User Comments on a Clinical Event Monitor

George Hripcsak, Paul D. Clayton
Columbia-Presbyterian Medical Center

Columbia-Presbyterian Medical Center's health care providers have access to alerts and interpretations generated by an Arden Syntax-based clinical event monitor. They have the opportunity to send comments to the clinical information services staff. Over a period of 26 months, they sent 126 comments. The comments were analyzed using the critical incident technique, resulting in a hierarchy of categories that summarizes user concerns. The majority of comments (65) indicated that the messages were actually (8) or at least potentially useful (57). A minority (28) indicated that they were unhelpful (27) or actually harmful (1). Another group (27) made suggestions or asked questions. The comments have been very helpful for the maintenance of Medical Logic Modules (MLMs) and the clinical event monitor itself.

INTRODUCTION

Clinical event monitors, reminder systems, and other forms of automated decision-support systems are experiencing increasing visibility, trust, and use [1]. Such systems improve provider compliance with accepted guidelines [2-4], but response is generally less than 100%, and often as low as 20 to 40%. Finding out what causes health care providers not to follow a guideline is critical to improving event monitors and the guidelines that they implement. Potential causes include: incorrect, outdated, or controversial guidelines; incorrect guideline implementation; confusing, inappropriate, or untimely presentation; extenuating or unusual patient circumstances; etc.

Questionnaires have been used to ascertain the attitude of providers toward automated decision-support systems. Rind and co-authors mailed questionnaires to 622 physicians who had received computer-generated reminders [5]. They reported that 53% of respondents (of 288 who offered an opinion) found the reminders helpful, and 31% (of 312) found the reminders annoying; respondents also judged the appropriateness of reminder criteria. Teach and Shortliffe used questionnaires to assess physicians' general attitudes toward automated decision-support systems [6]. Henderson and co-authors compared providers' compliance with "correct" versus "incorrect" computer-generated instructions and with active versus passive instructions [7]. Other groups have analyzed physicians' assessment of computer-generated consultation advice [8-9].

We took a more open-ended approach. Health care providers at Columbia-Presbyterian Medical Center (CPMC) have access to computer-generated alerts and interpretations on their patients [10]. We set out to obtain feedback (in the form of electronic comments) from providers about the alerts and interpretations, in whatever form and on whatever aspect the provider desired. The goal was to obtain information about what concerned the providers, without biasing their responses with our preconceptions. We then used the critical incident technique [11] to organize their responses into a logical framework. The technique has been used in other contexts, including the analysis of the use of MEDLINE for clinical problem solving [12].

METHODS

CPMC provides automated decision-support through its clinical event monitor [10], which is based upon the Arden Syntax for Medical Logic Modules [13]. Clinical events in the medical center (e.g., admissions, laboratory results) trigger pertinent rules called Medical Logic Modules (MLMs). Each MLM reads data from the clinical database, evaluates medical criteria, and, if appropriate, generates a message. Alert messages warn of emergent situations. For example, one MLM warns of new or worsening renal insufficiency. Alerts are presented to providers when they sign onto the clinical information system to review a patient's data. Interpretation messages convey passive information. For example, one MLM calculates the creatinine clearance from blood and urine laboratory values. Interpretations are presented with their corresponding data values.

During the study, providers reviewed messages on screens like the one in Figure 1. If they wanted to send a comment, they pressed the F6 key. They were then presented with a comment screen, which instructed them to:

> Enter your comments about the computer-generated message, and press F3 when you are finished. They will be forwarded to Clinical Information Services.

The comments were forwarded to the Clinical Information Services staff and logged. Comments

0195-4210/94/$5.00 © 1994 AMIA, Inc.

were collected from the time the system was turned on for clinical care (March, 1992) to the present (May, 1994).

Comments were analyzed using the critical incident technique [11]. This technique has been used to determine the reasons for success or failure of a process involving human beings, based upon narrative answers to open-ended questions. By categorizing the answers, the technique produces a framework from which one can better grasp the issues involved. In this study, we used the open-ended request on the comment screen (quoted above).

