

Research Paper ■

## Simulating an Integrated Critiquing System

---

MANON M. KUILBOER, MD, MSc, JOHAN VAN DER LEI, MD, PhD,  
JOHAN C. DE JONGSTE, MD, PhD, SHELLEY E. OVERBEEK, MD,  
BEN PONSIOEN, MD, PhD, JAN H. VAN BEMMEL, PhD

**Abstract Objective:** To investigate factors that determine the feasibility and effectiveness of a critiquing system for asthma/COPD that will be integrated with a general practitioner's (GP's) information system.

**Design:** A simulation study. Four reviewers, playing the role of the computer, generated critiquing comments and requests for additional information on six electronic medical records of patients with asthma/COPD. Three GPs who treated the patients, playing users, assessed the comments and provided missing information when requested. The GPs were asked why requested missing information was unavailable. The reviewers reevaluated their comments after receiving requested missing information.

**Measurements:** Descriptions of the number and nature of critiquing comments and requests for missing information. Assessment by the GPs of the critiquing comments in terms of agreement with each comment and judgment of its relevance, both on a five-point scale. Analysis of causes for the (un-)availability of requested missing information. Assessment of the impact of missing information on the generation of critiquing comments.

**Results:** Four reviewers provided 74 critiquing comments on 87 visits in six medical records. Most were about prescriptions ( $n = 28$ ) and the GPs' workplans ( $n = 27$ ). The GPs valued comments about diagnostics the most. The correlation between the GPs' agreement and relevance scores was 0.65. However, the GPs' agreements with prescription comments (complete disagreement, 31.3%; disagreement, 20.0%; neutral, 13.8%; agreement, 17.5%; complete agreement, 17.5%) differed from their judgments of these comments' relevance (completely irrelevant, 9.0%; irrelevant, 24.4%; neutral, 24.4%; relevant, 32.1%; completely relevant, 10.3%). The GPs were able to provide answers to 64% of the 90 requests for missing information. Reasons available information had not been recorded were: the GPs had not recorded the information explicitly; they had assumed it to be common knowledge; it was available elsewhere in the record. Reasons information was unavailable were: the decision had been made by another; the GP had not recorded the information. The reviewers left 74% of the comments unchanged after receiving requested missing information.

**Conclusion:** Human reviewers can generate comments based on information currently available in electronic medical records of patients with asthma/COPD. The GPs valued comments regarding the diagnostic process the most. Although they judged prescription comments relevant, they often strongly disagreed with them, a discrepancy that poses a challenge for the presentation of critiquing comments for the future critiquing system. Requested additional information that was provided by the GPs led to few changes. Therefore, as system developers faced with the decision to build an integrated, non-inquisitive or an inquisitive critiquing system, the authors choose the former.

■ JAMIA. 1998;5:194–202.

Decision-support systems have been shown to be able to provide users with support.<sup>1–3</sup> Most of these systems, however, have failed to get incorporated into

daily clinical practice.<sup>4,5</sup> The main reason for this failure is the failure to meet the specific requirements of the future users, resulting in a mismatch between

problem and solution.<sup>6</sup> For example, the system requires special data entry that interferes with normal practice, it is too time-consuming for daily use, the system's timing does not fit the clinical routine, or it ignores the physician's intelligence.<sup>7-9</sup>

Researchers have argued that decision-support systems need to be integrated with electronic medical records to improve these systems' chances to be incorporated into the physician's daily routine.<sup>7,10</sup> Such an integration with the electronic medical record allows a decision-support system to review or critique the physician's treatment using the data already available in the electronic medical record. In The Netherlands, over 50% of general practitioners have been using an electronic medical record for several years, making the time ripe for the development of integrated decision-support systems.<sup>10</sup> We are developing a particular kind of integrated decision-support systems, *critiquing systems*, that generate critiquing comments based on the user's actions as recorded in these medical records.<sup>11-13</sup>

Integrated critiquing systems aim to support physicians based on facts already entered in the electronic medical record, thus avoiding the problem of double data entry.<sup>4</sup> We are building integrated systems that will not ask the physician for additional data: *non-inquisitive* critiquing systems. The downside of this approach is the limited availability of data.<sup>14,15</sup> That is, the ability of such a system to critique diagnosis and treatment is limited by the data available in the electronic medical record. If the electronic medical records do not contain sufficient data, the concept of an integrated, non-inquisitive critiquing system is unfeasible. To determine the feasibility of such a system, we need insight into the number and the nature of comments that can be made based upon the information in the electronic medical record.

