to measure the quality of care objectively, quantitatively, accurately, and inexpensively.

---

[1] Wennberg, J.E.: Dealing with medical practice variations; Health Aff, 3: 6-32, 1984.
[2] Donabedian, A.: The quality of care; JAMA, 260: 1743-8, 1988.
[3] Blumenthal, D.: Quality of healthcare; New Engl J Med, 335: 891-3, 967-70, 1060-3, 1146-9, 1227-31 and 1328-31, 1996.

Brook and McGlynn of Rand Corporation and Cleary of Harvard categorized five methods by which quality has been assessed.[5] A physician may review a patient record and answer the following questions:

- Was the process of care adequate?
- Could better care have improved the outcome?
- Considering both of the above, was the overall quality of care acceptable?

These three methods are obviously more subjective than scientific. The examiner's background and personal opinion are affecting the methods. No effective standards for comparison exist.

The fourth method of assessing quality is much more illness-specific. For example, in the case of a patient with diabetes mellitus, the examiner could check the visit record for information on retina examination, foot examination, and so on: The explicit standard of care for diabetes drafted by the American Diabetes Association could be used as the standard, but it has not become popular.

The fifth method is even more progressive. A preset list of criteria is matched with the expected clinical course, such as near-normal glycohemoglobin levels in diabetes or blood pressure findings consistently below 140/90.

Despite the growing interest in quality of care, progress in this area is slow. This is probably because of the lack of cost-effective technology and the lack of convincing reports. An early attempt by the Health Care Financing Administration (HCFA) involved publishing mortality statistics of Medicare patients. After years of publishing these data, HCFA recognized the flaw: Without linking the relevant clinical data of each patient to the mortality statistics, the raw mortality data were inaccurate and misleading. This program was abandoned.

---

[4] Berwick, D.M.: Payment by capitation and the quality of care; New Engl J Med, 335: 1227-31, 1996.
[5] Brook, R.H., McGlynn, E.A., and Cleary, P.D.: Measuring quality of care; New Engl J Med, 335: 966-70, 1996.

5-2. Using Complete Electronic Medical Records for Assessing Quality of Care

As an example, a most effective test for hypertension management was done by us through testing the effectiveness of care extracted from the electronic record. As a basic premise, we assume that by keeping hypertensive patients consistently in normotensive range, the risk of coronary heart disease, stroke, and other complications of hypertension would be considerably reduced, as reported by numerous investigators.

CATEGORY:	**Systolic Pressure**	**Diastolic Pressure**
Normal range	120 mm Hg	80 mm Hg
High normal range	130-139 mm Hg	85-89 mm Hg
Hypertension:		
Stage 1 (mild)	140-159 mm Hg	90-99 mm Hg
Stage 2 (moderate)	160-179 mm Hg	100-109 mm Hg
Stage 3 (severe)	180-209 mm Hg	110-119 mm Hg
Stage 4 (very severe)	≥ 210 mm Hg	≥ 120 mm Hg

In a recent study, we analyzed the electronic patient records for the blood pressure measurements at revisits of established hypertensive patients. Since our study was based on single office visits, we could not perform a temporal analysis, but some valuable information could be derived by simply ranking the blood pressures of the 76 cases at a random office visit.

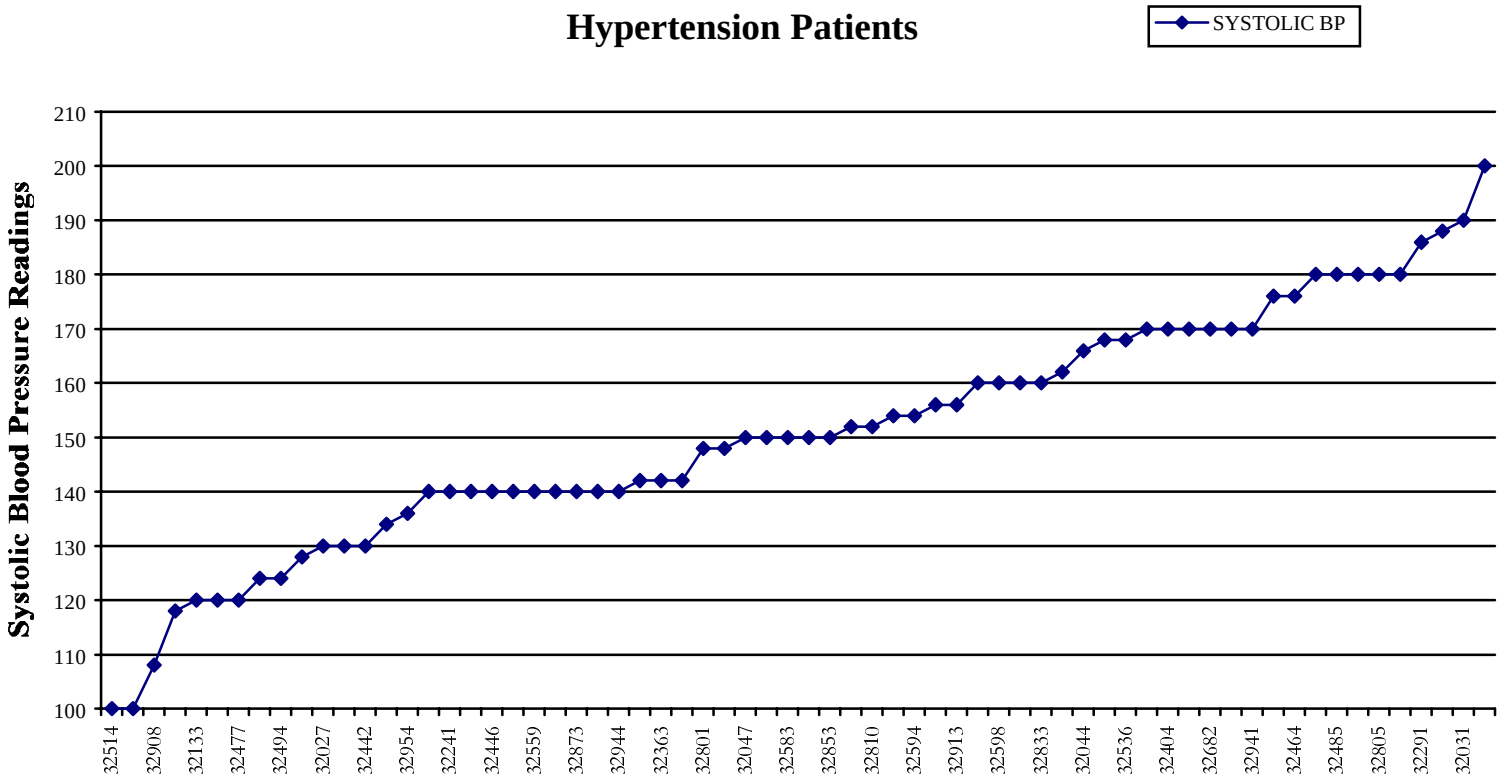


Figure 2. Systolic blood pressure findings of 76 established hypertensive patients measured during office visits.

This analysis shows a substantial number of patients with blood pressure readings above the targeted normal range (140 mm Hg). Interestingly, no comment was in any medical record about the apparent failure of therapy when the blood pressure was over 170 mm Hg. It is my personal impression that if such study results would be presented to physicians, these would improve the quality of clinical care.

Comparisons of these results with results from physicians in other HMOs could be easily done if we had similar records maintained as complete electronic systems. Importantly, this type of quality of care monitoring is very inexpensive once the complete electronic medical record system is installed. And it is very powerful.

6. OUTCOME STUDIES

Outcome studies are another important way to quantify quality of care. These studies can measure quality in some particular clinical conditions, such as survival of breast cancer cases, postoperative course of hip replacement, or frequency of complications requiring hospitalization in patients with diabetes mellitus or coronary heart disease.

When Congress approved the Medicare Act, it also ordered HCFA to monitor the quality of care provided for under the Medicare Act. This led to the creation of peer review organizations (PROs) as we know them today. After numerous critical reports and major national controversy it can be concluded today that any such manual review is extremely expensive, unreliable, and, at best, as good as a manual review of a small sample (less than 10 percent) can be. Importantly, for a fraction of the present PRO operations, HCFA could monitor all Medicare cases and readily recognize the cases with substandard outcomes, by automated analysis of the computerized records.

The great pioneer of outcome studies was Ernest Amory Codman. This iconoclast Boston surgeon bravely stated that success of surgery should not be limited to the postoperative period ending when the patient leaves the hospital alive.[6] Codman introduced the "end result" idea, the ultimate result of the clinical intervention. This idea was embraced by the College of Surgeons. It is receiving much lip service, but manual longitudinal follow-up is operationally difficult, resource-intensive, and very expensive.

Complete electronic medical records are eminently suited for longitudinal follow-up. Algorithmic linking of records on the same patient could be a routine procedure, allowing convenient evaluation of the long-range benefits of some procedures such as cardiac revascularization or hip replacement.

[6] Codman, E.A.: The Shoulder; Thomas Todd, Boston, 1934. (Codman paid for the printing, to avoid pressure from the publisher to leave out the harangues about contemporary medicine.)

