Original Investigations

JAMIA

Research Paper ■

Evaluation of Long-term Maintenance of a Large Medical Knowledge Base

DARIO A. GIUSE, DRNG, NUNZIA B. GIUSE, MD, MLS, RANDOLPH A. MILLER, MD

Abstract **Objective:** Evaluate the effects of long-term maintenance activities on existing portions of a large internal medicine knowledge base.

Design: Five physicians who were not among the original developers of the knowledge base independently updated a total of 15 QMR disease profiles; each updated submission was modified by a review group serving as the "gold standard," and the pre- and post-study versions of each updated disease profile were compared.

Measurements: Numbers and types of changes, defined as any difference between the original version and the final version of a disease profile; reason for each change; and bibliographic references cited by the physicians as supporting evidence.

Results: A total of 16% of all entries were modified by the updating process; up to 95% of the entries in a disease profile were affected. The two most common modifications were changes to the frequency of an entry, and creation of a new entry. Laboratory findings were affected much more often than were history, symptom, or physical exam findings. The dominant reason for changes was appearance of new evidence in the medical literature. The literature cited ranged from 1944 to the present.

Conclusions: This study provides an evaluation of the rate of change within the QMR medical knowledge base due to long-term maintenance. The results show that this is a demanding activity that may profoundly affect certain portions of a knowledge base, and that different types of knowledge (e.g., simple laboratory vs expensive or invasive laboratory findings) are affected by the process in different ways.

■ JAMIA. 1995;2:297-306.

Computerized medical knowledge bases must be revised constantly, and can never be considered com-

pletely finished. The natural evolution of the systems that use a particular knowledge base can cause older portions of the knowledge base to become outdated, or give rise to completely new subcomponents. Even more important than these intrinsic reasons for change, however, is the rapid evolution of medical knowledge itself. New diagnostic and therapeutic procedures are developed and evaluated regularly, and the results of such efforts are reported in the literature weekly. Consequently, even the best medical knowledge bases are subject to obsolescence unless a careful maintenance and updating process is implemented.

Affiliation of the authors: Vanderbilt University, Nashville, TN.

Supported in part by research grant no. R01 LM04622 from the National Library of Medicine. At the time that the work reported in this article was conducted, the authors were at the University of Pittsburgh, Pittsburgh, PA.

Correspondence and reprints: Dario A. Giuse, Drng, Division of Biomedical Informatics, Vanderbilt University Medical Center, Es-kind Biomedical Library, 2209 Garland Avenue, Nashville, TN 37232-8340. e-mail: dario.giuse@mcmail.vanderbilt.edu

Received for publication: 1/25/95; accepted for publication: 4/21/95.

The term "long-term maintenance," as used here, describes the ongoing process required to handle this constant change. Long-term maintenance involves three components: creation of new portions of a knowledge base; modification of existing portions of a knowledge base; and systematic checking to ensure consistency with externally available medical knowledge and better performance. In this paper the term "knowledge base updating" is used for the second component, i.e., the modification of existing portions of a knowledge base. The purpose of this study is to provide a quantitative evaluation of the amount of change produced over time by knowledge base updating in a large internal medicine knowledge base, and thus to give an indication of the total amount of effort expected.

Background

Interest in the creation of knowledge bases for medicine is a central theme in medical informatics,^{1,2} especially in the area of medical expert systems.³ Much of the research in this field has been in the classic area of knowledge acquisition, typically viewed as the initial creation of a set of rules or basic components⁴ by eliciting relevant information from a domain expert.⁵⁻⁷ However, from the beginning it became clear that the activities that follow the initial creation of a knowledge base require time and effort.⁸⁻¹¹

The area of knowledge base research known as knowledge refinement^{12,13} deals with the modification of rule bases after the initial version of an expert system has been deployed. The main body of knowledge refinement research has been confined to nonmedical domains of artificial intelligence (AI) research; however, two medical examples are contained in references 14 and 15. As generally defined, knowledge refinement is a subcomponent of long-term maintenance. The emphasis in knowledge refinement is typically on locating and modifying individual, incorrect rules,¹⁶ generally in response to failures in the performance of the expert system. This approach retains the flavor of some earlier work on rule-based systems, notably SEEK and SEEK-II,¹⁷⁻¹⁹ in which libraries of cases with known solutions were used to suggest to the domain expert potential improvements in the rule base. In most such systems, changes are typically dictated by failures of the system to perform as expected, rather than by the desire to keep a knowledge base up to date with external sources of knowledge. Another example of this approach is GARVAN-ES1,²⁰ an expert system for the interpretation of thyroid hormone measurements. This system uses ripple-down rules²¹ to effectively create

families of related rules, making it less likely for changes to one portion of the rule base to adversely affect other portions.