If the lack of patient data in the record prohibits the development of a non-inquisitive critiquing system,

---

Affiliations of the authors: Department of Medical Informatics, Erasmus University Rotterdam (MMK, JvdL, JHvB); Sophia Children's Hospital Rotterdam (JCdJ); Dijkzigt Academic Hospital Rotterdam (SEO); and Department of Family Practice, Erasmus University Rotterdam (BP); all in The Netherlands.

Presented in part at the AMIA Fall Symposium, Nashville, Tennessee, November 1997. Supported by The Netherlands Asthma Foundation (# 92.62).

Correspondence and reprint requests to: M. M. Kuilboer, MD, Erasmus University Rotterdam Ee2100, PO Box 1738, 3000 DR Rotterdam, The Netherlands. e-mail: <kuilboer@mi.fgg.eur.nl>.

Received for publication 2/18/97; accepted for publication 10/29/97.

we can consider a separate module that requests additional information. To determine the viability of a separate data collection module, we need insight into the availability of information missed from the record for the critiquing task. Such a module would be useful only when physicians are able to provide the required information. In addition, we have to gain insight into the relevance of this information. When the impact of additional information on the generation of comments is small, obtaining the additional data may require too much effort on the part of the general practitioner.

Whether a critiquing system will be rejected or accepted is also determined by the users' judgment of its critiquing comments. To determine which critiquing comments might be perceived as inappropriate, builders of a critiquing program need insight into general practitioners' responses to these critiques.<sup>12</sup>

Before building an integrated non-inquisitive critiquing system, a system builder thus has to face a number of questions that center around two issues:

- Will it be possible to generate critiquing comments based on the information available in the electronic medical record, and how will general practitioners judge them?
- How much information is missing? Can general practitioners provide the missing information? Why and why not? Does provided information make a difference for the generation of critiquing comments?

In the past, we addressed these issues by building and evaluating prototypes.<sup>16</sup> This process, however, is very time-consuming. An alternative to building prototypes is to perform a simulation study in which humans play the role of the system. To our surprise, we have not found examples of studies using such an approach. The closest comparable technique is used in the field of human-computer interface (HCI). It is called the "Wizard-of-Oz" technique; to reveal important aspects of an interface design, humans play the role of a computer.<sup>17</sup> The user's commands are interpreted by humans, who, invisible to the users, generate the expected responses. The difference of our approach from the Wizard-of-Oz technique is that we do not blind our users for the fact that humans play the role of the computer system.

In this article, we report the results of a small-scale simulation study that attempted to answer the system builders' questions with regard to a critiquing system that supports general practitioners in the diagnosis and treatment of patients with asthma/chronic obstructive pulmonary disease (COPD).

## Methods

In this simulation study, we reviewed six medical records of patients who had been diagnosed as having chronic respiratory disease (asthma/COPD) by their general practitioners. The records were randomly selected from the electronic medical record systems of three general practitioners. In The Netherlands, most general practitioners make use of electronic medical records that adhere to the national standard prescribing the data elements that an electronic medical record should contain (WCIA).<sup>18</sup> For our study, we worked with physicians who were using the general practitioners' information system ELIAS, one of the most commonly used information systems for general practitioners in The Netherlands.<sup>10</sup>

The role of the computer system was played by four reviewers with special interest in asthma/COPD: two specialists (one pulmonologist and one pediatric pulmonologist) and two general practitioners. The role of "users" was played by the same three general practitioners who provided the medical records. The simulation was conducted in three phases as illustrated in Figure 1.

### Reviewers' Comments and Requests for Further Information

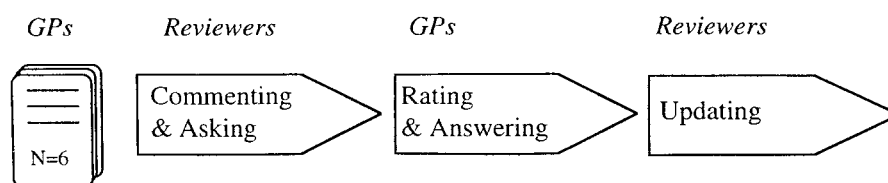
In the first phase of the study (see Figure 1), we provided each reviewer with the medical records. For each visit documented in the record, we asked the reviewer to formulate suggestions for changes in the physician's patient management—*critiquing comments*. Also, we asked the reviewer to verify whether the record contained sufficient information to comprehend the general practitioner's interventions. If the reviewer felt that information was missing, we asked him to formulate this as a *request for additional information*. As each reviewer worked independently, they sometimes used different formulations of essentially the same comment. To enable comparison, we mapped those comments to a single comment. Sub-

sequently, we asked the reviewers to verify the mapping. Finally, we submitted all comments to all reviewers, and we asked each reviewer to indicate for each comment whether he agreed with it.