7. DRUG THERAPY PROBLEM

In 1976, a front page headline in *The New York Times* read: "Thousands a Year Killed by Faulty Prescriptions."[7] This was one of the early alarms, but these early concerns lacked substantive data. These were followed by more factual reports[8] pleading for "better training of physicians" in the discipline of drug ordering.

Timothy Lesar[9] and his colleagues in Albany, New York, found "significant risk to patients from medication orders, and more than half the mistakes could have harmed the patient if not caught." Interestingly, this study also indicated that prescription mistakes were not dominated by less-experienced residents. The errors were equally widespread among all prescribers. According to this study, there were roughly three errors per thousand medication orders.

Seven years later, Lesar published another study showing, if anything, a worsening situation. In this second study, there were four errors per thousand medication orders, of which 70 percent could have seriously harmed the patient if not caught.

A 1989 headline in *Hospitals*, published by the American Hospital Association, read, "Drug Errors: Dangerous, Costly and Avoidable."

Another major study by T.A. Brennan, an attorney and physician, and his team reported that negligence was responsible for nearly one-third of the problems regarding medications. About one in 25 hospital admissions resulted in a drug-injured patient.[10]

The findings of an Australian research group were even worse: 16 percent of admissions to Australian hospitals include an adverse drug effect, leading to 14,000 preventable deaths each year, as presented to the nation's parliament on June 1, 1995.[11]

In another disturbing study report, 110 nurses of varying experience levels took a written test of their ability to calculate medication doses. Eight out of 10 nurses made

---

[7] Rensberger, Boyce: Thousands a year killed by faulty prescriptions; NY Times, 28January1976, p. 1.
[8] Folliet, Hugo L., et al.: Medication error prevention by clinical pharmacists in two children's hospitals; Pediatrics, 79: 718-22, 1987.
[9] Lesar, T.S.: Factors related to errors in a teaching hospital; JAMA, 263: 2329-34, 1990.
[10] Brennan, T.A.: New Engl J Med, 324: 370-6, 1991.
[11] Cordner, S.M.: Australia's preventable hospital deaths; Lancet, 345: , 1995.

calculation mistakes at least 10 percent of the time, while four out of 10 nurses made mistakes 30 percent of the time. Moreover, the error rate for dosages of drugs administered intravenously was "significantly higher" than for oral medications.[12]

Automation may also cause error. At Boston's Children's Hospital, a five-year-old boy with seizures was supposed to receive BuSpar, an antianxiety drug. The pharmacist followed standard procedure and entered the first three letters of the prescription (B + U + S) into his computer. Because BuSpar was not in the hospital formulary, instead of BuSpar the drug Busulfan came out, a drug used mostly in chronic myelogenous leukemia. When the error was discovered, the hospital claimed that the child would suffer no long-term effects, but others cautioned that Busulfan could cause bone marrow failure.[13]

7-1. Scope of Drug-Related Information

In ambulatory care, as well as in a large part of all clinical care, drug prescription is the major therapeutic instrument. Physicians begin memorizing the drugs in medical school and are expected to keep their knowledge base current. There are 50-70 new drugs available each year, and indications for drugs are also changing constantly. Using the anti-inflammatory and analgesic drug Ibuprofen as an example, the following section illustrates the drug-related facts that are relevant for physicians to consider as they make prescribing decisions.

Brand Names. Today there are around 1,000 generic drugs available for physicians to prescribe, and many of these generic drugs are also marketed with different brand names. For example, the generic drug Ibuprofen is also marketed under five different brand names: Advil, Motrin, Nuprin, Rufen, and Trendar.

Contraindications. In the case of Ibuprofen, contraindications are as follows:

- Active peptic ulcer,
- Chronic inflammation of gastrointestinal tract,
- Bleeding disorders of the gastrointestinal tract,
- Hypersensitivity to aspirin,

[12] Bindler, Ruth, and Bayne, T.: Medication calculation ability of registered nurses; Image: J Nurs Sch, 23: 221-3, 1991.
[13] Medical Mixups a Growing Concern; Boston Globe, 7 April 1995, p. 1.

- Hypersensitivity to other nonsteroidal anti-inflammatory agents,
- Impaired renal or hepatic function,
- Predisposition to fluid retention, and
- Third trimester of pregnancy (may cause malformation of fetal cardiovascular system such as premature closure of ductus arteriosus).

Dosage Recommendations. The recommended dosages of Ibuprofen are as follows:

- Acute rheumatoid arthritis, adults: 300-800 mg 3-4 times/day;
- Mild/moderate pain or primary dysmenorrhea, adults and the elderly: 400 mg every 4-6 hours;
- Fever, minor aches, adults and the elderly: 200-400 mg every 4-6 hours; and
- Children: 5-10 mg/kg/dose, maximum 40 mg/kg/day.

Drug Interactions. Ibuprofen may interact with certain drugs taken simultaneously:

- Increasing the effects of oral anticoagulants,
- Increasing the effects of heparin,
- Increasing the effects of thrombolytics,
- Decreasing the effects of antihypertensives,
- Decreasing the effects of diuretics,
- Increasing risks of gastrointestinal side effects of salicylates,
- Increasing risk of bone marrow suppression with depressants,
- Increasing toxicity of lithium, and
- Increasing toxicity of methotrexate.

Effect on Laboratory Results. Ibuprofen may alter the results of some diagnostic laboratory tests, such as the following:

- May prolong bleeding time,
- May affect blood glucose level,
- May increase blood urea nitrogen result,
- May increase creatinine result,
- May increase potassium result,
- May elevate liver function tests, and
- May decrease hemoglobin and hematocrit results.

Adverse Reactions. Ibuprofen may cause the following adverse reactions:

- <u>Acute side effects</u>:
  a) Metabolic acidosis
  b) Peptic ulcer and
  c) Gastrointestinal bleeding;

- Chronic toxicity:
  a) Hepatic dysfunction, such as cholestasis and jaundice
  b) Nephrotoxicity, such as dysuria, hematuria, and proteinuria; and
  c) Hypersensitivity reaction.

Patient Education. Patients taking Ibuprofen should be aware of the following information:

- Avoid aspirin;
- Avoid alcohol;
- If gastric upset is noted, take it with food, antacid or milk; and
- Report blurred vision.

Summary: Inspection of this list of facts relevant to Ibuprofen is impressive:

- 5 brand names,
- 8 contraindications,
- 4 dosage recommendations,
- 9 drug interactions,
- 7 effects on laboratory data results,
- 6 adverse reactions, and
- 4 patient education points.

This makes the total number of relevant facts for Ibuprofen 43. Ibuprofen may not be the best example to illustrate the amount of drug-related information to be memorized by the physician, but it is obvious that it would be an exercise in futility to expect 30,000-40,000 similar pieces of drug-related information to be memorized by each practitioner.

A recent article in the *Journal of the American Medical Association* presents a clear picture of the problems with drug therapy, urging medicine to use high technology for better control of this painful problem.

7-2. Potential Use of Information Technology in Managing Drug Therapy

The essential first step in using information technology in managing drug therapy should be, of course, to liberate the patient's clinical fact data from the constraints of the paper record. Thus, the first step is to extract and digitize the clinical data automatically. In the ambulatory care setting, the 55 ± 15 clinical facts include the following:

- Medications the patient is taking,
- Allergies and drug sensitivities of the patient,
- Diagnosis and/or impression,
- New drugs ordered, and
- Lab tests ordered.

These data could be automatically transmitted to the pharmacist's computer for matching with the prestored drug knowledge bank to detect dosage error, drug interaction, and allergy problems, and it could promptly return any concerns to the physician's office computer.

This type of system should put the practitioner at ease that the proposed drug therapy is not violating any of the established risk reduction rules. It should also lower malpractice insurance premiums and the cost to the HMO in the form of treatment of drug-induced diseases. Last, but not least, this would protect the patient from drug errors.

HMOs with their own pharmacies would find it particularly easy to implement such a safe drug therapy system.

8. COMPUTER-AIDED MANAGEMENT OF CHRONIC DISEASES

Recent reliable studies[14] have reported that 34 million Americans have a chronic illness that limits their normal activities adjusted for their age group. Forty percent of these cases are under age 65, and their medical care is around 48 percent of the total national budget! Further, these cases account for 60 percent of all hospital admissions.[15]

[14] LaPlante, M.: The demographic of disability; Milbank Q, Suppl: 55-77, 1991.
[15] Lugne, J.L., et al.: Quality Management in Health Care; 5 (1): 17-24, 1996.

It is widely recognized among physicians that a *formal effective clinical paradigm* is lacking for chronic diseases and that creative thinking is needed for a clinically useful information infrastructure for optimizing care management decisions.

The term "chronic disease" covers a broad spectrum of clinical conditions. For these diseases there is currently no effective cure, and characteristically the condition leads to progressive decline of the patient's health, until death.

For this discourse, we have compiled a partial "short list" of more frequent chronic diseases. This list is certainly incomplete, representing only the major illnesses. It's assembled only to illustrate the breadth of the clinical diagnosis.