The QMR Knowledge Base

The authors' previous research has used the framework of the INTERNIST-I/QMR knowledge base (henceforth called the "QMR knowledge base"), an extensive internal medicine knowledge base whose history dates back to the early 1970s.²² The long-term process of constructing and maintaining a medical knowledge base of that size and complexity has provided much of the motivation and insight behind this research.

Work on long-term maintenance of the QMR knowledge base began in the early 1970s under the direction of Jack D. Myers, MD. From the project's inception, the creation and maintenance of the INTERNIST-I and subsequently QMR knowledge bases were modular and separate from the function of the INTERNIST-I and QMR diagnostic algorithms (Clancey²³ pointed out the benefits of such a separation in rule-based systems such as MYCIN). As such, the primary objective of knowledge base construction and maintenance was to create an academic repository of useful information regarding the diagnosis of various disorders, which could also serve as the substrate for diagnostic programs. The fact that essentially the same knowledge base could be used to support two disparate approaches to diagnostic decision support²⁴ suggests that an emphasis on medical knowledge representation, rather than on program performance, is correct. The performance of the program, thus, is but one of many feedback mechanisms used to maintain the knowledge base.

The methods used for long-term maintenance of the QMR knowledge base included four main components. First, any new description of a disease was thoroughly tested by creating and using a set of classic, "textbook" cases of that disease, and then against challenging cases with known (e.g., anatomically verified) diagnoses. Second, clinicians involved in the project updated the knowledge base as they read new articles, attended conferences, etc. Third, the clinicopathologic conference (CPC) cases discussed in *The New England Journal of Medicine*, as well as CPC cases from local and other hospitals, were entered into QMR weekly for testing. This resulted in periodic, random attention (via additional literature review) to previously completed disease profiles; failure to reach the correct diagnosis would prompt a manual revision of any incorrect disease profile. Fourth, the evolution of the knowledge base—such as the creation of a new finding—would typically prompt the re-

vision of all diseases to which the new finding should be added, again based on literature review. Changes to a finding would often need to be propagated to all diseases in the finding's differential diagnosis.

The processes described above were conducted continuously from the early 1970s to the time of this study. Additionally, on certain occasions, an existing disease profile would be systematically reviewed, through a literature survey and consultation with experts, as a concentrated activity. Unlike the other processes, this process traditionally took place only on an as-needed basis, such as when understanding of legionnaires' disease advanced during the 1980s.

Research in knowledge acquisition for the QMR knowledge base has addressed different components of the knowledge acquisition and long-term maintenance processes. An initial quantitative study documented the errors encountered during the initial creation of QMR disease profiles by physicians who had varied degrees of expertise in knowledge acquisition.²⁵ This was followed by the development of a multicenter strategy to broaden knowledge acquisition to geographically separate areas,²⁶ providing more clinical expertise and manpower for this demanding task. The feasibility of this strategy was shown not only pragmatically, but also by a quantitative study of the inter-author reproducibility of knowledge base construction under controlled circumstances.²⁷ That study showed statistically significant agreement in several key areas among physicians working independently on the same task. The implementation of a tool that helps to enforce the internal consistency of the knowledge base was reported.^{28,29} The question of what sources of medical knowledge are available for knowledge base construction was explored,³⁰ examining the depth and breadth of coverage of general and specialized medical textbooks compared with the QMR knowledge base.

Definitions

The following definitions of phrases related to the QMR knowledge base are used in the remainder of this paper. The basic unit of the QMR knowledge base is the disease profile. A disease profile contains 1) all the findings associated with a disease; and 2) links to other diseases associated with a disease. For simplicity, the term "entry" is used to refer generically to both findings and links. All QMR findings are classified as one of six types³¹: history findings, symptoms, objective signs, and three types of laboratory results that range from Lab0 (routine, inexpensive) to Lab2 (expensive and/or highly invasive). Each entry is associated with two quantitative measures, the evoking strength and the frequency.²²

The evoking strength is analogous to positive predictive value, and ranges from 0 (lowest) to 5 (highest). The frequency gives an indication of how often an entry is seen in patients who have the disease (sensitivity), and ranges from 1 (rare) to 5 (essentially universal).

Methods

Participants and Disease Profiles Selection

Five licensed, clinically active physicians volunteered to participate in the knowledge base updating process. The group included a board-certified internist who had about five years of postcertification clinical experience; two physicians in their final year of residency training in internal medicine; a board-certified neurologist who had over five years of academically based clinical experience; and a board-certified internist who was completing additional fellowship training in occupational medicine.

