Research Paper ■

The Granularity of Medical Narratives and Its Effect on the Speed and Completeness of Information Retrieval

HUIBERT J. TANGE, MD, PHD, HARRY C. SCHOUTEN, MD, PHD, ARNOLD D. M. KESTER, PHD, ARIE HASMAN, PHD

Abstract **Objective:** Using electronic rather than paper-based record systems improves clinicians' information retrieval from patient narratives. However, few studies address how data should be organized for this purpose. Information retrieval from clinical narratives containing free text involves two steps: searching for a labeled segment and reading its content. The authors hypothesized that physicians can retrieve information better when clinical narratives are divided into many small, labeled segments ("high granularity").

Design: The study tested the ability of 24 internists and 12 residents at a teaching hospital to retrieve information from an electronic medical record—in terms of speed and completeness—when using different granularities of clinical narratives. Participants solved, without time pressure, predefined problems concerning three voluminous, inpatient case records. To mitigate confounding factors, participants were randomly allocated to a sequence that was balanced by patient case and learning effect.

Results: Compared with retrieval from undivided notes, information retrieval from problem-partitioned notes was 22 percent faster (statistically significant), whereas retrieval from notes divided into organ systems was only 11 percent faster (not statistically significant). Subdividing segments beyond organ systems was 13 percent slower (statistically significant) than not subdividing. Granularity of medical narratives affected the speed but not the completeness of information retrieval.

Conclusion: Dividing voluminous free-text clinical narratives into labeled segments makes patient-related information retrieval easier. However, too much subdivision slows retrieval. Study results suggest that a coarser granularity is required for optimal information retrieval than for structured data entry. Validation of these conclusions in real-life clinical practice is recommended.

■ **J Am Med Inform Assoc.** 1998;5:571–582.

The primary function of the medical record is to supply information for direct clinical practice. The medical record originally served as a memory-aid for the individual practitioner but has evolved to become a communication tool used by different professionals sharing responsibility for an individual patient. Computerization of the medical record is considered inevitable, but several impediments exist. The inability of existing electronic medical record (EMR) systems to incorporate physician-gathered clinical narratives satisfactorily is seen as one of the major impediments.[1] "Clinical narratives" are defined as clinical textual data that are traditionally recorded as natural prose. This article deals with the following types of clinical narratives: previous history, medical history (i.e., history of present illness and review of systems), physical examination, and progress notes. In classic medical record systems this information is usually stored as free text.[2–5]

Affiliation of the authors: Maastricht University, Maastricht, The Netherlands.

Correspondence and reprints: Huibert J. Tange, MD, PhD, Department of Medical Informatics, Maastricht University, P.O. Box 616, NL-6200 MD Maastricht, The Netherlands; e-mail: ⟨huibert.tange@mi.unimaas.nl⟩.

Received for publication: 6/19/97; accepted for publication: 7/13/98.

Previous EMR research focused predominantly on problems of entering and storing clinical narratives. Few published studies concern how clinical narratives should be organized within an EMR to facilitate the retrieval of information. Information retrieval from free-text clinical narratives is a two-step process. First, the user searches through the clinical record to find segments that might contain relevant information, employing segment labels and note entry times as searching keys. The medical record can support this step by providing a searching structure in which clinical data are presented in a two-dimensional (label × time) matrix. Such a searching structure is known as a flow sheet.[5,6] Second, the user reads the content of pertinent segments to retrieve the information needed. This content can be textual (e.g., clinical narratives), numeric (e.g., laboratory results, temperature measurements), or otherwise (e.g., images).

Typical of textual data is the variable level of detail in which the text can be divided into segments. Clinical narratives containing a large number of segments have a "high granularity." The more detailed these segments, the more specifically one can search by their segment label and the less one has to read. On the other hand, the less detailed these segments, the less specifically one can search, but the more informative is their readable content. Such clinical narratives have a "coarse granularity." The granularity of clinical narratives determines the mixture of searching and reading and as a consequence may influence the clinicians' ability to retrieve information.

Is there an optimal balance between searching effort and reading effort when retrieving information from an EMR? The literature indicates that, among existing EMR systems, different granularities are used.[7] Since the authors could not find comparative studies investigating the granularity of clinical narratives, it was decided to investigate the influence of granularity on clinicians' information retrieval. The authors hypothesized that physicians would retrieve information faster and more completely when using clinical narratives divided into more numerous, smaller segments.

Method, Materials, and Test Design

The study evaluated physicians' ability to retrieve information from an EMR—in terms of speed and completeness—when using different granularities of clinical narratives. To control other factors that might influence the retrieval of information in clinical practice, a laboratory study was conducted. A number of physicians were asked to solve a set of predefined clinical problems concerning three preselected, voluminous, inpatient cases. The authors conducted this study in the domain of internal medicine.