For the analysis, we assigned each comment to one of four categories:

- *Diagnostic* comments dealt with the diagnostic part of the doctor–patient encounter (examples: “Before the diagnosis of asthma can be established, the presence of allergies should be investigated” and “The child has an upper respiratory tract infection; she should have her ears, nose, and throat examined”).
- *Workplan* comments dealt with the physician's proposed therapeutic strategy (examples: “The patient is using too many bronchodilating agents; anti-inflammatory therapy is indicated” and “The child is taking ketotifen, which is not indicated for children older than 4 years without frequent symptoms”).
- *Prescription* comments dealt with prescription specifications (examples: “The prescription frequency is too high” and “The prescription of different routes of administration is irrational”).
- *Follow-up* comments dealt with the timing of a follow-up (examples: “The patient should return in six weeks instead of three months because his condition is instable” and “The follow-up is insufficient because the effect of this nasal corticosteroid should be checked”).

Because the reviewers worked independently, different reviewers could also request identical additional information using slightly different wording. We mapped these requests from more than one reviewer to a single request. For the analysis, we assigned the requests for additional information to one of three categories. Two of the three categories dealt with missing facts, and one category dealt with missing reasoning:



**Figure 1** Four reviewers analyzed six medical records. The reviewers generated comments and requested further information when needed. The general practitioners rated these comments and provided the missing information. When information was not available, they were asked to explain why. Finally, the reviewers updated their comments, taking the additional information into account.

- Requests about *Factual patient data* dealt with missing data of the medical history, physical examination, diagnosis, or additional tests (for example, "What did the pulmonary examination reveal?", "What are the patient's symptoms after this period of two years?", "What is the patient's condition after treatment with inhalation corticosteroids?").
- Requests about *Factual therapeutic data* dealt with the physician's therapeutic interventions (for example, "What was the exact amount of medication?", "Which medication has been continued?", "How much corticosteroids has the patient been instructed to take per dosage?").
- Requests about *Motivation* dealt with missing information about the general practitioners' motivation for their policies (for example, "Why did the physician change the medical device?", "What was the indication for oxazepam—nocturnal asthma?", "Why doesn't the doctor do anything?").

### General Practitioners' Ratings and Answers

In the second phase of the study, we asked the general practitioners to consider each individual comment and to rate its correctness (on a five-point scale ranging from complete disagreement to complete agreement) and its relevance (on a five-point scale ranging from completely irrelevant to completely relevant). The relevance of a comment was defined as "being relevant for this situation." In addition, we asked the general practitioners to answer the reviewers' requests for additional information. This question could result in one of two situations: 1) the physician could not provide the requested information, in which case he was asked to explain why; 2) if he could provide the requested information, he was asked to explain why he had not recorded the information in the first place.

### Reassessment of the Comments by the Reviewers

In the third phase, we asked the reviewers to reassess their initial comments. We provided the reviewers with the original records, their comments, their requests for additional information, and the additional information given by the general practitioners. We subsequently gave the reviewers the opportunity to retain, withdraw, or change comments, or to add new comments.

### Analysis

For analysis, we counted the comments per category and the requests for additional information per cate-

Table 1 ■

Frequencies and Percentages of Comments Made by Reviewers per Category

Category	Frequency	Percentage
Diagnostics	13	18
Workplan	27	36
Prescription	28	38
Follow-up	6	8
TOTAL	74	100

gory. As an indication of agreement among the reviewers, we counted per comment the number of reviewers that agreed with that comment. To explore the comments' relevance and correctness as given by the general practitioners, we used descriptive analysis and calculated the correlation coefficient. To analyze the causes for requested information to be (un-)available, we counted the reasons given by the general practitioners per category. To analyze the impact of additional information, we counted the number of comments that the reviewers left unchanged, withdrew, changed, or added.