Based on their interests and expertise, the physicians selected from the list of candidates one or more disease profiles for updating. The list of candidate QMR disease profiles had been generated at the beginning of the study; all candidate disease profiles met the following criteria. First, candidate profiles had been created before 1985, the year when the INTERNIST-I/QMR team began preserving bibliographic references for all new disease profiles. Second, they had not been systematically revised since 1985 (although they might have been altered incrementally as part of the normal maintenance activities described in the Background section). Thus, candidate profiles had no bibliographic references, and had not been systematically updated for at least eight years.

Each physician was given the original version of the disease profile(s) he or she had selected, and a copy of the QMR-KAT knowledge acquisition tool,³² which he or she used for all subsequent editing. This tool allows a disease profile to be edited interactively with a simple menu-and-mouse interface, and performs a number of consistency checks²⁹ as a disease profile is modified. To ensure that study participants had a uniform level of familiarity with the process, the five physicians were trained at the Section of Medical Informatics, University of Pittsburgh, by one of the authors (NBG), who has trained all physicians creating new disease profiles over the past several years. Training for each physician lasted one day, as customary, and included an introduction to the QMR knowledge base, a few hours for learning how to use QMR-KAT, and, finally, the hands-on creation of a mock disease profile.

Updating Activity

The physicians worked individually on updating the disease profiles. During this process they were allowed to consult any available source of medical information (such as MEDLINE, academic medical libraries, textbooks, monographs, and expert colleagues), and typically located relevant literature references after conducting extensive bibliographic searches. In addition, they could rely on their own clinical knowledge of specific items when necessary, although use of external references was strongly encouraged.

Starting from the original version of a disease profile, the physicians added and modified information using the QMR-KAT editor. The physicians used QMR-KAT to annotate each entity and to record complete references to the articles used as evidence; annotations and bibliographic references became an integral portion of each disease profile.

Each updated disease profile submitted by a study participant was reviewed via the customary INTERNIST/QMR knowledge base group review process.²⁵ During this process, each participant met with the group of clinicians (Myers, Miller, N. Giuse, and other relevant clinicians based on subject matter) who have been responsible for maintaining the QMR knowledge base over the past 20 years. In each group review, which lasted several hours, a participant discussed each proposed finding or link change and the relevant evidence from the literature. When the knowledge base maintainers had doubts about a specific change, they asked the participant to produce further evidence until a consensus was reached. Following the methodology used in constructing the knowledge base, the purpose of each group review was not to add new knowledge, but to ensure that each proposed change was fully justified by literature evidence. Moreover, the reviewers checked for consistency of each proposed change with the rest of the knowledge base.³² Once the group review had been completed, the updated version of each disease profile replaced the original in the knowledge base. This process, as mentioned earlier, focuses primarily on academic medical knowledge regarding diagnosis of a disorder, not on system performance.

Analysis

For the purpose of this study, a change was defined as any difference between the original (pre-study) version of a disease profile and the final (post-study) version in the knowledge base, excluding the creation of references, which were absent from the original version. Thus, changes included addition or deletion

of a finding or link and modification of the numeric values associated with a finding or link. Changes included both those proposed directly by the participating physicians and those added during the final group review process. This definition is consistent with the methodology used in a previous study.²⁷

After all disease profiles had been added to the knowledge base, two of the authors (NBC, DAG) analyzed all changes, determined the reason for each change, and classified reasons for change into a number of categories (developed retrospectively). Reasons for change were determined based on annotations and literature references created by the five physicians in the updated disease profiles and original notes taken during the group review process. This analysis, which requires an understanding of the clinical reasoning leading to each change in a disease profile, was conducted manually.

In addition to the reason for each change, the authors also recorded the total number of changes by disease profile and by type of entry (six finding types, plus links). The effects of the updating process on the numeric values of evoking strength and frequency were expressed as the average rate of change for all entries in which one or both were modified. Because a numeric value can either increase or decrease during the updating process, the authors used absolute values of change for this analysis.

Finally, data were collected about literature references used by the five physicians as supporting evidence for proposed updates. This analysis used all the literature references cited by at least one physician under at least one entry that had been modified by the updating process. Using the year of publication of each reference as an indicator of how recently the relevant information had appeared in the medical literature, references were tabulated as a function of the type of entry under which they appeared.

Results

The group of five physicians selected and updated a total of 15 disease profiles (about 2.5% of all existing QMR disease profiles). In the case of carpal tunnel syndrome, two physicians who had different backgrounds—a board-certified internist who was completing a fellowship in occupational medicine and a board-certified neurologist—independently updated the same disease profile. The two updated versions were merged and considered as one during the group review process. All the remaining disease profiles were updated by one physician each.

As is the case for most new or updated disease pro-

files, the physicians worked on this task on a part-time basis. Consequently, exact values for the total time required are not available. Based on the physicians' subjective evaluations, we estimate that the average time to complete the task was around 12 days full-time equivalent. This figure is consistent with our previous experience for disease profile construction.²⁷ All of the physicians reported that most of the time was spent reading and analyzing the medical literature, again consistent with previous results.