1.	AIDS	20.	Head trauma with brain injury
2.	Alcoholism/drug dependence	21.	Hemophilias and clotting disorders
3.	Alzheimer's disease	22.	Osteoarthritis requiring hip replacement or knee replacement
4.	Asthma	23.	Huntington's disease
5.	Chromosome disorders	24.	Hypercholesterolemia, dyslipidemias
6.	Chronic anemias	25.	Inborn errors of metabolism
7.	Chronic cephalalgias	26.	Ischemic heart disease, cardiac revascularization
8.	Chronic infectious diseases, such as tuberculosis	27.	Malignancies
9.	Chronic low back pain	28.	Multiple sclerosis
10.	Chronic nephropathies	29.	**Muscular dystrophy**
11.	Chronic obstructive pulmonary disease	30.	Myasthenia gravis
12.	Chronic psychiatric diseases, such as depression and schizophrenia	31.	Organ transplantation, such as liver, kidney, or heart
13.	Chronic toxicity, such as lead	32.	Osteoporosis
14.	Congestive heart failure	33.	Parkinsonism
15.	Coronary heart disease	34.	Peptic ulcer disease
16.	Crohn's disease and other inflammatory bowel diseases	35.	Psoriasis
17.	Diabetes mellitus	36.	Rheumatoid arthritis/rheumatic diseases
18.	Diverticular disease of colon	37.	Sickle cell disease
19.	Epilepsy, seizure disorders	38.	Systemic lupus erythematosus

Most of these chronic diseases require a *health care team*: (1) primary physician(s) to optimize functional capability, (2) special physician(s) to manage pain and discomfort, (3) pharmacists to dispense and monitor drug therapy, (4) psychologist support, (5) social assistance, (6) environmental help, and (7) spiritual guidance.[16]

Typically, all seven of these different care services operate quite independently. The patient's electronic record should be the primary depository of all information generated by the health care team. Presently, however, paper records pose numerous communication problems: access often difficult, manual record transfer often fails, poor legibility, high cost (labor-intensive) of record storage, record pulling and refiling; all this is usually disjointed and unorganized.

Electronic patient records offer many potential advantages for managing chronic diseases.

- Access: Regardless of the actual site where various records are generated, the complete electronic record is instantly available from any site.
- **Information organization: The records generated by different care providers would be algorithmically organized and integrated to form a single information continuum, that must be easy to read and comprehensive. A longitudinal record, which is algorithmically coordinated, is an essential step.**
- **Life-long health care file: The record automatically links all documentation on the same person, regardless of the geographic site or time of origin of care.**
- Automated risk monitoring and quantification of disease progression: The records provide serial monitoring of biochemical and functional status of organs and functions that are characteristically affected by the particular illness. For example, patients with

[16] Phillips-Harris, C.: Integration of primary care and case management in chronic disease; Qual Manage

coronary heart disease should be monitored for increasing cardiac damage, and patients with diabetes mellitus should be monitored for albuminuria, glycohemoglobin level, cardiovascular functions, and the like. The computer would algorithmically compare past measurements with the current findings and apply a conventional trend calculation as a function of time. The slope of the trend would indicate the pace of progression of the illness. The primary physician's office computer could carry out the trend calculations, at no extra cost.

- Benefits: Such an aggressive information infrastructure should gradually convert the management of chronic diseases from art to science. The members of the health care team could know the comments and observations of all other members, a great help in coordinated patient care. The accumulating clinical experience could also be used for developing *optimal management protocols*, in a parsimonious manner, guided by collective experience with many cases.
- **Cost considerations and feasibility: Until recently, a complete automated electronic patient record system was merely an aspiration, a dream. Recently, this essential first step has become a reality. Further software development will be required for trend analysis, since each disease will require an illness-specific approach. The technology is now available. The result would be spectacular, the cost saving immense, and human suffering much reduced.**

9. AUTOMATION OF PATIENT EDUCATION

In the traditional paper record-based clinical care management system, the physician's verbal instructions to patients are usually only briefly documented, expecting the patient to absorb the instructions and remember them accurately at home. This centuries old method of patient instructions is a weak part of our health care. Patients are often nervous

---

Health Care, 5: 1-6, 1996.

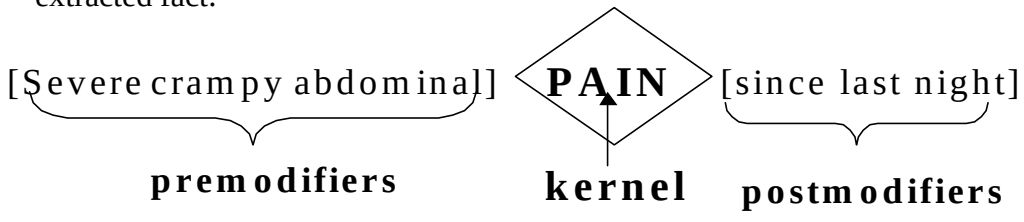
or tense in the doctor's office, their attention is reduced and much of the medical explanation and instruction is complex; it becomes inadequately memorized. Further, this patient instruction segment may use up eight to ten minutes, between 30 and 50 percent of the valuable encounter time.

A sound plan would be to automate the patient instruction segment. Presorted instructions could be algorithmically retrieved as the physician dictates his documentation of the office visit, and the pharmacist's computer could print out the drug-related instructions, including information about the drugs dispensed, their therapeutic purpose, the way to take the medications, and so on; these could be integrated with the physician's instructions. The benefit of this low-cost service would not only improve the effect of clinical care, but patient participation would be strengthened and patient satisfaction enhanced.

10. DATABASE OF ELECTRONIC CASE HISTORIES

Before describing the ways user questions will be answered by a database storing the electronic medical record system, it may be useful to see how the database is constructed. The Medical Text Processor first finds and extracts each clinical fact present in the patient record. Then, as the last step, it uniquely codes all the component elements of the fact. In more specific terms, the Medical Text Processor begins with positive identification of the type of record, such as ambulatory record, discharge summary, or pathology report; then the extracted facts receive the appropriate markers. These out of context ("naked") facts extracted consist of three components:

- The kernel: the medical term(s);
- The premodifiers: modifiers preceding the kernel; and
- The postmodifiers: modifiers following the kernel. Here is an example of an extracted fact:

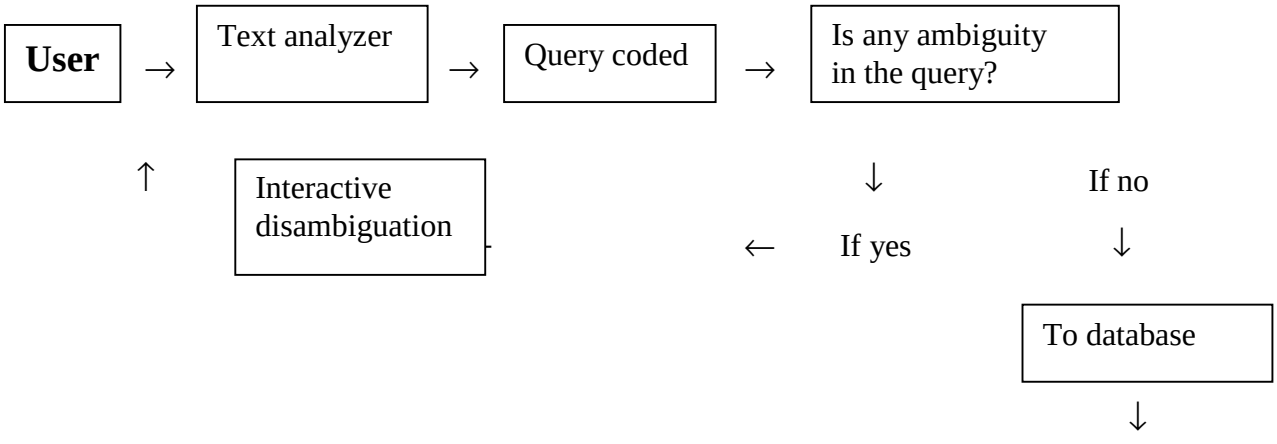


<u>Experience retrieval</u>, as we see it, will be limited primarily to the kernel of the facts, the medical term. We need not tool up for any selective modifier retrieval since nobody would ask for all facts that are "severe" or "occasional," but the database should respond to "severe headache" as a subset of "headache" with the modifier "severe."

Recently we built a modest experimental fact bank, mostly for demonstration purposes, with only about 400,000 clinical facts for about 4,000 patients. The user may choose the starting point (a symptom, a sign, a diagnosis, or a drug) and ask for a target, such as Zestril → diagnosis, or Minipress → diagnosis. Our plan, however, is more ambitious. Since we have an automated medical text interpreter, we can place it at the user interface, allowing the user to formulate his query with the wording of his choice. But we also know that formulation of a query is not an easy task. For example, we have received a query: "What is the best drug for treatment of acute severe diarrhea?" The database can generate the list of antidiarrheal drugs, but the modifier "best" is still ambiguous:

- Best because least expensive?,
- Best because of minimal side effects?, or
- Best because it is most potent?

This line of reasoning led us to the following format for handling queries:



Compile answer

Figure 3. Analysis of a query.

It is apparent that the construction of an "intelligent" clinical experience database that is able to return useful answers, including statistical inferences, is an exciting challenge involving both retrieval expertise and clinical know-how.

THE SUMMARY PAGE

After the computer completes the processing of an ambulatory care record, it algorithmically compiles the summary page. The structure of the summary page is displayed below.