## Results

The six patient records covered 87 visits, on average 14.5 (range: 5–24) visits per record. The reviewers made a total of 74 different comments, on average 0.9 per visit.

### Reviewers' Comments and Requests for Further Information

#### Categories of Comments Made by the Reviewers

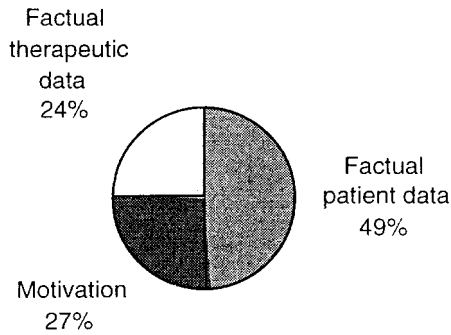
The number of reviewers' comments per category is shown in Table 1. The largest categories of comments were related to *Prescriptions* ( $n = 28$ ; 38%) and the physician's *Workplan* ( $n = 27$ ; 36%).

#### Missing Information

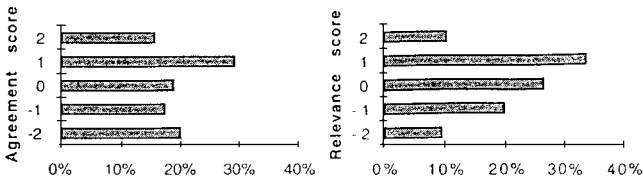
The reviewers stated a total of 132 requests for additional information, which we mapped to 90 single requests. The percentage of each category of requests for additional information is shown in Figure 2.

#### Assessment of Agreement among the Reviewers

Of the 74 comments made by the reviewers, 45% were endorsed by all four reviewers, 31% by three, 12% by two, and 12% by only one expert. In two of the 74 comments, the reviewer who had stated the comment subsequently disagreed with his own comment.



**Figure 2** Summary of information missed in six electronic medical records by reviewers. Three categories of missing information could be identified: *Factual patient data* ( $n = 44$ )—any request for additional information related to a patient’s medical history, physical examination, diagnosis, or additional test; *Factual therapeutic data* ( $n = 22$ )—requests asking the physician about his or her therapeutic strategy; *Motivation* ( $n = 24$ )—requests asking for the physician’s motivation for his or her interventions.



**Figure 3** Distribution of the individual agreement scores and relevance scores ( $n$  scores = 424) of three general practitioners for comments ( $n$  comments = 74) generated by reviewers. The vertical axes shows the range of the scores that the general practitioners could assign (−2 representing *complete disagreement* to +2 representing *complete agreement* and −2 representing *completely irrelevant* to +2 representing *completely relevant*, respectively). The horizontal axes show the percentages with which each score was assigned.

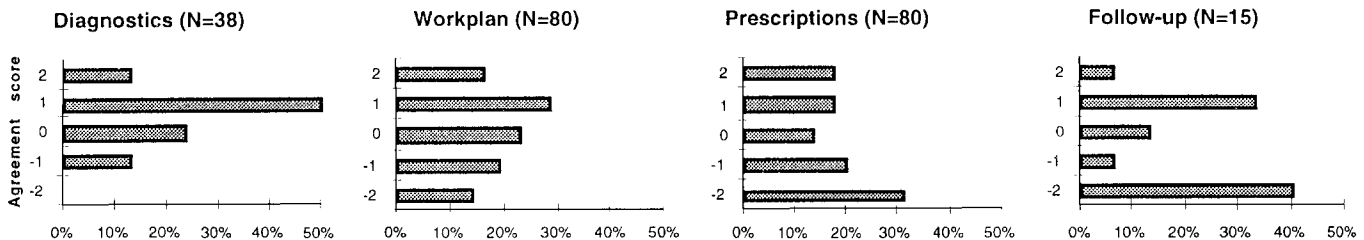
**General Practitioners’ Ratings of Comments**

Each of the three general practitioners rated each individual comment for correctness and relevance on a five-point scale, resulting in 222 judgments of correctness and 222 judgments of relevance. Of these judgments, the general practitioners explicitly had no opinion in 9 (correctness) and 11 (relevance) cases. These judgments were excluded from further analysis.