Table 1 shows the numbers of findings and links for each of the 15 disease profiles, before and after updating. As a result of updating, the average number of findings in each disease profile increased by 4 (about 4.5%), from 90.3 to 94.4. The average number of links also increased, from 12.8 to 14.1 (an increase of about 10%). For comparison, the average number of findings per disease profile over the entire knowledge base is 80, and the average number of links per disease profile is about 15.

Table 2 shows the numbers of changes by four categories (change to evoking strength, change to frequency, addition of an entry, and deletion of an entry) and by disease profile. The average number of changes per disease profile was 25.6. This corresponds to an average of 25.2 changed entries per profile, because a total of six entries included changes in both evoking strength and frequency.

Changes in frequency outnumbered changes in evoking strength by a factor of 6 (148 to 24). The average magnitude of change in evoking strength was 1.04,

Table 1 ■

Numbers of Findings and Links for Each of the 15 Disease Profiles, before and after Updating

Profile	Findings		Links	
	Before	After	Before	After
Amyotrophic lateral sclerosis/progressive muscular atrophy	96	98	6	6
Anemia of decreased vitamin B ₁₂ absorption	94	98	8	8
Botulism	113	120	4	4
Carpal tunnel syndrome	26	39	14	14
Goodpasture's syndrome (pulmonary component)	71	74	3	3
Goodpasture's syndrome (renal component)	78	81	5	5
Guillain-Barré syndrome	133	131	16	25
Iron deficiency anemia	83	84	50	51
Myasthenia gravis	110	97	13	12
Pernicious anemia	130	135	17	19
Pleural effusion chylous	40	51	2	4
Renal tubular acidosis distal (type 1)	51	56	16	21
Renal tubular acidosis proximal (Fanconi syndrome)	38	42	7	7
Tuberculosis pleurisy	74	80	5	6
Wegener's granulomatosis	218	230	26	27
TOTAL	1,355	1,416	192	212

indicating that the typical affected entry had its evoking strength adjusted by 1 upwards or downwards (the smallest possible change). The average magnitude of change in frequency was very similar (1.05).

Table 2 ■

Numbers of Changes for All Entries (Findings and Links), by Four Categories and by Disease Profile

Profile	Evoking Strength Change	Frequency Change	Addition	Deletion	Total Changes
Amyotrophic lateral sclerosis/progressive muscular atrophy	2	15	10	8	35
Anemia of decreased vitamin B ₁₂ absorption	0	0	4	0	4
Botulism	0	20	11	6	37
Carpal tunnel syndrome	6	8	19	6	39
Goodpasture's syndrome (pulmonary component)	1	2	3	0	6
Goodpasture's syndrome (renal component)	1	8	2	0	11
Guillain-Barré syndrome	0	21	19	15	55
Iron deficiency anemia	0	0	5	3	8
Myasthenia gravis	6	26	12	26	70
Pernicious anemia	1	0	7	0	8
Pleural effusion chylous	0	8	14	1	23
Renal tubular acidosis distal (type 1)	4	12	9	1	26
Renal tubular acidosis proximal (Fanconi syndrome)	1	3	4	0	8
Tuberculosis pleurisy	1	5	8	1	15
Wegener's granulomatosis	1	20	14	4	39
TOTAL	24	148	141	71	384
RANGE	0-6	0-26	2-19	0-26	4-70

Table 3 ■

Numbers of Entries Changed in All Profiles, by Entry Type and by Type of Change*

Entry Type	Changed Evoking Strength (%)	Changed Frequency (%)	Addition (%)	Deletion (%)	Total Changed (%)	Total Entries
History	0 (0.0)	21 (6.3)	29 (8.7)	4 (1.2)	54 (16.3)	332
Symptom	1 (0.4)	14 (5.0)	5 (1.8)	15 (5.3)	35 (12.4)	282
Sign	7 (0.9)	67 (8.9)	25 (3.3)	32 (4.2)	131 (17.3)	756
Lab0	1 (0.3)	8 (2.4)	2 (0.6)	3 (0.9)	14 (4.2)	336
Lab1	5 (2.0)	19 (7.7)	51 (20.6)	10 (4.0)	85 (34.3)	248
Lab2	2 (2.4)	7 (8.5)	12 (14.6)	3 (3.7)	24 (29.3)	82
Link	8 (2.4)	12 (3.6)	17 (5.1)	4 (1.2)	41 (12.4)	331
TOTAL	24 (1.0)	148 (6.3)	141 (6.0)	71 (3.0)	384 (16.2)	2,367

*All percentages are relative to the total number of entries of each type, listed in the last column.