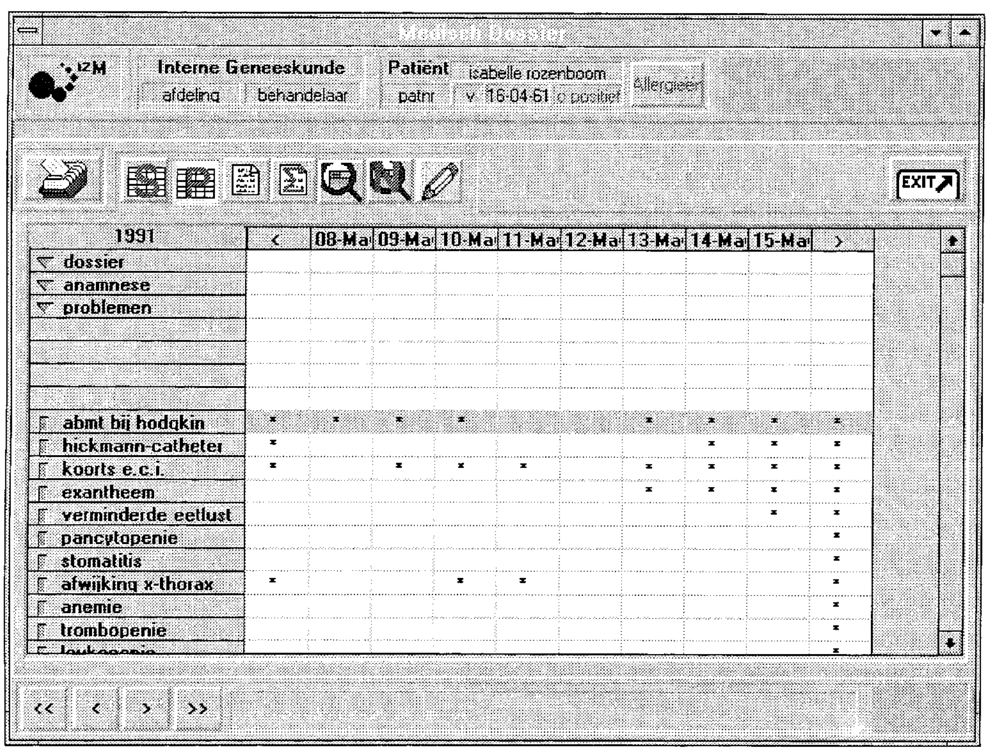

Figure 1 Flow sheet presentation showing problem list with corresponding progress notes. In this example, the problem "abmt bij Hodgkin" (autologous bone-marrow transplantation concerning Hodgkin's disease) has been selected over the entire range of time.

Table 1 ■

Default Organization of Segment Labels*

Coarse Granularity Set	Intermediate Granularity Set	Fine Granularity Set
Background data	Background data	Background data
Previous history	*Previous history*	*Previous history*
Medical history	Medical history *General well-being* *Circulatory system* *Respiratory system* . . .	Medical history . . . Circulatory system *Palpitations* *Chest pain* . . .
Physical examination	Physical examination *General impression* *Head and neck* *Heart* *Lungs* . . .	Physical examination . . . Lungs *Percussion* *Auscultation* . . .
Progress notes	Progress notes *Problem 1* *Problem 2* . . .	Progress notes *Problem 1* *Problem 2* . . .
Tests	Tests	Tests
Treatment	Treatment	Treatment

*The three granularity sets differ only with respect to medical history, physical examination, and progress notes. Labels in italics are not further subdivided and may contain patient data. Subdivisions of the labels "Background data," "Tests," and "Treatments" are not shown.

Granularity Sets and Medical Record System

The study compared three granularities of medical history and physical examination (undivided, divided into organ systems, and divided into separate questions or observations) and two granularities of progress notes (undivided and problem-partitioned). The authors defined three "granularity sets" combining the most encountered granularities of medical history, physical examination, and progress notes, as found in the literature. Previous-history notes were equal in all three granularity sets. These sets were as follows:

- **Coarse**. This structure was an exact copy of the paper clinical record as it is used at the Department of Internal Medicine of Maastricht University Hospital. Medical history, physical examination, and progress notes were all undivided. Medical history and physical examination data gathered after intake were recorded as part of the progress notes.
- **Intermediate**. Medical history and physical examination records were divided into organ systems, and progress notes were divided into problems. Medical history and physical examination data gathered after intake were recorded after the corresponding data from the intake.
- **Fine.** Medical history and physical examination records were further divided into separate questions and observations, whereas progress notes remained divided into problems. A further SOAP-like division (i.e., subjective–objective–assessment–plan format) of progress notes was not used. Medical history and physical examination data gathered after intake were recorded after the corresponding data from the intake.

The study presented the clinical record as a flow sheet (Figure 1) in which clinical data were directly accessible by two indexes: segment label (vertical axis) and registration time (horizontal axis). This flow sheet was used for searching. Clicking a cell (or an area of cells) in the flow sheet would open a text window in which the content of these clinical data could be read. An asterisk denoted the presence of data behind a cell. The segment labels were organized in two ways. By default, they were hierarchically organized according to source, organ system, and type of observation (Table 1). The top level of the hierarchy consisted of the

Table 2 ■

Alternative Organization of Segment Labels Pertinent to the Problem "Liver Cirrhosis"

Coarse Granularity Set	Intermediate Granularity Set	Fine Granularity Set
Background data	Background data	Background data
Previous history	*Previous history*	*Previous history*
Medical history	Medical history	Medical history
	General well-being	General well-being
	Digestive system	*Tired?*
	Urogenital system	Digestive system
	Dermatologic system	*Color stool?*
	Neurologic system	Urogenital system
		Color urine?
		Dermatologic system
		Itching?
		Neurologic system
		Dizziness?
Physical examination	Physical examination	Physical examination
	General impression	General impression
	Abdomen	*Ill-looking?*
	Skin	*Jaundice?*
		Abdomen
		Percussion liver
		Percussion spleen
		Percussion elsewhere
		Palpation liver
		Palpation spleen
		Skin
		Scrataching marks?
Progress notes	Progress notes	Progress notes
	Liver cirrhosis	*Liver cirrhosis*
Tests	Tests	Tests
Treatment	Treatment	Treatment