Figure 3 shows the overall distribution of the general practitioners’ judgments. The correlation coefficient between the three general practitioners’ agreement scores and their relevance scores was  $r = 0.65$ . The most frequently assigned scores were *agreement* (code: +1) and *relevant* (code: +1). In almost 20% of the cases, the general practitioners *completely disagreed* with a comment, but only 10% of the comments were judged *completely irrelevant*.

Figures 4 and 5 show the general practitioners’ judgments for the four individual categories of comments *Diagnostics*, *Workplan*, *Prescription*, and *Follow-up*. Overall, the general practitioners rated the category of comments regarding *Diagnostics* positively, both for their agreement with a comment as well as their judgment of its relevance.

The agreement scores and relevance scores of the comments regarding the general practitioners’ *Workplan* were also generally positive, even though 14% (11/80) of these judgments were *complete disagreement* and an equal percentage (also 11/80) were judged *completely irrelevant*. The general practitioners gave a relatively large number of comments in the category *Prescriptions* a negative agreement score (*complete disagreement*: 31% (25/80)). In contrast, the general practitioners were less negative about the relevance of these comments (*completely irrelevant*: 9% (7/78)).



**Figure 4** Agreement scores of general practitioners ( $N = 213$ ) for comments ( $n = 74$ ) generated by reviewers. The results are shown by the four categories of comments: *Diagnostics* ( $n = 13$ ), *Workplan* ( $n = 27$ ), *Prescription* ( $n = 28$ ), and *Follow-up* ( $n = 6$ ). For each category, the distribution of the agreement scores is shown by the horizontal bars. The vertical axes show the ranges of the scores that the general practitioners could assign (−2 representing *complete disagreement* to +2 representing *complete agreement*). The horizontal axes show the frequencies with which the scores were given.

### General Practitioners' Answers to the Requests for Further Information

The reviewers had stated 90 different requests for additional information. The general practitioners were able to provide information responding to 58 (64%) of the reviewers' requests. The reasons the information had not been recorded in the medical record are summarized in Table 2. In 54% of the 58 answered requests, the physician indicated that the requested information had not been explicitly recorded in the medical record (e.g., why something had not been done). In 22%, the requested information had been assumed to be known (e.g., "fever" means a temperature above 38.5°C). In 17% of the cases, the requested information had been recorded elsewhere in the electronic medical record (e.g., information recorded as a personal note to the record). In 5% of the cases, the information had not been recorded in the electronic medical record yet, but had been available in the paper-based record. (In The Netherlands, most general practitioners use electronic medical records, while in the past, they used paper-based records. During the transition from paper-based records to electronic medical records, the two types of records temporally co-exist until all relevant medical data have been recorded electronically.) Finally, in 2%, the information was provided by an external individual (e.g., a family member).

For 32 of the 90 requests (36%), the general practitioners were not able to provide the requested information. The reasons requested information was unavailable are summarized in Table 3. In 41% of these 32 cases, the general practitioner indicated that the decision had been made by another individual (most commonly the general practitioner on call during the night or on weekends). In 37%, the physician did not know the answer to the request, nor did he know where to locate the missing information (e.g., infor-

Table 2 ■

Reasons Information That Was Available When Requested (*n* = 58), Had Not Been Recorded in the Electronic Medical Record

Frequency		Reason
No.	(%)	
31/58	(54)	Not explicitly recorded
13/58	(22)	Assumed to be known
10/58	(17)	Registered elsewhere in the electronic medical record
3/58	(5)	Registered in the paper-based record
1/58	(2)	Other source

Table 3 ■

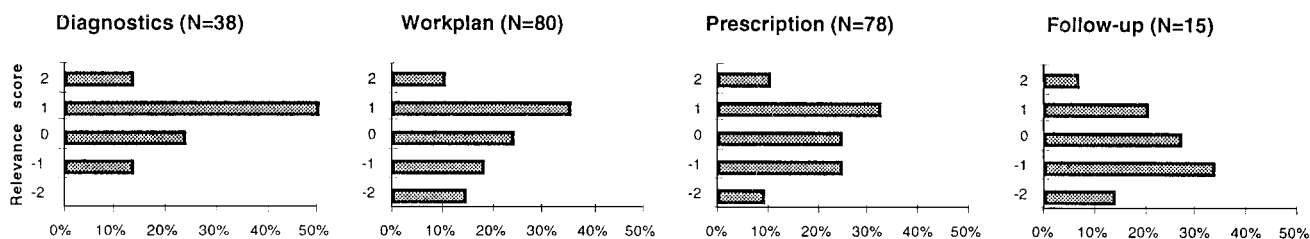
Reasons Requested Information was Unavailable (*n* = 32)

Frequency		Reason
No.	(%)	
13/32	(41)	Other decision maker
12/32	(37)	Information not known
6/32	(19)	Too much effort required
1/32	(3)	Request unclear

mation about the physical examination had not been recorded at the time of the visit). In 19% of the cases, the physician knew where to find the information, but had not made the effort to retrieve it (e.g., in the paper-based record). In 3%, the request could not be answered because it was unclear to the physician.