We recommend that those summary pages be stored in a clinical data bank, to be available night and day. When patients receive care by another physician or are hospitalized with an acute condition, the patient can remind the physician that their past history can be accessed with a simple phone call. For security, a protocol has to verify that the call is legitimate, and made with the patient's consent.

Special summary pages for specialties such as obstetricians, ophthalmologists, cardiologists, etc. are being developed.

The summary page bank can be used

- by consultants
- for correspondence
- emergency care
- administrative statistics
- actuarial calculations
- quality of care monitoring

The design of the Summary Page is as follows:

Patient identifiers

Current medications	**Allergies**
Last visit date	**List of previous diagnoses/problems,**
Previous visit dates	**each with current status**
Open	

GENERATION OF ELECTRONIC RECORD FOR DIABETIC PATIENTS

Following is an example of the computer processing and office record of a patient with diabetes.

The names of both the patient and the physician are removed to protect confidential data.

ID # 471792

Admission Date: 7-10-94
Discharge Date: 8-9-94
Chief Complaint: Ulceration of left toe, fever

History of Present Illness: This is one of several admissions. This 49 year old black male has a history of insulin-dependent diabetes mellitus and chronic ulcerations of lower extremities. Past history of right toe ulcer, followed by right below-knee amputation and prosthesis one year ago. Admitted this time for ulceration of left fifth toe. Has three Day history of persistent fever. Denies chills, night sweats.

Past Medical History: Insulin-dependent diabetes mellitus times one year. Denies any history of myocardial infarction, congestive heart failure, liver disease, tuberculosis, or significant other history.

Medications: Insulin 20 units Nph every morning. No known allergies.

Family History: Positive for diabetes.

Social History: Negative for ethyl alcohol and smoking.

Physical Examination: Vital signs stable. Febrile 100.4.
Heent: Within normal limits.
Neck: Supple, without lymphadenopathy, thyromegaly or masses. No bruits. No jugular venous distention.
Lungs: Clear to auscultation and percussion.
Heart: Regular rate and rhythm, no murmurs, rubs or gallops.
Abdomen: Good bowel sounds, not distended, no organomegaly.
Rectal: No masses; stools heme negative.
Extremities: Right below-knee amputation. Stump is viable, warm. Incision is well healed.
Left Foot: Ulceration between fourth and fifth digits. Web space is draining purulent material. Skin covering this area is somewhat necrotic. Femoral and popliteal pulses 1+ and equal bilaterally.
Left Leg: No dorsalis pedis or posterior tibialis pulses palpable. Left lower extremity is cool to touch from approximately the tarsal bones distally.
Neurology: Intact.

Laboratory Tests: Upon admission tests within normal limits. Blood sugars between 150 and 200.

Hospital Course: Was placed on Mefoxin 2 grams every 6 hours. Because the ulceration on left foot continued to advance proximally and showed

no signs of improvement. the antibiotics were changed to Cleocin 600 mg every six hours and every six hours and Gentamicin. In addition wound care with Betadine dressings four times a day.

On July 30th it was decided that because the necrotic area had continued to advance proximally, showing no sign of improvement, we will take him to the operating room for left below-knee amputation.

Amputation went quite well. Postoperative course was unremarkable. Six days later, the patient was transferred to rehabilitation medicine. Amputation stump was warm, viable. He will be followed up 3 times a week while in rehabilitation medicine.

The office record, in narrative text form, is not compatible with a computer. To make it compatible, we must extract the clinical facts and change their mode of representation. To achieve this goal, we must use a very “intelligent” computer with profound medical knowledge.

Following is the listing of clinical facts extracted by the Medical Text Processor, with added headings for easier review.

HEADER:

<1> **ID NUMBER:** 471792
<2> **ADMISSION DATE:** 7/10/1994
<3> **DISCHARGE DATE:** 8/9/1994
<4> **LENGTH OF STAY:** 30 DAYS
<5> 49 YEAR-OLD
<6> BLACK MALE

PERI-ADMISSION STORY:

CHIEF COMPLAINT:

<7> ULCERATION OF LEFT TOE
<8> FEVER

HISTORY OF PRESENT ILLNESS:

<9> ONE OF SEVERAL ADMISSIONS
<10> HISTORY OF INSULIN-DEPENDENT DIABETES MELLITUS
<11> HISTORY OF CHRONIC ULCERATIONS OF LOWER EXTREMITIES
<12> PAST HISTORY OF RIGHT TOE ULCER
<13> FOLLOWED BY RIGHT BELOW-KNEE AMPUTATION AND PROSTHESIS 1 YEAR AGO
<14> ADMITTED THIS TIME FOR ULCERATION OF LEFT FIFTH TOE
<15> THREE DAY HISTORY OF PERSISTENT FEVER
<16> DENIES CHILLS
<17> DENIES NIGHT SWEATS

PAST HISTORY:

<18> INSULIN-DEPENDENT DIABETES MELLITUS X 1 YEAR
<19> DENIES ANY HISTORY OF MYOCARDIAL INFARCTION
<20> DENIES ANY HISTORY OF CONGESTIVE HEART FAILURE
<21> DENIES ANY HISTORY OF LIVER DISEASE
<22> DENIES ANY HISTORY OF TUBERCULOSIS
<23> DENIES ANY SIGNIFICANT OTHER HISTORY

<24> **MEDICATIONS:** INSULIN NPH 20 UNITS EVERY MORNING

<25> NO KNOWN ALLERGIES

<26> **FAMILY HISTORY:** POSITIVE FOR DIABETES

SOCIAL HISTORY:

<27> NEGATIVE FOR ETHYL ALCOHOL
<28> NEGATIVE FOR SMOKING

PHYSICAL EXAMINATION:

<29> VITAL SIGNS STABLE
<30> FEBRILE 100.4

<31> **HEENT:** WITHIN NORMAL LIMITS

NECK:

<32> NECK SUPPLE
<33> NECK WITHOUT LYMPHADENOPATHY
<34> NECK WITHOUT THYROMEGALY
<35> NECK WITHOUT MASSES
<36> NECK NO BRUITS
<37> NO JUGULAR VENOUS DISTENTION

<38> **LUNGS:** CLEAR TO AUSCULTATION AND PERCUSSION

HEART:

<39> REGULAR RATE AND RHYTHM
<40> NO MURMURS RUBS OR GALLOPS

ABDOMEN:

<41> GOOD BOWEL SOUNDS
<42> ABDOMEN NOT DISTENDED
<43> NO ORGANOMEGALY

RECTAL:

<44> RECTAL NO MASSES
<45> STOOLS HEME NEGATIVE

EXTREMITIES:

<46> RIGHT BELOW-KNEE AMPUTATION
<47> STUMP VIABLE, WARM
<48> INCISION WELL-HEALED
<49> LEFT FOOT: ULCERATION BETWEEN FOURTH AND FIFTH DIGITS
<50> WEB SPACE DRAINING PURULENT MATERIAL
<51> SKIN COVERING THIS AREA IS SOMEWHAT NECROTIC
<52> FEMORAL AND POPLITEAL PULSES 1+ AND EQUAL BILATERALLY
<53> LEFT LEG: NO DORSALIS PEDIS OR POSTERIOR TIBIALIS PULSES PALPABLE
<54> LEFT LOWER EXTREMITY IS COOL TO TOUCH FROM APPROXIMATELY TARSAL BONES DISTALLY

<55> **NEUROLOGY:** INTACT

INITIAL DIAGNOSTIC WORK-UP:

LABORATORY TESTS UPON ADMISSION:

<56> TESTS WITHIN NORMAL LIMITS
<57> BLOOD SUGARS BETWEEN 150 AND 200

CLINICAL COURSE:

<58> PLACED ON MEFOXIN 2 GRAMS EVERY 6 HOURS
<59> BECAUSE ULCERATION ON LEFT FOOT CONTINUED TO ADVANCE PROXIMALLY
<60> ULCERATION ON LEFT FOOT SHOWED NO SIGN OF IMPROVEMENT
<61> ANTIBIOTICS WERE CHANGED TO CLEOCIN 600 MG EVERY 6 HOURS AND GENTAMICIN
<62> IN ADDITION WOUND CARE WITH BETADINE DRESSINGS 4X A DAY
<63> ON 7/30/1994 IT WAS DECIDED THAT BECAUSE NECROTIC AREA HAD CONTINUED TO ADVANCE PROXIMALLY
<64> NECROTIC AREA SHOWING NO SIGN OF IMPROVEMENT
<65> WE WILL TAKE HIM TO OPERATING ROOM FOR LEFT BELOW-KNEE AMPUTATION
<66> AMPUTATION WENT QUITE WELL
<67> POSTOPERATIVE COURSE WAS UNREMARKABLE

DISCHARGE INSTRUCTIONS:

<68> SIX DAYS LATER TRANSFERRED TO REHABILITATION MEDICINE
<69> AMPUTATION STUMP WAS WARM VIABLE
<70> WILL BE FOLLOWED UP 3X A WEEK WHILE IN REHABILITATION MEDICINE

After extracting all these facts, the same computer automatically codes each clinical fact. The code must be unique for each clinical fact, and accurate, in order to accurately conserve the intent of the physician.