NOTE: The three granularity sets differ only with respect to medical history, physical examination, and progress notes. Labels in italics are not further subdivided and may contain patient data. Subdivisions of the labels "Background data," "Tests," and "Treatments" are not shown.

following "sources": background data, previous history, medical history, physical examination, progress notes, tests, and treatment. The user could enter a subdivision of a source by clicking the segment label. Clicking "physical examination," for example, would result in a list of regions and body parts, and clicking the label of one of these would result in a list of separate observations (as far as the granularity permitted). Alternatively, the user could switch to an organization in which segment labels of any kind (source, organ system, etc.) were listed according to their relevance to clinical problems (Table 2). For the study, the authors used an electronic medical record system developed in Visual Basic 3.0 in the Microsoft Windows 3.11 environment. Segment labels as well as their organizational structures were specified in a Btrieve 5.0 database in which the patient data also were stored. The authors used Btrieve because it is very fast and flexible. Clinical narratives were stored as free text. Advanced text-processing functions were not available to facilitate searching. Additional details about the system have been provided elsewhere.[8]

Patient Cases and Problems Addressed

The authors conducted the study in the domain of internal medicine, because it is the "mother" of many other specialties and harbors a substantial collection of voluminous patient cases. The authors filled the study EMR system with three large and complex patient cases. These cases were originally documented in paper medical records and transcribed into the computer afterwards. Patient X was a 30-year-old woman who had Hodgkin's disease for many years. She was hospitalized to undergo autologous bone marrow transplantation. Patient Y was a 47-year-old man with cirrhosis of the liver due to alcohol abuse. He was hospitalized because of suspected venous thrombosis, which could not be verified. During hospitalization, he developed end-stage renal failure. Patient Z was a 53-year-old woman who was hospitalized for suspected leukemia. During hospitalization, the diagnosis "acute myeloid leukemia" was established and the woman received chemotherapy for the first time.

The authors asked the clinician supervising each case to select for the case some instances where medical practices were (re)considered or reconciled: when on call, on rounds, or on consultation; when composing a discharge summary; or for clinical-audit purposes. Table 3 shows which problems were addressed. For each problem, the clinician formulated one or more specific questions that were or could have been raised about the case (see Appendix). A total of 19 questions were formulated: 7 about case X, 7 about case Y, and 5 about case Z. For each question, the clinician supervising the case had formulated a gold-standard list of essential data elements that should be retrieved.

The questions could be divided into three categories, according to the clinical narrative part involved (see Appendix, last column). For answering the three questions of Category I, only information from the previous history was needed. This section was equally structured in each granularity set, so the study employed these questions as a control. For answering the 11 questions of Category II, information from progress notes was needed. This section was equally structured in the intermediate and fine sets but was different in the coarse set. For answering the five questions of Category III, information from medical history or physical examination was needed. These sections were differently structured in each granularity set.

Table 4 ■

Balanced Incomplete Block Design*

Physician	Series 1	Series 2	Series 3
1	AX	BY	CZ
2	AX	CY	BZ
3	BX	AY	CZ
4	BX	CY	AZ
5	CX	AY	BZ
6	CX	BY	AZ
7	AX	BZ	CY
8	AX	CZ	BY
9	BX	AZ	CY
etc.	...	...	...

*Notice that in one block of six physicians the granularity sets (A, B, and C) are balanced over the three series. In six blocks, the patient cases (X, Y, and Z) are also balanced. The number of physicians needed is 6 × 6 = 36. For granularity sets, A indicates coarse; B, intermediate; C, fine. In the body of the table, AX indicates the combination of granularity set A and patient case X. The series numbers indicate the presentation order in which a combination of granularity set and patient case is offered.

Outcome Measures and Cross-over Design

The authors were interested in the effect of the granularity set used on the speed of information retrieval and the completeness of information retrieved by a physician. For each physician participating in the study, the authors recorded, for each question, the time spent on searching and the proportion of essential data elements retrieved. For each physician, the study used as outcome measures the averages of the measurements for the questions across each granularity set.

The study used an outcome model with four main effects that could influence the outcome measures: granularity set, physician, patient case, and learning effect. The authors also included three effect modifiers (factors that could interact with the granularity effect) in the model: patient case, learning effect, and clinical expertise, since experienced physicians tend to use other thinking strategies than novices.[9] To balance the main effects, the authors used a cross-over design for three treatments.[10,11] A completely balanced design—in which each physician would be given all possible combinations of granularity set and patient case—was not acceptable, because the physicians would have to be confronted with the same patient and the same questions three times. Hence, the study used an allocation pattern known as balanced incomplete

Table 3 ■

Problems Addressed in the Questions Concerning Cases X, Y, and Z

Problem	Case X	Case Y	Case Z
Fever, cause unknown	Yes	Yes	–
Protocol compliance	Yes	–	Yes
Stomatitis	Yes	–	–
Liver cirrhosis	–	Yes	–
Renal failure	–	Yes	–
Fluid retention	–	–	Yes
Thrombopenia	–	–	Yes
Leg ulcer	–	–	Yes
Heart failure	–	–	Yes

Table 5 ■

ANOVA Test of Significance for Sources of Variance (Main Effects) and Effect Modifiers (Interactions) Regarding Speed of Information Retrieval