Table 3 presents the numbers of changes by type of entry. Findings are divided into the six customary types (history, symptom, sign, Lab0, Lab1, and Lab2); links are listed as a separate type. There are marked differences in the percentages of change for each of the six types; for example, Lab1 findings were 11 times more likely to be added to a disease profile than were symptoms. Overall, Lab0 findings (that is, results from inexpensive laboratory tests that are performed routinely in most hospitals) showed the least degree of change, whereas Lab1 and Lab2 findings changed most often as part of the updating process.

Table 4 lists the reasons, as well as their corresponding short identifiers, for changes that occurred during the updating process. Category *new-link* indicates those cases in which a newly created link to another disease profile partially explains the occurrence of a finding in some of the patients who have the disease. To avoid double-counting of the finding by the diagnostic algorithm in such cases, the original frequency of the finding is reduced accordingly. The category *earlier-pres* indicates that, due to better diagnostic techniques or improved screening, the typical initial presentation of the disease now occurs at an earlier stage. This fact tends to reduce the frequency of those findings that characterize the more advanced stages of the disease. Examples of category *new-tech* were disease profiles that originally contained an x-ray finding, and to which the corresponding computerized tomography (CT) finding was added (CT scanning regarding a specific disorder was not commonly described in the literature at the time many disease profiles were originally created). Finally, to understand the meaning of category *freq-implic*, consider that the frequency for the general finding "abdomen pain present" cannot be lower than the frequency for the more specific "abdomen pain epigastrium,"

because the latter implies the former. An increase in frequency for the specific finding requires an increase in frequency for the more general finding, and this change is classified under the *freq-implic* category.

Changes to all entries, classified according to the reason for change, are listed in Table 5. As an example of the process that leads to updates, consider the finding "dyspnea at rest" in the disease profile

Table 4 ■

Reasons, and Their Corresponding Identifiers (ID), for Changing an Entry During the Updating Process

ID	Reason for Change
new-evid	New evidence from the literature
no-support	More recent literature does not support the original entry
new-finding	A new finding was created and added to the knowledge base
place-hold	A finding was a placeholder for a disease not previously in the knowledge base, and was removed once the disease was added
new-link	Frequency was changed because a newly added link partially explains the finding
diff-diag	Evoking strength was changed for consistency while reconsidering the whole differential diagnosis of a finding
earlier-pres	Actual changes in the initial presentation of the disease from the time the original profile was created
reviewers	Changed during group review because of reviewers' consensus
link-explains	Finding was unnecessary because it was fully explained by an existing link
freq-implic	Frequency adjusted to match logical implication from a more specific finding in the same disease
new-tech	Same finding reported using a different diagnostic technique

Table 5 ■

Numbers of Changes for All Entries, by Reason for Change and by Type of Change

ID*	Changed Evoking Strength	Changed Frequency	Addition	Deletion	Total Changes (%)
new-evid	13	121	81	12	227 (59.1)
no-support	1	12	0	54	67 (17.4)
new-finding	0	1	59	0	60 (15.6)
place-hold	0	0	0	2	2 (0.5)
new-link	0	1	0	0	1 (0.3)
diff-diag	9	0	0	0	9 (2.4)
earlier-pres	0	5	0	0	5 (1.3)
reviewers	1	5	1	1	8 (2.1)
link-explains	0	0	0	2	2 (0.5)
freq-implic	0	2	0	0	2 (0.5)
new-tech	0	1	0	0	1 (0.3)
TOTAL	24	148	141	71	384 (100)

*The meaning of each identifier (ID) appears in Table 4.

for amyotrophic lateral sclerosis/progressive muscular atrophy. The finding was originally marked as having a frequency value of 2, indicating occurrence in a significant minority of cases. After the knowledge updating process, the frequency value was changed to 1, indicating that dyspnea at rest is seldom seen in patients who have amyotrophic lateral sclerosis. The physician who updated the disease profile had to go back to the early literature to find primary data about this finding. This is shown by the annotation "Uncommon at onset but nearly universal in the latter stages of the disease," in which he referenced an article published in 1944.³³ In his annotation, the physician also stated that the article reported dyspnea at rest in only two of 81 patients. As usual, the physicians assigned a so-called "pertinence number" to each reference to each article.³² This number, which is kept as a part of all disease profiles created with QMR-KAT, expresses a physician's evaluation of the reliability and "pertinence" of each cited article to a fact in the disease profile. The pertinence number allows well-described randomized clinical trials, for example, to be flagged as more relevant and trustworthy than simple case reports.

Figure 1 shows the total numbers and the years of publication (grouped in 5-year intervals) of literature references cited by at least one physician in support of at least one changed entry. References cited in support of entries that were not modified do not appear in the figure. Figure 1A shows references cited in support of changes to history findings, symptoms, and signs. Figure 1B shows references cited for the three types of laboratory findings (Lab0, Lab1, and Lab2). Figure 1C shows references cited in support of changes to disease links.