The extracted clinical facts shown on the previous pages were automatically encoded to make the clinical information computer compatible.

To demonstrate the accuracy of the codes, the assigned meaning of each code was printed next to each code.

==========

HEADER:

```
<1>    2-1-2-4-1-1:8-20#471792
           2-1-2-4-1-1 = PATIENT'S ID ...
           8-20#471792 = NUMERIC QUALIFIER
<2>    6-1-25-4-4-4:8-20#7/10/1994
           6-1-25-4-4-4 = ADMITTED ON ...
           8-20#7/10/1994 = NUMERIC QUALIFIER
<3>    6-1-29-1-4-4:8-20#8/9/1994
           6-1-29-1-4-4 = DISCHARGED ON ...
           8-20#8/9/1994 = NUMERIC QUALIFIER
<4>    6-1-2-3-1:8-20#30
           6-1-2-3-1 = LENGTH OF STAY ... DAY(S)
           8-20#30 = NUMERIC QUALIFIER
<5>    2-3-1-1:8-20#49
           2-3-1-1 = ... YEARS OLD
           8-20#49 = NUMERIC QUALIFIER
<6>    2-16-8-1-2-1
           BLACK MALE
```

PERI-ADMISSION STORY:

```
<7>    4-2-1-12-1-1-1:8-12-1
           4-2-1-12-1-1-1 = ULCERATION OF TOE
           8-12-1 = LEFT
<8>    4-1-16-1
           FEVER

<9>    6-1-25-4-2-2-9-1
           ONE OF SEVERAL ADMISSIONS
<10>   4-3-6-1-2-1-1:8-33
           4-3-6-1-2-1-1 = INSULIN-DEPENDENT DIABETES MELLITUS
           8-33 = HISTORY OF
<11>   4-3-1-1-19-6:8-28-10:8-33
           4-3-1-1-19-6 = LEG ULCER
           8-28-10 = CHRONIC
           8-33 = HISTORY OF
<12>   4-2-1-12-1-1-1:8-12-2:8-33-2-1-1
           4-2-1-12-1-1-1 = ULCERATION OF TOE
           8-12-2 = RIGHT
           8-33-2-1-1 = PAST HISTORY OF
<13>   6-4-3-12-8-1-2-2:6-4-21-1-108:8-12-2:8-1-7-2-12:8-1-13-7
           6-4-3-12-8-1-2-2 = BELOW-KNEE AMPUTATION
           6-4-21-1-108 = PROSTHESIS
           8-12-2 = RIGHT
           8-1-7-2-12 = 1 YEAR AGO
           8-1-13-7 = FOLLOWED BY
<14>   6-1-25-4-2-2#2:8-1-29-3
           6-1-25-4-2-2#2 = ADMITTED FOR {4}
           8-1-29-3 = AT THIS TIME
       4-2-1-12-1-1-1:1-2-9-5-8-5:8-12-1
           4-2-1-12-1-1-1 = ULCERATION OF TOE
           1-2-9-5-8-5 = FIFTH TOE
           8-12-1 = LEFT
```

```
<15>   4-1-16-1:8-6-6-1:8-33-4-2:8-20#3
           4-1-16-1 = FEVER
           8-6-6-1 = PERSISTENT
           8-33-4-2 = ... DAY HISTORY OF
           8-20#3 = NUMERIC QUALIFIER
<16>   4-1-16-3:8-29-20-2
           4-1-16-3 = CHILLS
           8-29-20-2 = DENIES
<17>   4-1-1-4-8:8-29-20-2
           4-1-1-4-8 = NIGHT SWEATS
           8-29-20-2 = DENIES
```

PAST HISTORY:

```
                                    4-2-31-1
                                        PAST MEDICAL HISTORY
<18>   4-3-6-1-2-1-1:8-33:8-1-7-4-3
           4-3-6-1-2-1-1 = INSULIN-DEPENDENT DIABETES MELLITUS
           8-33 = HISTORY OF
           8-1-7-4-3 = FOR 1 YEAR
<19>   4-3-4-2-3-1-1-4:8-29-20-6
           4-3-4-2-3-1-1-4 = MYOCARDIAL INFARCTION
           8-29-20-6 = DENIES ANY HISTORY OF
<20>   4-3-4-1-1-4:8-29-20-6
           4-3-4-1-1-4 = CONGESTIVE HEART FAILURE
           8-29-20-6 = DENIES ANY HISTORY OF
<21>   4-3-5-2-4:8-29-20-6
           4-3-5-2-4 = DISEASES OF LIVER
           8-29-20-6 = DENIES ANY HISTORY OF
<22>   4-3-22-1-35-1:8-29-20-6
           4-3-22-1-35-1 = TUBERCULOSIS
           8-29-20-6 = DENIES ANY HISTORY OF
<23>   8-29-20-6-1
           DENIES ANY SIGNIFICANT OTHER HISTORY

<24>   6-13-7-3-1-2-4-X:6-12-14-1-10:8-20#20:6-12-14-4-9:6-12-11-24-4
           6-13-7-3-1-2-4 = NPH INSULIN
           6-12-14-1-10 = ... UNITS
           8-20#20 = NUMERIC QUALIFIER
           6-12-14-4-9 = EVERY MORNING
           6-12-11-24-4 = CURRENT MEDICATIONS

<25>   4-3-15-4-11
           NO KNOWN ALLERGIES

<26>   4-1-26-1-4-3:8-15-1-1-2
           4-1-26-1-4-3 = FAMILY HISTORY OF DIABETES
           8-15-1-1-2 = POSITIVE FOR

<27>   2-16-6-4
           NONDRINKER
<28>   2-16-5-6
           NON-SMOKER
```

==========

PHYSICAL EXAMINATION:

<29> 4-2-0-15-0-1
VITAL SIGNS STABLE
<30> 4-2-0-15-1-2-4-1:8-20#100.4
4-2-0-15-1-2-4-1 = FEVER ... DEGREES F
8-20#100.4 = NUMERIC QUALIFIER

<31> 4-2-0-5-10-0
HEENT NORMAL

<32> 4-2-0-6-0-1
NECK: SUPPLE
<33> 4-2-0-6-0-6-1
NECK: NO LYMPHADENOPATHY
<34> 4-2-0-6-0-8
NECK: THYROID NOT ENLARGED
<35> 4-2-0-6-0-6-2
NECK: NO MASSES
<36> 4-2-0-6-0-4
NECK: NO BRUITS
<37> 4-2-0-6-0-3-2
NECK: NO JUGULAR VENOUS DISTENTION

<38> 4-2-3-7-0-2-2
LUNGS CLEAR TO AUSCULTATION AND PERCUSSION

<39> 4-2-4-0-8
HEART REGULAR RATE AND RHYTHM
<40> 4-2-4-0-6
HEART WITHOUT MURMURS, RUBS OR GALLOPS

<41> 4-2-5-5-0-6-4
GOOD BOWEL SOUNDS
<42> 4-2-5-5-9:8-29-2
4-2-5-5-9 = ABDOMINAL DISTENTION
8-29-2 = NOT
<43> 4-2-5-5-0-3
ABDOMEN WITHOUT ORGANOMEGALY

<44> 4-2-5-6-0-3
RECTAL EXAMINATION: NO MASSES
<45> 4-2-5-6-0-2
RECTAL EXAMINATION: HEME NEGATIVE STOOLS

<46> 4-2-0-11-2-3-2:8-12-2
4-2-0-11-2-3-2 = BELOW-KNEE AMPUTATION
8-12-2 = RIGHT
<47> 4-2-0-11-2-3-2-1
STUMP VIABLE, WARM
<48> 4-2-1-4-2-1-1
INCISION WELL-HEALED

```
=====================================================================
       <49>   4-3-1-1-1-6-1#2:8-20-4:8-20-5:8-12-1
                  4-3-1-1-1-6-1#2 = FOOT ULCERATION BETWEEN DIGITS
                  8-20-4 = FOURTH
                  8-20-5 = FIFTH
                  8-12-1 = LEFT
       <50>   4-2-45-7-1-1
                  WEB SPACE DRAINING PURULENT MATERIAL
       <51>   4-2-1-2-2-5-2:8-3-2-4:8-32-33231-1
                  4-2-1-2-2-5-2 = SKIN NECROSIS
                  8-3-2-4 = SOMEWHAT
                  8-32-33231-1 = COVERING THIS AREA
              4-2-1-4-9:8-3-2-4:8-32-33231-1
                  4-2-1-4-9 = SKIN NECROSIS
                  8-3-2-4 = SOMEWHAT
                  8-32-33231-1 = COVERING THIS AREA
       <52>   4-2-4-4-6-8-1-1
                  FEMORAL PULSE 1+ AND EQUAL BILATERALLY
              4-2-4-4-6-9-1-1
                  POPLITEAL PULSE 1+ AND EQUAL BILATERALLY
       <53>   4-2-4-4-6-12-3:1-2-9#2:8-12-1
                  4-2-4-4-6-12-3 = DORSALIS PEDIS PULSES NOT FELT
                  1-2-9#2 = LEG
                  8-12-1 = LEFT
              4-2-4-4-6-11-7:1-2-9#2:8-12-1
                  4-2-4-4-6-11-7 = NO POSTERIOR TIBIAL PULSE PALPABLE
                  1-2-9#2 = LEG
                  8-12-1 = LEFT
       <54>   4-2-0-11-2#1-2:8-12-1:1-3-1-3-2-2-5:8-32-1251-1:8-24-13-3
                  4-2-0-11-2#1-2 = LOWER_EXTREMITY COOL TO TOUCH FROM {1-
3-1-3-
                    2}
                  8-12-1 = LEFT
                  1-3-1-3-2-2-5 = TARSAL BONES
                  8-32-1251-1 = APPROXIMATELY
                  8-24-13-3 = DISTALLY

       <55>   4-2-11-0#2
                  NEUROLOGICALLY INTACT
```