	Speed				
Sources of Variance	SS	DF	MS	*F*	*P*
Main effects					
Physician	0.42	35	0.01	1.81	0.023
Presentation order	0.12	2	0.06	9.33	0.000*
Patient case	0.66	2	0.33	49.73	0.000*
Granularity set	0.12	2	0.06	8.73	0.000*
Interactions					
Patient case × granularity set	0.03	4	0.01	1.29	0.318
Presentation order × granularity set	0.03	4	0.01	1.04	0.394
Expertise × granularity set	0.03	2	0.01	2.16	0.125
Individual error	0.37	56	0.01	–	–

NOTE: SS indicates sums of squares; DF, degrees of freedom; MS, mean squares; *F*, *F* ratio; *P*, significance of *F*.
*Significant at $P < 0.05$.

block design, in which each physician was given three combinations with alternating granularity sets and patient cases (Table 4). A balanced set of outcomes would be reached with 36 test persons.

Test Population and Physician-sampling Procedure

The study was restricted to an academic setting. This setting encompasses a broad spectrum of clinical experiences and pays relatively greater attention to clinical documentation. The authors recruited the test physicians from the Department of Internal Medicine of Maastricht University Hospital. Candidates were physicians with "sufficient" experience in internal medicine, including all specialists (35 candidates) and those residents who were qualified as such by one of their supervisors (18 candidates). The authors stratified the set of test persons needed into 24 specialists (four blocks) and 12 residents (two blocks). The candidates were selected from a randomized list and contacted by telephone. The recruiting procedure stopped as soon as the sample size imposed by the study design was reached. At the beginning of the test, each participant was randomly allocated to one of the sequences in the block design.

Test Configuration and Data-sampling Procedure

To support the test design, the authors extended the EMR system with some specific functions: a procedure for randomly allocating a physician to one of the sequences in the block design, a mouse-skill test, an automated protocol for guiding the physician through the sequence of questions and patient cases, and a procedure with which all computer actions of the physician could be recorded for further analysis. The system was installed on a PC workstation in a separate room. To avoid disturbances, the physicians had to switch off their pagers. Each physician performed the test individually and was supervised by an investigator (first author).

At the beginning of a session, each physician answered some questions about clinical experience and computer skills and performed a short mouse-skill exercise. During the actual test period, the protocol automatically guided the physician from one question to the next and from one patient to the other. Each question was introduced by a brief description of the reason for consultation. Clinical information recorded after the period covered by the question was not presented. Measurements started when the physician switched to the clinical record to find the information needed. From that time, all user-driven actions were logged and time-stamped. If the physician found information essential as evidence, he or she had to highlight this information with the mouse. All computer actions, highlighted information, and possible additional comments were recorded in a Microsoft Access 1.1 research database for further analysis.

The test configuration was validated on six randomly selected specialists. The measurements obtained with these six specialists were excluded from further analysis.

Statistical Analysis

The measurement data were analyzed in SPSS 5.1 by conducting a many-way analysis of variance

Table 6 ■

ANOVA Test of Significance for Sources of Variance (Main Effects) and Effect Modifiers (Interactions) Regarding Completeness of Information Retrieved

	Completeness				
Sources of Variance	SS	DF	MS	*F*	*P*
Main effects					
Physician	0.38	35	0.01	1.52	0.079
Presentation order	0.03	2	0.01	1.94	0.000*
Patient case	0.61	2	0.31	42.79	0.000*
Granularity set	0.00	2	0.00	0.19	0.828
Interactions					
Patient case × granularity set	0.01	4	0.00	0.36	0.835
Presentation order × granularity set	0.06	4	0.01	2.00	0.107
Expertise × granularity set	0.01	2	0.00	0.59	0.574
Individual error	0.40	56	0.01	–	–

NOTE: SS indicates sums of squares; DF, degrees of freedom; MS, mean squares; *F*, *F* ratio; *P*, significance of *F*.
*Significant at $P < 0.05$.

(ANOVA). Search speed was analyzed using a log-transformed scale. Significant sources of variance ($P < 0.05$) were further analyzed by means of a pairwise simple contrast analysis to determine which individual values were responsible for this variance,[12] using Bonferroni's correction for multiple comparisons.[13] If, for example, the granularity set was shown to have a significant effect on search speed, the differences between the individual granularity sets (coarse–intermediate, coarse–fine, intermediate–fine) would be tested.

Results

Overall, 80 percent of the specialists approached and 73 percent of the residents approached agreed to participate in the study. Noncooperation was, in all cases, due to lack of time. The average duration of an entire session, including explanation and preparatory exercise, was 50 min (range, 38–68 min), of which 44 min (range, 32–63 min) were spent on the actual test. On average, residents performed the test in 42 min, specialists in 45 min. All questions were answered satisfactorily, according to the physicians themselves.

Granularity

Analysis of variance demonstrated that the granularity set used was a significant source of variance for the speed of information retrieval ($P < 0.05$, Table 5), but not for the completeness of information retrieved ($P > 0.05$, Table 6). Table 7 shows that the physicians could uncover approximately 80 percent of the essential data elements in each granularity set. Table 7 also demonstrates that information retrieval with the coarse granularity set (152 sec) was significantly slower than with the intermediate and fine granularity sets ($P < 0.017$). Information retrieval with the intermediate set (124 sec) and fine granularity set (131 sec) did not differ significantly.