### Reassessment of the Comments by the Reviewers

After the general practitioners had provided the requested additional information, the reviewers received the medical records, their comments, their re-



**Figure 5** Relevance scores of general practitioners (*N* = 211) for comments (*n* = 74) made by reviewers. The results are shown by the four categories of comments: *Diagnostics* (*n* = 13), *Workplan* (*n* = 27), *Prescription* (*n* = 28), and *Follow-up* (*n* = 6). For each category, the distribution of the relevance scores is shown by the horizontal bars. The vertical axes show the ranges of the scores that the general practitioners could assign (-2 representing *completely irrelevant*, to +2 representing *completely relevant*). The horizontal axes show the frequencies with which the scores were given.

quests for additional information, and the provided additional information to review their comments. The reviewers left 55 (74%) comments unchanged, withdrew 11 of them (15%), changed 8 (11%) comments, and made 15 new ones.

## Discussion

We performed a simulation study to gain insight into some of the issues that determine the feasibility and effectiveness of a computer-based critiquing system that will support general practitioners in the treatment of their patients with chronic lung diseases. In this simulation study, we focused on issues that center around the availability of medical data for critiquing, and the role of missing information. In addition, we investigated the kinds of comments that could be made and the general practitioners' assessments of these comments.

Our study scope was small and thus the potential for an extensive analysis was limited. A more extensive design would have made a more extensive analysis possible, but it would have cost more time and effort; the physicians need time and patience to work through the medical data, comments, and changes.

However, a more extensive study would have made it possible to analyze generated comments in relationship to the general practitioners' assessments on a more detailed level, making more detailed recommendations possible. In addition, instead of being purely descriptive, the analysis could have been extended to a statistical analysis of changes in comments. In retrospect, an extension of the study to include the general practitioners' reassessments of the edited comments would have been valuable.

The design of a simulation study depends on the lessons that need to be learned from it. An advantage of a simulation study of a computer system is that feedback is possible on issues that, when prototyping, could have emerged only at a very late stage. For example, the role of additional information could have been investigated only when additional modules had been programmed, and sufficient functionality would have been available for which this information would have made a difference. From our study, we could draw the conclusions that we required to determine some of the core aspects of our system design.

Critiquing systems that are integrated with a general practitioner's information system work with medical data as they are currently available in general practitioners' electronic medical records. Therefore, the available data for the system will be limited to the data that a general practitioner is able and willing to

enter into the electronic medical record. This is a potential limitation that may impair the generation of useful critiquing statements in clinical domains such as chronic lung diseases. However, as P. Miller pointed out, it remains to be seen whether a limited availability of data necessarily limits the effectiveness of a critiquing system.<sup>19</sup> In other words, there may be reasons why it is good to be generic. The acceptability of a system may improve when comments are less specific, because, for example, comments are less likely to be wrong.

As system developers, we are faced with the decision between a *non-inquisitive* and an *inquisitive* system design. Our simulation study showed that it is possible for human reviewers to generate critiquing comments (on average, one comment per visit), despite the fact that the reviewers in our study often missed information (90 requests for further information were stated over 87 visits). The majority of the comments (74%) were left unchanged by the reviewers after the participating general practitioners had provided additional information for 64% of the requests. On the other hand, 26% of the comments were changed and 15 new comments were made, showing that additional information may change some comments or give rise to additional ones.

To assess the feasibility of an *inquisitive* design, we assessed the availability of missing information. In our study, one-third of the requests could not be answered. To explain why requested information was unavailable, the general practitioners most often mentioned that the decision that had been asked about had been made by a decision maker other than the patient's personal general practitioner. Therefore, the general practitioner could not provide the requested information. Even though in Dutch health care general practitioners function as gatekeepers, this observation illustrates the fact that a single patient receives care from an increasing number of different health care workers. This increase in number of health care providers creates a need for a better management of health-care information.