INITIAL DIAGNOSTIC WORK-UP:

```
       <56>   3-1-4-2
                  TEST WITHIN NORMAL LIMITS
       <57>   3-4-5-1-0-1-4:8-20#150:8-20#200
                  3-4-5-1-0-1-4 = BLOOD SUGARS BETWEEN ... AND ...
                  8-20#150 = NUMERIC QUALIFIER
                  8-20#200 = NUMERIC QUALIFIER
```

==========

CLINICAL COURSE:

4-11-23
HOSPITAL COURSE

<58> 6-13-10-1-2-2-6-1.1-X:6-12-14-1-1:8-20#2:6-12-14-4-7-1:6-12-11-1-7
6-13-10-1-2-2-6-1.1 = CEFOXITIN SODIUM
6-12-14-1-1 = ... GRAMS
8-20#2 = NUMERIC QUALIFIER
6-12-14-4-7-1 = Q6H
6-12-11-1-7 = PLACED ON MEDICATION

<59> 4-3-1-1-1-6-1:8-12-1:10-1-4-102#1-3:8-24-3-1-1:8-10-6-3
4-3-1-1-1-6-1 = FOOT ULCER
8-12-1 = LEFT
10-1-4-102#1-3 = CONTINUED TO ADVANCE
8-24-3-1-1 = PROXIMALLY
8-10-6-3 = BECAUSE

<60> 4-3-1-1-1-6-1:8-12-1
4-3-1-1-1-6-1 = FOOT ULCER
8-12-1 = LEFT
4-11-16#2-1
{4} SHOWED NO SIGN OF IMPROVEMENT

<61> 6-13-10-1-4-2-2-X:6-12-14-1-1-1:8-20#600:6-12-14-4-7-1:6-12-11-48-2#1-2
6-13-10-1-4-2-2 = CLINDAMYCIN
6-12-14-1-1-1 = ... MG
8-20#600 = NUMERIC QUALIFIER
6-12-14-4-7-1 = Q6H
6-12-11-48-2#1-2 = ANTIBIOTICS CHANGED TO {6-13}
6-13-10-1-6-1:6-12-11-48-2#1-2
6-13-10-1-6-1 = GENTAMICIN
6-12-11-48-2#1-2 = ANTIBIOTICS CHANGED TO {6-13}

<62> 6-13-8-10-7-9-3-2-X:6-12-14-2-16:6-12-14-4-7
6-13-8-10-7-9-3-2 = POVIDONE-IODINE
6-12-14-2-16 = DRESSING
6-12-14-4-7 = QID
6-4-17-9-1#1:8-2-2-1-2-1
6-4-17-9-1#1 = WOUND CARE WITH {6-13}
8-2-2-1-2-1 = IN ADDITION

<63> 4-2-1-2-2-5-2-1:10-1-4-102#1-3:8-24-3-1-1:8-10-6-3:8-1-46:8-20#7/30/1994:10-1-4-1738-5
4-2-1-2-2-5-2-1 = AREA OF NECROSIS
10-1-4-102#1-3 = CONTINUED TO ADVANCE
8-24-3-1-1 = PROXIMALLY
8-10-6-3 = BECAUSE
8-1-46 = ON ...
8-20#7/30/1994 = NUMERIC QUALIFIER
10-1-4-1738-5 = IT WAS DECIDED THAT

<64> 4-2-1-2-2-5-2-1
AREA OF NECROSIS
4-11-16#2-1
{4} SHOWED NO SIGN OF IMPROVEMENT

==========

<65> 6-4-3-12-8-1-2-2:8-12-1
6-4-3-12-8-1-2-2 = BELOW-KNEE AMPUTATION
8-12-1 = LEFT
6-1-29-10#7-1
WE WILL TAKE PATIENT TO OPERATING ROOM FOR {6}
<66> 6-4-22-1-1#1
AMPUTATION WENT QUITE WELL
<67> 6-4-22-11-0
NORMAL POSTOPERATIVE COURSE

DISCHARGE INSTRUCTIONS:

<68> 6-1-20-3-6#1
REHABILITATION MEDICINE
6-1-29-2-1#2:8-1-4-7-13-1:8-20#6
6-1-29-2-1#2 = TRANSFERRED TO {6}
8-1-4-7-13-1 = ... DAYS LATER
8-20#6 = NUMERIC QUALIFIER
<69> 6-4-17-9-7-6-1
AMPUTATION STUMP WARM, VIABLE
<70> 6-1-20-3-6#1
REHABILITATION MEDICINE
6-1-29-2-1#3:8-1-5-8:10-1-5-383-124
6-1-29-2-1#3 = FOLLOWED UP WHILE IN {6}
8-1-5-8 = 3 TIMES A WEEK
10-1-5-383-124 = WILL BE

SOME TECHNICAL ASPECTS OF ELECTRONIC MEDICAL RECORDS

The core technology described in this document is an automated medical text analyzer applying simultaneously

- analytical linguistics (classical grammar),
- semantic interpretation, and
- clinical knowledge

to find the information content of the record.

This core technology detects every unit of clinical information present in the patient record, extracts the information carrying linguistic structures, and encodes all the information carriers to become digital data, as envisioned by the report of the Institute of Medicine.

THE FACT THEOREM

The free narrative medical text in patient records contains discrete semantically cohesive information carrying words as units, called CLINICAL FACTS. These facts are scattered within the text through the conventional grammar of sentences. The trained eye of the physician intuitively recognizes the clinical facts, dispersed in the text, carrying these units of information. The Medical Text Processor must simulate the physician's rapid fact recognition skill and distill the clinical facts from the sentential format medium.

CHARACTERISTICS OF CLINICAL FACTS

- A clinical fact may be a word, or a semantically cohesive group of words representing a discrete clinical information entity.
- Each clinical fact has a "kernel", which is a medical term.
  - The kernel may be preceded and/or followed by modifying words.
  - These modifiers alter the lexical meaning of the kernel.
  - The total clinical information is distributed among all the word elements of the clinical fact, that is the kernel and all its modifiers.
  - To conserve the full intended meaning of the clinical fact, all component words of the fact must be kept together.

An example of a clinical fact:

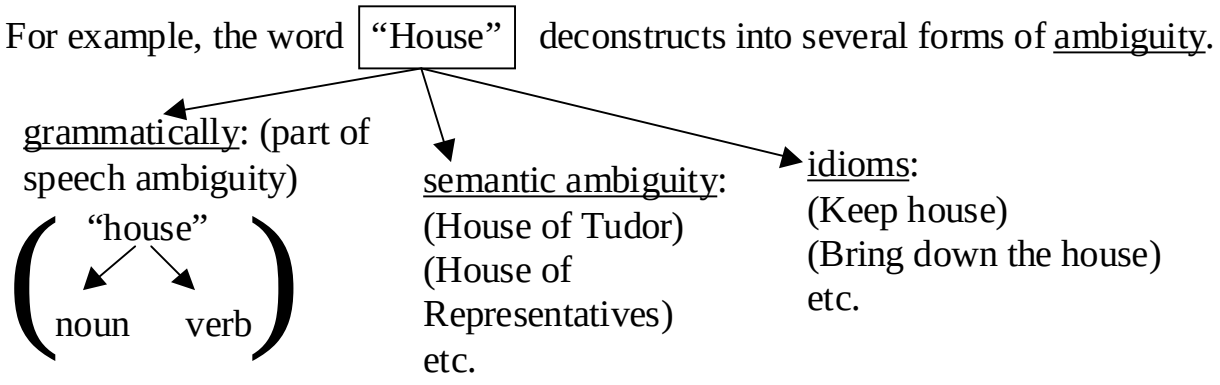
Mrs. Smith complained of severe crampy abdominal pain in lower right quadrant since last evening.

To a clinically oriented eye, this is a single unit of information. The **Kernel** is **Pain**.

In this example, the five modifiers customize the generic term "pain" with regard to its severity, kind, site, and duration in the particular clinical case, severe

The Medical Text Processor must keep the entire clinical fact together as a single information unit, so as to conserve the intended meaning of the physician

The challenge of any free natural language text processor is the inherent ambiguity of the English language, as 39% of all English words are ambiguous.



The Medical Text Processor must disambiguate all such ambiguous words using a highly developed, complex rule-based analysis to determine the correct grammatical role and/or the correct use of a word or phrase that fits the sentence semantically.

FUNCTIONS OF THE MEDICAL TEXT PROCESSOR

The Medical Text Processor analyzes the narrative text, one sentence at a time, using three engines simultaneously to find and delimit the clinical facts. Then the identified clinical facts are coded (digitized), converting the linguistic mode of the clinical facts into disciplined digital data ready for further computer processing.