To determine whether these differences were explained by the structure of the progress notes, the structure of the other narrative parts, or both, the authors conducted another analysis of variance, including also the question category (see Appendix) as an effect modifier of the granularity set. Its modifying effect was indeed significant ($P < 0.05$), which means that the kind of notes influences which granularity set enables the fastest searching. Table 8 shows that for previous-history notes (Category I) granularity does not make a difference. This was expected, because these notes were equally structured in all three granularity sets. Retrieving information from progress notes (Category II) when using a problem-partitioned structure was significantly faster than when using an undivided structure ($P < 0.017$). Information retrieval from medical history and physical examination, when divided into organ systems, was significantly faster than when further divided into separate questions and observations ($P < 0.017$). The physicians scored between when using undivided medical history and physical examination notes.

Other Sources of Variance

Strong sources of variance for both speed and completeness were patient case and order of presentation

Table 7 ▪

Completeness and Speed of Information Retrieval*

	Completeness		Speed			
	Average %	SE	Average sec	SE	Difference	*P* Value
Granularity set						
A (coarse)	81	2.0	152	9.0	A–B	0.000†
B (intermediate)	81	2.0	124	8.0	A–C	0.016†
C (fine)	79	2.0	131	5.8	B–C	0.102
Patient case						
X	87	1.5	122	5.3	X–Y	0.326
Y	85	1.3	117	4.5	X–Z	0.000†
Z	70	1.8	174	8.8	Y–Z	0.000†
Presentation order						
1	81	2.0	149	7.8	1–2	0.020
2	78	2.0	134	8.2	1–3	0.000†
3	82	2.0	123	7.5	2–3	0.060

*Completeness (proportion of essential data elements retrieved, in percentage) and speed (duration of information retrieval, in seconds) were averaged over the 36 physicians as a function of granularity set, patient case, and presentation order. For speed the geometric average was used. SE indicates standard error of the mean.
†Significant at $P < 0.017$.

(Tables 5 and 6; $P < 0.05$). Especially for Patient Z, the physicians needed more time (174 sec) and reached less complete answers (70 percent) than for the other two patients (Table 7; $P < 0.017$ for speed). A presentation-order effect (which is an indication of a learning effect) could be demonstrated only for speed: the last presented granularity set took less time than the first (Table 7; $P < 0.017$). The averages for completeness did not show any indication of a learning effect. The physicians differed significantly in speed (Table 5; $P < 0.05$) but not in completeness (Table 6). Clinical expertise did not influence the effect of the granularity set used, either for speed or for completeness (Tables 5 and 6).

Discussion

The risk of investigating more than one outcome measure is that the results may point in different directions. The authors found that the granularity of clinical narratives affects the speed of information retrieval but not the completeness of information retrieved. However, the two measures are not independent. In the simulated setting of the authors' laboratory, the physicians were allowed to complete all questions, however long it took. In daily practice, physicians often work under time pressure. That such a situation can lead to less complete answers (when searching takes too much time) was illustrated by Tang et al.,[14] who studied the completeness of patient information retrieved by physicians during 168 return outpatient visits. In 81 percent of these visits, during which the physicians could use both the hospital information system and a paper shadow record, the physicians stated that they could not find all the relevant patient information. Of the missing data, 31 percent were from clinical narratives. So the authors of the current study have reason to believe that in real-life practice, a study similar to this one would lead to less-complete answers. In that case, it is likely that completeness would decrease most when the granularity that takes the most time is used.

The authors are not aware of other comparative timing studies concerning the effect of different granularities in EMR systems. A comparable study using paper-based clinical records was performed by Fries in 1974.[15] He compared the speed of information retrieval among 26 physicians when using time-oriented, traditional (source-oriented), and problem-oriented clinical records. In the time-oriented clinical record, Fries used a flow sheet presentation of data, with a fine-granularity searching structure for medical history and physical examination but with undivided progress notes. The physicians in his study consulted the time-oriented clinical record four times faster than the other record structures. A comparison of present study findings with Fries' results cannot be completed, however, because the studies differ in several respects. The most important difference (besides the paper medium) is the type of search questions. Fries used five "common clinical questions"—two about blood pressure and one each about body weight, medication supply, and a laboratory test result. These questions concerned quantitative information that can be found in sections of the clinical record that were structured identically among the granularity sets used in the current study.

Progress Notes

In the late 1960s, Lawrence Weed—promoting the problem-oriented medical record (POMR)—maintained that a division of progress notes into problem-related segments would help physicians to focus on one problem at a time, without neglecting the overview.[16] Nowadays, the progress notes of most EMR systems are divided into problem sections.[1] The findings of the current study suggest that problem-partitioning makes information retrieval easier, indeed. Study physicians retrieved information from progress notes faster when using a problem-partitioned structure as compared with an undivided structure. The time gained—using multiple-pathology cases—was 22 percent, a level the authors believe is worthwhile.

However, one swallow does not make a summer. Problem partitioning is not the only characteristic of the POMR. The authors did not employ the SOAP structure* in the present study, a concept that has been much criticized.[17–19] They also did not investigate the consequences of problem partitioning for the ease of data entry, which most consider to be problematic. The data entry problem might explain why some developers seem to have lost confidence in the problem-oriented approach and have returned to undivided progress notes.[20] Nevertheless, having demonstrated the superior information retrieval capability of problem-partitioned progress notes, the authors believe there is a serious reason for spending more research effort on overcoming data entry problems of the POMR.