Discussion

Medical knowledge bases capture the information used by medical decision support systems.³⁴ The data and knowledge they contain can also be used for other purposes, such as data discovery, medical education³⁵ and continuing education, and satisfying health workers' information needs.³⁶⁻³⁸ Clearly, the study of how medical knowledge bases are constructed and maintained will remain a central issue in medical informatics.

The emphasis of this paper is on the activities that occur after the initial version of a medical knowledge base has been tested and deployed, the process of long-term maintenance of a knowledge base. The type and structure of the knowledge base have important implications for the way long-term maintenance occurs. Compare, for example, the area of knowledge refinement (mentioned earlier) with the knowledge base updating process described in this paper. The main differences are the philosophies underlying medical knowledge base construction, the knowledge sources used, and the scope and purpose for change. Systems that deal with very circumscribed domains may be able to rely on the knowledge of a single expert, and may require relatively limited changes over time. On the other hand, broad-scope medical knowledge bases generally use the published literature as their main knowledge source, and thus require a systematic, long-term process to track advances in the knowledge described in the literature. Such knowledge bases are often based not on rules, but on larger units of knowledge—such as descriptions of entire diseases or pathophysiologic clusters³⁹—that must be handled as a whole to en-

Figure 1 Numbers of references cited as supporting evidence for all changed entries, by year of publication and by type of entry. (A) History, symptom, and sign findings. (B) Laboratory findings. (C) Links. The year of publication is actually a 5-year period beginning with the year given (e.g., 1940 refers to references published between January 1940 and December 1944).

sure consistency. The philosophical underpinnings of a medical knowledge base are important. If the sole purpose of a knowledge base is to support the function of a medical decision support system, the most important metric of knowledge base content is system behavior. If, on the other hand, the purpose of the knowledge base is to serve as an academic repository of diagnostic information, then system performance, while clearly important, may be more secondary. The survival of a medical knowledge base over time is linked to its external anchors; if the knowledge sources are widely accessible, then the knowledge base is more likely to survive beyond its original creators and location. This is the primary motivation for the updating activity for the QMR-knowledge base described in this paper. While the diagnostic failures of the QMR expert system are handled on an ad hoc basis, as in most other expert systems, the long-term knowledge base updating process is aimed at reconciling the knowledge base with the long-term evolution of accepted clin-

ical knowledge, as embodied in the medical literature.

The results of the study show the magnitude of effort required to keep medical knowledge bases of this kind up to date. The 384 changes observed in the study affected a total of 379 entries (16% of all the entries in the updated disease profiles); in other words, one in six entries was modified by the process. Assuming that 50% of a target 800 medical disease descriptions were to be updated every 5 years, at an average of 80 entries per profile, long-term maintenance would alter more than a thousand entries per year.

Another observation from this study is that the amount of change in each disease profile is difficult to predict. As a percentage of the number of entries in the original version of each disease profile, the number of changed entries ranged from a minimum of 3.9% to a maximum of 97.5% (average 19.4%, standard deviation 25.5%). The disease profile for carpal tunnel

syndrome was practically rewritten by the two physicians who updated it: 97.5% of the initial entries were modified in some way. This was an atypical disease profile, with the original (baseline) version only about one-third as long as the average profile, and we had expected it to be revised substantially. Additionally, carpal tunnel syndrome has recently been the object of much interest in the medical community, especially in occupational medicine, resulting in a considerable amount of new publications. In any event, changes of this magnitude indicate that certain portions of any existing medical knowledge base are likely to be affected profoundly by long-term maintenance.

The two most common types of change were 1) changes in the frequency of an entry, with an average of almost 10 changes per disease profile; and 2) addition of a new entry, with an average of 9.5 additions per disease profile. Both types of change are a direct consequence of the evolution of the medical literature, as indicated by the analysis of the reasons for change. Changes in the frequency of an entry were overwhelmingly due to the appearance of new information in the peer-reviewed literature (category *new-evid*), or the lack of supporting evidence for data that had originally been based on expert opinion alone (category *no-support*). Additions of new entries were almost universally dictated by the appearance of new information in the literature.

Changes in the evoking strength of an entry were the least common type of change. This is consistent with the observation that the medical literature does not commonly report data that can be used to determine the evoking strength (analogous to the positive predictive value). Given this absence, it is not surprising that evoking strength values tended to remain fairly stable through the updating process.