Building a Lexicon for Automated Medical Text Processing

Medical Term Listings, such as ICD, CPT, DSM, and SNOMED, have permeated the world of American medicine, each serving different goals. Some of these vocabularies assist the reimbursement process (ICD or CPT); others are simply medical word listings (SNOMED). Unfortunately,

- each is limited to the purposes for which it was created...
- each is disconnected from the others by the absence of a unifying lexicon…, and
- each is unsuited for the demands of automated medical text processing.[17,18,19,20,21,22,23]

[17] Basic Terms of Anatomy and Physiology, 2nd ed.; W.B. Saunders Co., 1986.
[18] Dictionary of Medical-Legal Terms; Parthenon Publishing Group, 1997.

The Gabrieli Medical Lexicon was built specifically to support a medical text analyzing system, with linguistic and taxonomic coherence. It is a semantic network to support automated medical text processing and allows true relational database searching.

A medical lexicon constructed to support the automated medical text analyzer must have the following characteristics:

- It must be COMPREHENSIVE: it must list all medical terms, phrases, and expressions that may occur in a patient record. This includes not only the canonical (academically proper) terms, but also layman's expressions ("head cold") or outright vernacular metaphors ("my marriage ran out of gas").

In addition to the close to 400,000 medical term entries, the medical lexicon must be supported by a family of auxiliary lexicons listing all the non-medical words, phrases, idioms, abbreviations, medical-legal terms, administrative terms, and so on.

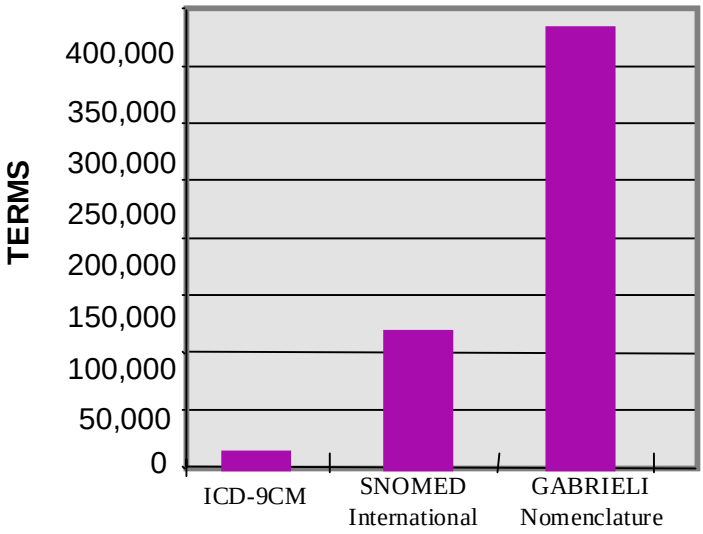


Figure 3. Scope of ICD-9CM, SNOMED and the Gabrieli lexicon.

---

[19] Kaplan-Saddock: Comprehensive Glossary of Psychiatry and Psychology; Williams and Wilkins, 1991.
[20] Pyle: Current Medical Terminology, 7th ed.; Health Professions Institute, 1998.
[21] Dictionary of Epidemiology, 3rd ed.; Oxford University Press, 1995.
[22] AHIMA: Glossary of Healthcare Terms, revised 1994.
[23] Stedman's Electronic Medical Dictionary, version 3.0; Williams and Wilkins, 1996.

- This type of lexicon must be TAXONOMICALLY SOUND. This means that the classification of the terms follows logical laws that govern the position of a term within the lexicon. Some lexicons are organized alphabetically, others by meaning and relationship to other terms with certain kinship by meaning. Dictionaries without systematization, such as the ICD dictionary, are not concerned with semantic kinships. For example, respiratory system disorders are listed in ICD under the heading of infectious diseases (by etiology) and also under neoplasms, diseases of the respiratory system, etc. This fragmentation of related terms would make information retrieval difficult. The lexicon should be organized largely as a semantic network.

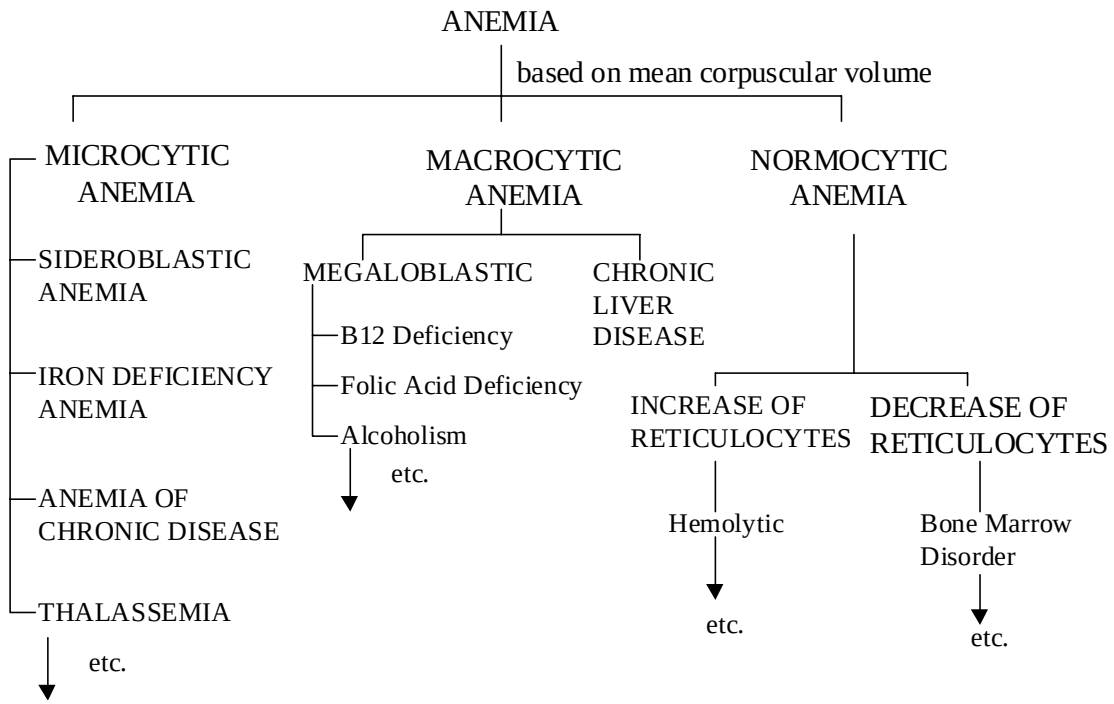
- This kind of lexicon must be SENSITIVE to medical ambiguities. A medical term may express a symptom, such as "shortness of breath", if stated by the patient, but the same expression may also mean a sign, observed by the examiner. Some medical terms indicate a medical conclusion, a diagnosis . Whenever a clinical term may indicate a symptom or a sign, or a sign as well as a diagnosis, the lexicon must list the different clinical meanings, and code them accordingly, in order to distinguish between them. It is the task of the text processor software to choose the appropriate code as the context indicates.

- This kind of lexicon must be DYNAMIC. Our medical vocabulary is an evolving entity, with 4,000 to 6,000 new terms per year. (The newest Dorland's Medical Dictionary, 28th edition reports 7,500 new terms!) Terms become obsolete, others represent new concepts. This lexicon must be easily updateable or revised.

- This kind of lexicon must render DIRECT MATCH, unambiguously, accurately.

- Our lexicon was built to be HIERARCHICALLY organized. Class terms were partitioned to increase homogeneity of the subterms. The result of the hierarchy is a semantic network.

- GRANULARITY refers to the range of terms, from the large class terms, such as "anemia" or "pain", to terms with highest specificity. To achieve such a wide range, each term must be examined for its clinical information content, with special attention to the homogeneity of the membership covered by the term. "Anemia", for example, is a class term, including different kinds of anemias. Once the term is categorized as a class term, the lexicographer keeps partitioning, resulting in subterms with increasing homogeneity of their membership. This is shown in the following illustration:



The parent term in this example is “anemia”, defined as decrease of red blood cells in circulation. Obviously, it is a class term classified either by red cell morphology or by the pathophysiologic mechanism (increased loss or inadequate production of red cells). The lexicographer’s decision is to choose one classification scheme or both. We have selected the latter.

DIGITIZING THE TERMS

At the outset, we accepted the traditional structure of medicine followed by medical schools, starting with the basics of anatomy, the foundation of the science of medicine. Following this line of reasoning, we divided the domain of medical sciences into the following six major segments:

1. Anatomy, including histology and ultrastructures;
2. Physiology, including biochemistry and biophysics;
3. Diagnostic modality, including diagnostic laboratory tests, X-rays, and the like;
4. Clinical terminology, including all abnormal clinical manifestations;
5. Etiology of illness, including microbiology, accident, hereditary factors and environment; and
6. Therapy, including surgery, drugs, therapeutic radiology, and the like.

Admittedly this primary classification scheme is neither original nor the only choice, but we found it a workable system.

THE CODE SCHEME

In our system, the first code unit indicates the primary category of the term, such as 1= anatomy or 6= therapy.