Medical History and Physical Examination

The authors found that retrieving information from medical history and physical examination notes divided into finely detailed, separate questions and observations was slower than when segments partitioned at the level of organ systems were used. These findings suggest that too-detailed partitioning has a negative effect on the speed of searching. Hence, the authors have to reject the initial study hypothesis that physicians retrieve information faster when using finer-grained clinical narratives. The authors have two possible explanations for this unexpected result. First, the time needed to search among a high number of detailed segment labels may outweigh the time needed to read the whole story of a coarse segment. Second, the content of a single detailed segment may become so restricted that it no longer covers the clinical context needed for a proper understanding.

*We do not consider the SOAP structure an issue of granularity, since it will dissolve automatically when using subdivided medical history and physical examination sections. In that case, the subjective and objective observations will be distributed among the pertinent parts, leaving only assessment and plan for the progress notes.

Table 8 ■

Effect of Question Categories on Speed of Information Retrieval*

Granularity Set	Speed: Average sec	SE	Difference	P Value
Category I (3 questions)				
A (coarse)	90	11.0	A–B	0.380
B (intermediate)	77	9.0	A–C	0.067
C (fine)	76	8.4	B–C	0.339
Category II (11 questions)				
A (coarse)	127	14.2	A–B	0.000†
B (intermediate)	101	10.5	A–C	0.000†
C (fine)	104	10.5	B–C	0.893
Category III (5 questions)				
A (coarse)	169	25.7	A–B	0.064
B (intermediate)	151	21.4	A–C	0.283
C (fine)	174	19.8	B–C	0.004†

*Geometric average is given in seconds. SE indicates standard error of the mean. Category I indicates previous history (control); Category II, progress notes; Category III, medical history or physical examination.
†Significant at $P < 0.017$.

Among existing EMR systems, three different granularities of medical history and physical examination are in use: 1) without partitioning,[21–23] 2) with segments on the level of organ systems,[20,24] and 3) partitioned into separate observations or even single findings.[23,25,26] None of these EMR systems, however, has been evaluated specifically on its information retrieval capabilities. As far as the authors know, the current study is the first to reveal empirical findings that support granularity at the intermediate level. Nygren et al.[24,27] chose this granularity on the basis of observations of the way physicians from different specialties read the medical record. Poon et al.[20,28] preferred this granularity for presenting clinical narratives despite the fact that in their system these narratives were entered and stored as separate, structured data.

This study suggests a discrepancy between the optimal granularity for clinical narratives for information retrieval and the granularity that is utilized in studies facilitating structured data entry. "Dynamic data entry," for example, implies the entry of clinical narratives as separate observations.[20,29–33] To address both needs, clinical data stored as isolated facts must be aggregated into a segmented, free-text presentation format. The authors do not expect that this poses a serious problem, given the long history of applications that are capable of natural-language generation.[4,28,31,33–35] The opposite approach—parsing free text into its factual components—would be much more difficult, because natural-language interpretation, although promising, is still an immature development.[36–38]

Generalizability

The way physicians consult the medical record depends on so many factors that it is difficult to isolate one of them. The laboratory setting the authors chose for the present study made it possible to identify and control effects other than granularity on the searching behavior of participating physicians. However, findings from a laboratory setting have a limited generalizability to daily practice. The absence of time pressure has already been mentioned, but also other factors were better controlled than in a real-life setting.

Patient case, and more specifically the question mixture, was a larger source of variance than granularity. The questions about Patient Z required on average more essential data elements than the questions about the other patients. Would a different sample of cases or questions have changed study conclusions? The authors tried to enhance the coverage of the study question mixture by instructing the clinician-author of the questions to choose situations that were representative of the cases and to keep in mind the four strategies of reading a medical record, as described by Nygren and Henrikson[27]: 1) to get an overview of an unknown patient case, 2) to refresh one's memory about a known patient case, 3) to search for specific data, and 4) to solve clinical problems (i.e., to generate and test hypotheses). More objective measures of the technical complexity of searching the medical record were not available. Nevertheless, the authors believe that the question mixture the clinicians generated was satisfactory, because no interaction with the granularity effect was demonstrated.

Would a similar study outside the domain of internal medicine, or in a nonacademic setting, lead to other results? The authors have reasons not to believe this. First, the present study did not show a difference between specialists and residents. Second, observations of the way physicians from different specialties read the medical record indicate that retrieving information from the medical record is a general skill that is an integrated part of clinical expertise.[27] Hence, the authors assume that if a similar study would lead to different results, it would be because of factors other than the clinical discipline or clinical setting.

A factor that truly limits the generalizability of present study findings is the limitation of the study to use of only voluminous cases. When information is retrieved from small cases ("thin" medical records), it could be expected that the searching of undivided clinical narratives would be less problematic. Another limiting factor was the absence of the physician's familiarity with the cases at hand, as a result of the presentation only of patient cases that were not previously known to any participant. In routine clinical practice, a part of the information in the medical record is memorized by the physician, thus reducing the need for consulting the record. The medical record may also serve as a trigger for the physician's memory: When finding some information, he or she will also remember other information again.[27] The larger a piece of text, the greater the diversity of formulations that can be expected. This will raise the chance of memory triggers. Hence, if memorization has an effect on the preferred granularity, the authors assume it may shift from a finer to a coarser structure rather than the opposite way.

Also, the EMR system the authors used had some limitations that restrict the validity of study findings. The software development kit that was used (Visual Basic 3.0) did not support the latest features for free-text searching. Hence, it was not possible to search by keywords within large text bodies. Free-text searching by keywords (by the use of search engines) has become popular with the growing use of World Wide Web browsers. For searching in clinical narratives with a coarse granularity, such a feature would certainly have an added value. When available in an EMR, this feature may also lead to a preference shift from a finer to a coarser structure. On the other hand, the added value of search engines should not be overestimated. When reading clinical narratives, the clinical context of clinical data also has a guiding role. Search engines do not take this context into account, which may lead to an erratic way of searching.