The analysis of change by type of finding entry showed how the evolution of medical knowledge affects the QMR knowledge base. Entries describing laboratory tests that are costly, not commonly available, or invasive (i.e., findings of types Lab1 and Lab2) were almost twice as likely to change as the average entry. Most changes in such findings were additions, reflecting the introduction of new diagnostic procedures (such as magnetic resonance imaging and new assay techniques) that were experimental or did not exist at the time the disease profiles were last updated. As these procedures are described in the medical literature and slowly gain common acceptance, medical knowledge bases have to track their development. The study clearly showed this effect and its magnitude: additions of new Lab1 and Lab2 findings alone (63 changes) accounted for one of every six

changes. Interestingly, the Lab0 category (i.e., inexpensive laboratory tests that are routinely available in most hospitals) was the most stable of all entry types, with a percentage of changes that was almost four times below the general average.

This study marked the first time that the knowledge base updating process had been systematically delegated to physicians who were not among the QMR knowledge base developers. As such, it provided the first objective measure of the complexity of the knowledge base updating process for a large medical knowledge base when performed by non-developer physicians. Because this strategy considerably broadens the available manpower and expertise, we believe that it will become the norm, rather than the exception.

While much of the published research in knowledge acquisition deals with the initial creation of a knowledge base, the current study demonstrates that long-term maintenance (and specifically knowledge base updating) represents a considerable portion of the total life cycle of a medical knowledge base and thus deserves a prominent role in both clinical practice and the literature. Overall, the results of this study paint the picture of long-term maintenance as a complex, challenging enterprise that is likely to profoundly affect existing and future medical knowledge bases. While a comprehensive theory of long-term maintenance remains to be developed, the results indicate that this activity is crucial and just as worthy of investigation as the original creation of medical knowledge bases.

The authors thank the following physicians for volunteering to participate in this study: Wayne Chen, MD; Angelo Constantino, MD, MPH; Patrick W. Jamieson, MD; Salvatore Mangione, MD; and Vance Wright-Browne, MBBS.

References ■

1. Mars NJI, Miller PL. Tools for knowledge acquisition and verification in medicine. In: Proceedings of the Tenth Annual Symposium on Computer Applications in Medical Care. Washington, DC: IEEE Computer Society Press, 1986:36-42.
2. Musen MA. Automated Generation of Model-Based Knowledge-Acquisition Tools. London: Pitman, 1989.
3. Shortliffe EH. MYCIN: A rule-based Computer Program for Advising Physicians Regarding Antimicrobial Therapy Selection. Stanford, CA: University, 1974.
4. Davis R. Interactive transfer of expertise: acquisition of new inference rules. *Artif Intell.* 1979;12:121-57.
5. Buchanan BG, Barstow D, Bechtal R, et al. Constructing an expert system. In: Hayes-Roth F, Waterman DA, Lenat D, eds. *Building Expert Systems*. Reading MA: Addison-Wesley, 1983:127-67.
6. Garg-Janardan C, Salvendy G. A structured knowledge elici-