The second code unit is semantic category dependent. For example,

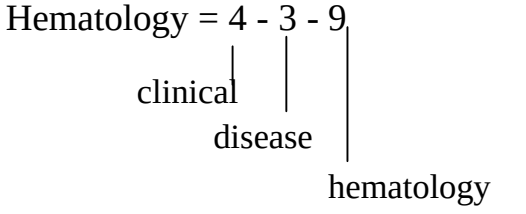
4-1 = symptom
4-2 = sign
4-3 = disease (diagnosis)
4-4 = syndrome
4-5 = abnormal courses
and so on.

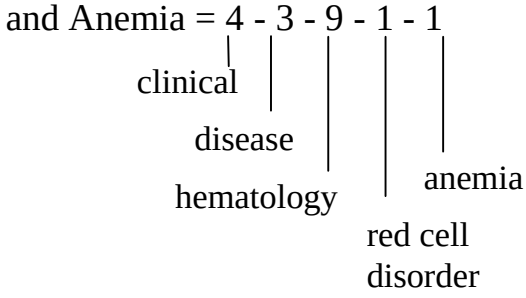
This subdivision of "4" differentiates the term by its source:

4-1 = stated by the patient: symptom
4-2 = observed by the examiner: sign
4-3 = summary conclusion: diagnosis

This source sensitivity was found essential in handling clinical facts. If the patient claims not having cardiac problems, this has a different clinical meaning than the examiner's statement of normal cardiac findings or if the EKG is "normal". The text processor must determine the source in order to assign the proper code.

The subsequent code units list the steps of subdivision leading to that particular term, such as





Thus the code chain describes the partitioning process until the term in focus is reached.

Code chains may be long—8 or 10 or more—but since they function as "medical machine language", the length is unimportant.

"LOCAL" CODE

The hierarchical code system is powerful because it describes the exact position of each medical term within the vast hierarchical tree, in a conceptual semantic space representing our entire medical language. This code places each medical term into the environment of related terms, which is most important in information retrieval. However, the hierarchical codes are unstable. A new "subdisease" to be added may disrupt a branch. Therefore, from the outset we have also assigned a "local" code to each term, an entirely random code, and all other codes are linked to this local code.

Other codes include grammatical code (part of speech), ambiguity code, cross reference to ICD, CPT, and so on; these are all attached to the local code.

THE LEXICAL SYSTEM

The lexical system includes (a) the medical lexicon and the family of nonmedical lexicons, and (b) software accepting questions from the Text Processor software and sharing all the information, medical, grammatical, semantic, and other knowledge attached to the term. The matching component of the software seeks first a perfect letter-by-letter match, then a "close match", such as antibiotics ⇒ antibiotic. Thus the lexicon is an integrated component of the Automated Text Analyzer, providing medical, syntactic, and semantic knowledge. Therefore, our medical lexical system is much larger than a dictionary, including numerous encyclopedic components as well.

AT THE OTHER END OF THE RAINBOW

At the turn of the century, American medicine is at the threshold of changing to electronic patient records. It may be appropriate to discuss briefly the expected benefit of this change.

In addition to the drastic cost saving simply because of the elimination of the manual paper records and the labor-intensive storage-retrieval of paper charts, perhaps the most important change will be the <u>experience-driven quality of clinical care</u>.

Around the turn of this century, the idea of quality of care was a non-issue. Standard medical care was what a physician—a graduate of a medical school—rendered. This implied that all physicians were equal and competent.

As medicine gradually changed, as the spectacular growth of the science and clinical care technology invited the period of malpractice suits, and as the proliferation of commercial insurance companies became molding forces, the concepts of accountability, appropriateness of care, and quality of care emerged. When the Medicare bill was passed, Congress mandated monitoring the care rendered, and thus the birth of the Professional Review Organizations. We soon learned that we lacked the basic tools for meeting the challenge.

For starters, the very definition of quality of care proved to be difficult, yet this definition should lead to assessment of the care rendered. In theory, good medical care should be in line with a gold standard which was conspicuously missing.

The leading scholar of the field, Avedis Donabedian,[24] described a gamut of components of quality of care, such as technical knowledge, judgement, and skill—elements that together result in quality of care.

Donabedian's interesting analysis led to the discussion of assessment of quality of care and he offered three categories:

- structure--the setting where the care is provided,
- process--what is actually done, and
- outcome--the result of the process.

This reasoning is interesting, but more scholarly than pragmatic.

A most important effort to summarize the current views of quality of health care was a six-part series of articles published by the New England Journal of Medicine with the leader of the field, David Blumenthal, at the helm.[25] The series clearly showed the enormous importance of making progress, but it is also apparent that a nationwide uniform upgrading, where needed, is still a desideratum rather than a reality. "To do what is right, at the right time, and in the right way" is not a pragmatic rule when we have not defined what is "right."

We know what is not right, such as mandatory continued medical education, guidelines drafted by experts, or samples of charts analyzed by physicians. The valiant effort of HCFA to monitor the quality of care of Medicare patients by massive chart review was educational, but too labor-intensive, expensive, not generalizable, and ineffective. We have attempted to circumvent the real obstruction—the fact that we are hostages of the paper records. Massive chart analysis is simply not cost effective.

[24] Donabedian, Avedis: The Quality of Care; JAMA, 260: 1743-8, 1988.
[25] Blumenthal, David: Quality of Health Care; N. Engl. J. Med., 335: 891-3, 1996.

The Mayo brothers left the important legacy of mandatory continuous review of their work in an effort to analyze the appropriateness of the indications, benefits, and negative outcomes. This rule is excellent when done on a small scale, but it is not realistic as a national program.

Practice of medicine is still a pragmatic profession. Yesterday's experience tells us what to do today. But current medicine has lost its footing, the guidance by statistically validated analysis of experience. Paper-bound experience, on a large scale, is not a realistic approach.

THE ELECTRONIC PATIENT RECORD

A. Automated low-cost computer-processing of the dictated patient records will make clinical care transparent, retrievable, and analyzable.
B. Anonymous clinical databanks will be created by removing all direct and indirect patient, physician, and care organization identifiers.
C. Algorithms will be created for each clinical entity to monitor the diagnostic appropriateness and the recommended therapy.
D. Algorithms will be created to derive the "best" diagnostic approach and the "best" therapy for a given illness at a given stage. These optimizing algorithms will be derived from the statistics of the comparison of diagnostic and therapeutic variants.
E. Physicians will be encouraged to make the patient records more detailed and personalized to enhance the accuracy of the statistical inferences.
F. Shared medical experience will be a dynamic information source, stimulating further research where knowledge is required.

As a closing comment, on a small scale such an anonymous electronic clinical experience bank has been tested to prove the hypothesis.

LONGITUDINAL HEALTH RECORDS

The life-long electronic patient record offers an entirely <u>new paradigm</u> for the management of patient care, feasible only with electronic care records.

Temporal reasoning, tomorrow's medicine, is not a new idea. In his autobiography, MY LIFE AND MEDICINE, Paul Dudley White, the famous Boston cardiologist, describes how he kept his medical records. Early in his career, he had had some large data sheets printed up, to facilitate recording the patient data. Each patient was accorded two horizontal lines. In the vertical columns, the personal data, diagnoses and etiology were recorded. Dr. White kept meticulous records, monitoring the course of the disease, and at the end, more than 15,000 longitudinal case histories

were documented, and frequently reviewed, looking for pattern similarities, relating etiology, diagnosis and therapy to the clinical course.

Changing from episodic snapshots to time-dependent clinical course will be a major change in the clinical view of chronic diseases.

In the traditional fee-for-service system, the office visit is based on a history, physical examination, and diagnostic tests to establish the symptoms and abnormal signs at the time of the encounter: a medical snapshot.

Longitudinal records allow an added temporal dimension: monitoring the PROGRESSION of the disease. In the case of diabetes mellitus, for example, the kidney, the coronary vessels, the retina, the peripheral vascular bed, and other organs may deteriorate at various rates. Electronic records can quantitate the progress of the illness. An algorithm can drive the computer to recognize the functional parameters of the specific organs affected, such as proteinuria and renal function tests for the kidneys or blood pressure measurement for hypertension.

Such an algorithm would do the following:

- ♦ link the new parameter data to the corresponding data of the past, then
- ♦ calculate the trend of progression,
- ♦ calculate the overall stage of the disease, and
- ♦ calculate the impact of adjusted therapy to change the trend.

This is a new dimension in clinical care!

The other aspects of this exciting new paradigm are these:

A. The care provider should measure the established parametric values for the particular disease, instead of the snapshot that is current and unfocused.
B. The care provider's primary role is to update the record, to act as a member of the invisible team, treating the case at a particular time.
C. Instead of the disjointed snapshots of the past, a clinical disease course continuum emerges, and the computer is an ACTIVE PARTNER in this.
D. Analysis of the accumulating databank of many longitudinal records should provide the needed information on
   - most cost-effective office visit data recording;
   - "best" therapy; and
   - **"optimal decision making", the much needed "gold standards" for many diseases.**

The longitudinal electronic health records will have the potential to drastically change the clinical management of numerous chronic diseases, such as diabetes, AIDS, parkinsonism, myasthenia gravis, hemophilia, diverticular diseases, inflammatory bowel diseases, organ transplants, and coronary heart diseases, and they will also deal with monitoring genetic data.