Summarizing the factors that limit the generalizability of the present study, the authors conclude that the optimal granularity for clinical narratives may in some situations be coarser than the study revealed, but never finer.

Conclusion

Electronic medical record systems may facilitate the retrieval of information to support direct patient care, especially when large and complex cases are concerned. The present study reveals that increasing the granularity of clinical narratives improves performance, but increasing granularity too much impairs performance. Most benefit can be expected from medical history and physical examination notes divided into organ systems and progress notes divided into problem segments. This conclusion has been reached through a laboratory study of large and complex patient cases. In some situations (e.g., with "thin" medical records), the use of undivided clinical narratives may provide an alternative, but retrieving information from more finely segmented, separated questions and observations will never be preferable. Since evidence suggests that highly structured data entry can improve the quality of stored data, and since this involves a higher level of granularity than is required for optimal retrieval, the application of natural-language generating techniques will be needed to collate and retrieve record contents stored as individual items for data quality purposes.

The best way to validate these conclusions under a variety of circumstances—such as searching under time pressure, searching in "thinner" medical records, searching information that may partly be memorized, and searching with the availability of full-text search engines—is to perform a comparative study in daily clinical practice. For such a study, the authors need physicians who routinely use an EMR system that includes clinical narratives. At present, such systems are not implemented on a large scale, at least in The Netherlands.

References ■

1. Dick RS, Stein EB. The computer-based patient record: an essential technology for health care. Washington, D.C.: National Academy Press, 1991.
2. McDonald CJ, Blevins L, Tierney WM, Martin DK. The Regenstrief medical records. MD Comput. 1988;5:34–47.
3. Whiting O'Keefe QE, Whiting A, Henke J. The STOR clinical information system. MD Comput. 1988;5:8–21.
4. Stead WW, Hammond WE. Computer-based medical records: the centerpiece of TMR. MD Comput. 1988;5:48–62.
5. Barnett GO. The application of computer-based medical-record systems in ambulatory practice. N Engl J Med. 1984; 310:1643–50.
6. McDonald CJ, Barnett GO. Medical-record systems. In: Shortliffe EH, Perreault LE (eds). Medical Informatics: Computer Applications in Health Care. Reading, Mass.: Addison-Wesley, 1990;181–218.
7. Tange HJ, Hasman A, de Vries Robbé PF, Schouten HC. Medical narratives in electronic medical records: a literature review. Int J Med Inform. 1997;46:7–29.
8. Tange HJ, Dreessen VAB, Hasman A, Donkers HHLM. An experimental electronic medical-record system with multiple views on medical narratives. Comput Methods Programs Biomed. 1997;50:157–72.
9. Custers EJFM, Boshuizen HPA, Schmidt HG. The influence of medical expertise, case typicality, and illness script component on case processing and disease probability estimates. Memory Cognit. 1996;24:384–99.
10. Senn S. Cross-over trials in clinical research. Chicester, England: John Wiley & Sons, 1993.
11. Jones B, Kenward MG. Design and analysis of cross-over trials. London, England: Chapman and Hall, 1989.
12. Norusis MJ. SPSS Advanced Statistics 6.1. Chicago, Ill.: SPSS, 1994.
13. Altman DG. Practical statistics for medical research. London, England: Chapman and Hall, 1991.
14. Tang PC, Fafchamps D, Shortliffe H. Traditional medical records as a source of clinical data in the outpatient setting. Proc 18th Annu Symp Comput Appl Med Care. 1994:575–9.
15. Fries JF. Alternatives in medical record formats. Med Care. 1974;12:871–81.
16. Weed LL. Medical records that guide and teach. N Engl J Med. 1968;278:593–600, 652–7.
17. Goldfinger SE. The problem-oriented record: a critique from a believer. N Engl J Med. 1973;288:606–8.
18. Donnelly WJ, Hines E, Brauner DJ. Why SOAP is bad for the medical record. Arch Intern Med. 1992;152:481–4.
19. Feinstein AR. The problems of the "Problem-Oriented Medical Record." Ann Intern Med. 1973;78:751–62.
20. Poon AD, Fagan LM. PEN-Ivory: the design and evaluation of a pen-based computer system for structured data entry. Proc 18th Annu Symp Comput Appl Med Care. 1994:447–51.
21. Stead WW, Hammond WE, Straube MJ. A chartless record —is it adequate? J Med Syst. 1983;7:103–9.
22. Safran C, Porter D, Lightfoot J, et al. ClinQuery: a system for online searching of data in a teaching hospital. Ann Intern Med. 1989;111:751–6.
23. van Ginneken AM, Stam H. Can one patient record accommodate the diversity of specialized care? Proc 19th Annu Symp Comput Appl Med Care. 1995:406–10.
24. Nygren E, Johnson M, Henriksson P. Reading the medical record, Part II: Design of a human-computer interface for basic reading of computerized medical records. Comput Methods Programs Biomed. 1992;39:13–25.
25. Bolens M, Borst F, Scherrer JR. Organizing the clinical data in the medical record. MD Comput. 1992;9:149–55.
26. Heathfield HA, Hardiker N, Kirby J, Tallis R, Gonsalkarale M. The PEN & PAD medical record model: development of a nursing record for hospital-based care of the elderly. Methods Inf Med. 1994;33:464–72.
27. Nygren E, Henriksson P. Reading the medical record, Part I: Analysis of physicians' ways of reading the medical record. Comput Methods Programs Biomed. 1992;39:1–12.
28. Poon AD, Fagan LM, Shortliffe EH. The PEN-Ivory project: exploring user-interface design for the selection of items from large controlled vocabularies in medicine. J Am Med Inform Assoc. 1996;3:168–83.
29. Heathfield HA, Kirby J, Hardiker NR. Data entry in computer-based care planning. Comput Methods Programs Biomed. 1995;48:103–7.
30. Huff SM, Pryor TA, Tebbs RD. Pick from thousands: a collaborative processing model for coded data entry. Proc 16th Annu Symp Comput Appl Med Care. 1992:104–8.
31. Gouveia Oliveira A, Raposo VD, Azevedo AP, et al. SI-SCOPE: a multiuser information system for gastrointestinal endoscopy. Endoscopy. 1991;23:272–7.
32. Moorman PW, van Ginneken AM, van der Lei J, van Bemmel JH. A model for structured data entry based on explicit