- tation methodology for building expert systems. In: Boose JH, Gaines BR, eds. *The Foundations of Knowledge Acquisition: Knowledge Based Systems*, vol. 4. London: Academic Press, 1988.
7. Miller PL, Blumenfrucht SR, Rose JJ, et al. Expert system knowledge acquisition for domains of medical workup: an augmented transition network model. In: *Proceedings of the Tenth Annual Symposium on Computer Applications in Medical Care*. Washington, DC: IEEE Computer Society Press, 1986:30-5.
8. Davis R, Buchanan BG, Shortliffe EH. Production rules as a representation for a knowledge-based consultation program. *Artif Intell.* 1977;8(1):15-45.
9. McDermott J. R1's formative years. *AI Magazine*. 1981;2(2):21-9.
10. McDermott J. R1 ('XCON') at age 12: lessons from an elementary school achiever. *Artif Intell.* 1993;59(1-2):241-7.
11. Guha RV, Lenat DB. CYC: a mid-term report (knowledge base). *Appl Artif Intell.* 1991;5(1):45-8.
12. Paris C, Gil Y. EXPECT: intelligent support for knowledge base refinement. In: Aussenac N, Boy G, Gaines B, et al., eds. *Proceedings of the Seventh European Workshop, EKAW '93: Knowledge Acquisition for Knowledge-Based Systems*. Toulouse and Caylus, France, 1993:220-36.
13. Lassez C, McAloon K, Port G. Stratification and knowledge base management. *J Symbolic Computation*. 1989;7(5):509-22.
14. Bacon E, McKenzie S, Smith M, Bingham J, Bramer M. GAUSS: a generalised approach to updating scanner strategies. In: Hunter J, Cookson J, Wyatt J, eds. *Proceedings of AIME 89: Second European Conference on Artificial Intelligence in Medicine*. London, UK, 1989:276-85.
15. Widmer G, Horn W, Nagele B. Automatic knowledge base refinement: learning from examples and deep knowledge in rheumatology. *Artif Intell Med.* 1993;5(3):225-43.
16. Agarwal R, Tanniru M. A structured methodology for developing production systems. *Decision Support Syst.* 1992;8(6):483-99.
17. Politakis PG. *Using Empirical Analysis to Refine Expert System Knowledge Bases*. New Brunswick, NJ: Rutgers University, Laboratory for Computer Science Research, 1982.
18. Politakis PG. *Empirical Analysis for Expert Systems*. Boston: Pitman, 1985.
19. Weiss SM, Politakis PG, Ginsberg A. Empirical analysis and refinement of expert system knowledge bases. In: *Proceedings of the Tenth Annual Symposium on Computer Applications in Medical Care*. Washington, DC: IEEE Computer Society Press, 1986:53-60.
20. Compton P, Jansen R. Diabetes expert systems: planning for long term use. *Horm Metab Res Suppl.* 1990;24:124-9.
21. Compton P, Edwards G, Kang B, et al. Ripple down rules: turning knowledge acquisition into knowledge maintenance. *Artif Intell Med.* 1992;4(6):463-75.
22. Miller RA, Pople HE, Myers JD. INTERNIST-I: an experimental computer-based diagnostic consultant for general internal medicine. *N Engl J Med.* 1982;307:468-76.
23. Clancey WJ. Epistemology of a rule-based expert system: a framework for explanation. *Artif Intell.* 1983;20:215-51.
24. Miller RA, Masarie FE. The demise of the "Greek Oracle" model for medical diagnostic systems. *Methods Inf Med.* 1990;29(1):1-2.
25. Giuse NB, Bankowitz RA, Giuse DA, Parker RC, Miller RA. Medical knowledge base acquisition: the role of the expert review process in disease profile construction. In: *Proceedings of the Thirteenth Annual Symposium on Computer Applications in Medical Care*. New York: IEEE Computer Society Press, 1989:105-9.
26. Giuse NB, Giuse DA, Miller RA. Computer assisted multi-center creation of medical knowledge bases. In: *Proceedings of the Twelfth Annual Symposium on Computer Applications in Medical Care*. Washington, DC: IEEE Computer Society Press, 1988.
27. Giuse NB, Giuse DA, Miller RA, Bankowitz RA, et al. Evaluating consensus among physicians in medical knowledge base construction. *Methods Inf Med.* 1993;32:137-45.
28. Giuse DA, Giuse NB, Bankowitz RB, Miller RA. Using the medical literature in the construction of medical knowledge bases. In: Chytil MK, Duru G, Eimeren W, Flagle CH, eds. *Proceedings of the Fifth International Conference on System Science in Health Care*. Prague, Czechoslovakia, 1992:1361-3.
29. Giuse DA, Giuse NB, Miller RA. Consistency enforcement in medical knowledge base construction. *Artif Intell Med.* 1993;5:245-52.
30. Giuse DA, Giuse NB, Bankowitz RA, Miller RA. Comparing contents of a knowledge base to traditional information sources. In: *Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care*. Washington, DC, 1993:626-30.
31. Masarie FE, Miller RA, Myers JD. INTERNIST-I properties: representing common sense and good medical practice in a computerized medical knowledge base. *Comput Biomed Res.* 1985;18:458-79.
32. Giuse DA, Giuse NB, Bankowitz RA, Miller RA. Heuristic determination of quantitative data for knowledge acquisition in medicine. *Comput Biomed Res.* 1991;24(3):261-72.
33. Wechsler LS, Saperstein MR, Stein A. Primary and symptomatic amyotrophic lateral sclerosis: a clinical study of 81 cases. *Am J Med Sci.* 1944;208:70-81.
34. Miller RA. Medical diagnostic decision support systems: past, present, and future: A threaded bibliography and brief commentary. *JAMIA.* 1994;1:8-27.
35. Clancey W. Tutoring rules for guiding a case method dialogue. *Int J Man-Machine Studies.* 1979;11:25-49.
36. Giuse NB, Huber JT, Giuse DA, Brown CWJ Jr, Bankowitz RA, Hunt S. Information needs of health care professionals in an AIDS outpatient clinic as determined by chart review. *JAMIA.* 1994;1:395-403.
37. Loftus J, Kowaloff H, Safran C, Cotton D, Crumpacker C. HIV-ProtoCall: use of a hospital clinical computer for AIDS clinical trials protocol information distribution. In: *Proceedings of the International Conference on AIDS*. 1990:283.
38. Safran C, Rind DM, Herrmann F, et al. The development of knowledge-based medical records for clinicians caring for patients with HIV infection. In: *Proceedings of the Seventh World Congress on Medical Informatics—MEDINFO '92*. Geneva, Switzerland, 1992:495-500.
39. Puppe B, Puppe F. A knowledge representation concept facilitating construction and maintenance of large knowledge bases. *Methods Inf Med.* 1988;27(1):10-6.