In our study, two-thirds of the requests could be answered. To explain why the requested information had not been recorded (i.e., the information was available upon request), the general practitioners most frequently (54%) indicated that they normally did not record that information explicitly. For example, the motivation for a particular choice of therapy may not be recorded. Information about a physician's reasoning was often recorded implicitly, and available only when asked for.<sup>13,20</sup> Some of the requested information turned out to be available elsewhere in the electronic medical record (17%). For example, information had

been recorded as a short personal note in free text (not necessarily understood by others). In other cases, the information was assumed to be known by the readers of the medical record (22%). To address these limitations, the current electronic patient record will have to be modified with emphasis on structured data entry. The challenge that such systems have to face is to try to combine complexity with clarity and ease of use.<sup>21,22</sup>

The fact that requested additional information was available in many cases supports the option to build an *inquisitive system*. About two-thirds of the missing information was available only when requested. However, an inquisitive system will have a much larger impact on the physician's normal routine, and therefore runs a larger risk of being rejected. Also, the majority of comments remained unchanged when the requested information became available, while we do not yet know the impact of the minority of changed comments. Therefore, awaiting the results of our further studies, we have chosen a *non-inquisitive* design.

Having discussed the implications of our finding that critiquing comments could be made by human reviewers based upon data as they are currently available, we now discuss the kinds of comments that could be made and the general practitioners' assessments of these comments.

The largest categories of comments were those critiquing the prescribed medication and the general practitioner's therapeutic strategy in general. Interestingly, the reviewers' comments about the diagnostic phase of the patient-doctor encounter (though not made very frequently) were judged very positively. The general practitioners' positive response to these comments may suggest a need for support during the diagnostic phase. This observation seems to be in contradistinction to studies that have shown that diagnostic systems have had little impact on daily clinical practice.<sup>23</sup> Possibly, the kind of diagnostic support that is appreciated by physicians (support with diagnostic work-up) differs from the kind of support that diagnostic systems have provided in the past (support with differential diagnosis). When describing major obstacles to the implementation of decision-support systems, Taylor identified "loss of clinical control" as one of the possible reasons diagnostic systems have achieved so little.<sup>9</sup> The fact that critiquing leaves the physician in control could account for our finding that general practitioners appreciated the diagnostic comments.

Prior to this study, we believed that if a general practitioner disagreed with the content of a critique, he would also judge that critique to be irrelevant. Overall, the agreement score and relevance score correlate,

with  $r = 0.65$ . We were surprised to find that in a number of cases the general practitioner strongly disagreed with the content of a comment, but did not judge the comment to be irrelevant. This was most pronounced in the category of prescription-related comments. In other words, the general practitioners could see that a comment was relevant, but they could still strongly disagree with its content. This observation may imply that comments regarding prescriptions are very much needed from the point of view of the quality of health care—comments about prescriptions were frequently made—but that it will be a challenge to get physicians to accept prescription-related recommendations.

More insight is needed into the reasons physicians reject critiquing comments in order to make the distinction between a reluctance of the physician to accept advice and a disagreement of the physician with the content of the advice.

## Conclusion

We performed a simulation study of a computer system in order to gain insight into issues that determine the feasibility and effectiveness of an integrated critiquing system. Even though the reviewers missed a considerable amount of information, our simulation study showed that it is possible for human reviewers (and therefore, theoretically feasible for computer algorithms) to generate critiquing comments based upon patient medical data as they are currently stored in electronic medical records. The largest categories of comments were about prescriptions and the physician's workplan. Comments regarding the diagnostic process are highly appreciated by the general practitioners. Interestingly, even though we investigated only a limited number of electronic medical records, the general practitioners judged prescription comments to be relevant, but often strongly disagreed with them. This discrepancy poses a challenge for the acceptability of critiquing comments that will be made by the future critiquing system. The general practitioners could provide answers to about two-third of the reviewers' requests for additional information. When this missing information was obtained, it led to changes in only a minority of generated comments. Until integrated decision-support systems have available more data than we observed, general practitioners' information systems will have to be developed that better support the structured entry of medical data. As a result of this study, we have started building the non-inquisitive critiquing system Asthma-Critic. AsthmaCritic will be subject to a field study, in which we will investigate the relationship between



general practitioners' opinions of comments' correctness and relevance, the role of missing information, and the system's effectiveness.