discriptional knowledge. Methods Inf Med. 1994;33:454–63.
33. Trace D, Naeymi Rad F, Haines D, et al. Intelligent Medical Record-Entry (IMR-E). J Med Syst. 1993;17:139–51.
34. McDonald CJ, Tierney WM, Blevins L. The benefits of automated medical record systems for ambulatory care. In: Orthner HF, Blum BI (eds). Implementing Health Care Information Systems. New York: Springer-Verlag, 1988.
35. Bernauer J. Conceptual graphs as an operational model for descriptive findings. Proc 15th Annu Symp Comput Appl Med Care. 1991:214–8.
36. Johnson K, Poon A, Shiffman S, Lin R, Fagan L. A history-taking system that uses continuous speech recognition. Proc 16th Annu Symp Comput Appl Med Care. 1992:757–61.
37. Rassinoux AM, Wagner JC, Lovis C, Baud RH, Rector A, Scherrer JR. Analysis of medical texts based on a sound medical model. Proc 19th Annu Symp Comput Appl Med Care. 1995:27–31.
38. Sager N, Lyman M, Bucknall C, Nhan N, Tick LJ. Natural language processing and the representation of clinical data. J Am Med Inform Assoc. 1994;1:142–60.

APPENDIX

Specification and Characterization of the Searching Questions
Specification and Characterization of the Searching Questions

Case (x)	*woman 30; Hodgkin's disease; autologous bone marrow transplantation*	*category*
day 10:	*You are on duty and a nurse of the haemato-oncology department calls you. Mrs. (x) has got fever.*	
1.	What was the indication for hospitalization?	I
2.	Which diagnoses and treatments were given in the past?	I
3.	What was the initial treatment plan after intake?	II
4.	History and physical examination are negative. Make up a differential diagnosis for the cause of the fever (at least 3 options).	II
day 29:	*Discharge*	
> discharge:	*Mrs (x) has been discharged. You saw her only the last two days and you must dictate the discharge letter.*	
5.	How many cells were in the bone marrow transplant?	II
6.	Which was, in retrospect, the cause of the stomatitis that started on the 15th day, and how was it treated.	II
7.	Which was, in retrospect, the cause of the fever that occurred on the 10th day?	II
Case (y)	*man 47; liver cirrhosis with alcohol abuse; renal failure*	*category*
day 2:	*Mr. (y) was hospitalised yesterday via the emergency ward. Today you see him for the first time, being the physician of the inpatient ward.*	
8.	Were there signs of ascites found in the emergency ward?	III
9.	Were there signs of ascites found in the past?	I
10.	Which were, yesterday, the considerations for requesting a bone marrow punction?	II
11.	Mr. (y) has symptoms at hands and feet. How were they designated and what was done to support this diagnosis?	II
day 21:	*You are on duty and you are called because Mr. (y) has got fever.*	
12.	What do you find during physical examination and what is the most likely cause of the fever ?	III
day 56:	*It is considered to transfer Mr (y) to a nursing home. You are asked for a second opinion.*	
13.	You read that Mr (y) has a renal failure. Which tests were performed to determine its aetiology, and which diagnosis has been established?	II
14.	You read about a positive blood culture with staphylococcus aureus. Yet, the treatment with flucloxicilline was interrupted and replaced by other antibiotics. Why?	II
Case (z)	*woman 53; acute myeloid leukaemia; diagnosis and first chemotherapy*	*category*
day 12:	*You are on weekend duty and are making rounds. Mrs (z) lost 13 kilos within 1 week..*	
15.	Did the diuretic treatment exceed its goal or not? Find also treatment considerations to support your answer.	III
day 23:	*You are on duty and read a remark on day 22 "no haematuria any more".*	
16.	Since when were there signs of haematuria, and what was considered as its cause?	III
day 27:	*Discharge*	
> discharge:	*Mrs (z) has been discharged. You saw her only the last two days and must dictate the discharge letter.*	
17.	What was considered the cause of the leg ulcers and how were they treated?	II
> discharge:	*Mrs (z) has been discharged. You are auditing protocol adherence.*	
18.	Did Mrs (z) receive GM-CSF, as a part of the protocol? Why was this treatment not accomplished conforming the protocol?	II
19.	You are looking for complications of the treatment. Were there signs of congestive heart failure, during the treatment or before?	III

Category: I = past history; II = progress notes; III = medical history or physical examination.