The non-author reviewer in this study, Prof. Dr. E. van der Does, and the three general practitioners, Drs. J. Brienen, C. Kunst, and J. van Wijngaarden, receive special thanks.

#### References ■

- McDonald CJ. Protocol-based computer reminders, the quality of care and the non-perfectibility of man. *N Engl J Med*. 1976;295:1351–5.
- McDonald CJ, Wilson GA, McCabe GJJ. Physician response to computer reminders. *JAMA*. 1980;244:1579–81.
- McDonald CJ, Hui SL, Smith DM, et al. Reminders to physicians from an introspective computer medical record. A two-year randomized trial. *Ann Intern Med*. 1984;100:130–8.
- Elson RB, Connelly DP. Computerized decision-support systems in primary care. *Primary Care*. 1995;22:365–84.
- Shortliffe EH, Buchanan BG, Feigenbaum EA. Knowledge engineering for medical decision making: a review of computer-based clinical decision aids. *Proc IEEE*. 1979;67:1207–24.
- Heathfield HA, Wyatt J. Philosophies for the design and development of clinical decision-support systems. *Meth Inform Med*. 1993;32:1–8.
- Miller RA. Medical diagnostic decision support systems—past, present, and future: a threaded bibliography and brief commentary. *J Am Med Informatics Assoc*. 1994;1:8–27.
- Shortliffe EH. Testing reality: the introduction of decision-support technologies for physicians. *Meth Inform Med*. 1989;28:1–5.
- Taylor TR. The computer and clinical decision-support systems in primary care. *J Fam Pract*. 1990;30:137–40.
- Lei van der J, Duisterhout JS, Westerhof HP, et al. The introduction of computer-based patient records in The Netherlands. *Ann Intern Med*. 1993;119:1036–41.
- Lei van der J, Musen M. A model for critiquing based on automated medical records. *Comput Biomed Res*. 1991;24:344–78.
- Lei van der J, Does van der E, Man in't Veld AJ, Musen MA, Bommel van JH. Response of general practitioners to computer-generated critiques of hypertension therapy. *Meth Inform Med*. 1993;32:146–53.
- Lei van der J, Musen M, Does van der E, Man in't Veld AJ, Bommel van JH. Comparison of computer-aided and human review of general practitioners' management of hypertension. *Lancet*. 1991;338:1505–8.
- Melker de RA, Jacobs HM, Kreuger FAF, Touw-Otten FWMM. Medische verslaglegging van huisartsen. *Huisarts en Wetenschap*. 1994;37(2):46–51.
- Gilliland AEW, Millis KA, Steele K. General practitioner records on computer—handle with care. *Fam Pract*. 1992;9:441–50.
- Shortliffe EHS, Perreault LE. *Medical Informatics. Computer Applications in Health Care*. Reading, MA: Addison-Wesley, 1990.
- Detmer WM, Shiffman S, Wyatt JC, Friedman CP, Lane CD, Fagan LM. A continuous-speech interface to a decision-support system: II. An evaluation using a Wizard-of-Oz experimental paradigm. *J Am Med Informatics Assoc*. 1995;2:46–57.
- Overbeeke van JJ, Westerhof HP (eds). *WCIA-HIS-Referentiemodel 1995*. Volume A. Utrecht, The Netherlands: NHG/LHV Utrecht, NL, 1996:236.
- Miller PL, Frawley SJ. Trade-offs in producing patient-specific recommendations from a computer-based clinical guideline: a case-study. *J Am Med Informatics Assoc*. 1995;2:238–42.
- Vlug AE, Lei van der J. Postmarketing surveillance with computer-based patient records. In: Greenes RA (ed). *MEDINFO*. Vancouver, BC, Canada: 1995:327–30.
- Ginneken van AM. Structured data entry in ORCA: the strengths of two models combined. In: Cimino JJ (ed). 1996 *AMIA Annual Fall Symposium*. Philadelphia, PA: Hanley & Belfus, 1996:797–801.
- Moorman PW, Ginneken van AM, Lei Van der J, Bommel van JH. A model for structured data entry based on explicit descriptive knowledge. *Meth Inform Med*. 1994;33:454–63.
- Kassirer JP. A report card on computer-assisted diagnosis—the grade: C. *N Engl J Med*. 1994;330:1824–5.