Comments were analyzed one by one. Comments that addressed similar issues were grouped into categories and given descriptive phrases. As each new comment was reviewed, it was placed into an existing category, or a new category was created for it. Occasionally, the descriptive phrase was altered slightly to accommodate a new comment. Categories themselves were grouped into higher level categories, thus creating a hierarchy. Whenever a category was added or changed, the number of comments analyzed so far was recorded. These data were later used to assess the reliability of the categories. Comments were ordered randomly to avoid biasing the category definitions due to temporal trends.

RESULTS

From March, 1992 to May, 1994, an average of 20 alert messages were generated per day. Interpretations did not begin until March, 1993; after that, 2000 interpretation messages were generated per day. MLMs were based upon laboratory results, demographics, admit-discharge-transfer data, and discharge diagnoses.

A total of 126 comments were collected during the study period from physicians (97), nurses (19), and medical students (10). The comments were instigated by both alerts (77) and interpretations (49). Comments addressed the computer-generated messages themselves, patient status, and decision-support systems in general. Only 0.5% of alerts led to comments, although they were more numerous when the system was first turned on (2%). Even fewer (0.01%) interpretations led to comments; this is not surprising, since interpretations are passive whereas alerts are active.

The categories generated by the analysis are shown in Table 1. The comments in category 1 indicated that MLMs are at least potentially useful. All eight reports of benefit (1a) involved detection of a missed laboratory abnormality, with a consequent early work up or change in plan. No comment provided proof that a patient outcome was actually improved due to the message. Nevertheless, one comment reported the early placement of a ureteral stint in the setting of acute urinary obstruction; what would have happened if the message had not been sent or if the stint had not been placed as quickly remains unclear. Comments in category 1b focused more on the patient than on the message itself (e.g., "The patient's potassium was repleted. Thank you."). They indicated that the MLMs' advice was appropriate but not specifically that the message made a difference for the patient. Category 1c, which contains simple expressions of gratitude (e.g., "thank you"), was the largest. Users did not specify whether the message actually helped a specific patient.

The comments in category 2 indicated that MLMs are not useful. There was one harmful episode due to a message (category 2a). The physician read an alert that warned of renal insufficiency and called the patient to come into the emergency room for a follow-up creatinine test. The blood test was obtained, and the value had not worsened further. The physician later reviewed the previous creatinines and realized that the MLM had detected a six month increasing trend; there was no need for emergent follow-up. This message did not cause physical harm, but it did incur a cost in patient worry, patient time, emergency room charges, and physician time. Category 2b indicated that the message was not useful for a particular patient (e.g., "Patient is no longer taking digoxin."). Category 2c contained six general statements that MLMs were not useful or helpful.

Categories 3 and 4 contained a number of useful suggestions and comments that led to changes in the MLMs, fixes to the system, and new MLMs. Similarly, six comments from category 2 led to changes in MLMs' behavior. One comment warned that a tuberculosis alert appeared to be incorrect. The laboratory had changed its vocabulary without alerting the clinical database administrator, and an MLM called "positive_TB_culture" generated 93 inappropriate alerts in 20 minutes. The problem was uncovered, and the inappropriate alerts were inactivated before most were seen. Sometime later, another alert by the same MLM reported a "positive tuberculosis culture" based upon finding the organism M. chelonei. A user comment indicated that this was not generally a pathogen, and therefore the culture was not really positive. It was still desirable to send the alert, but the message was rephrased to

read, "tuberculosis culture result." The logic of an MLM that detects renal insufficiency based upon serum creatinine was modified to account for serum ketones because of a user comment. A bug in a central database retrieval program was uncovered by another user comment. Occasionally, comments requested changes in formatting, such as altering spacing or using lowercase letters. Several users commented on the misspelling of |supersede,| which was actually spelled correctly. Some criticisms were difficult to remedy. For example, medications had to be inferred from blood level tests, since pharmacy data were not yet available. Therefore, these MLMs were unreliable. Overall, the comments were very useful for the maintenance of MLMs.

Category 5, which contained comments about other applications, reflected the lack of a suggestion box elsewhere in the clinical information system, and the fact that users see the event monitor as integrated with the rest of the clinical information system. Category 6 contained what appeared to be aborted comments (e.g., word fragments).

First-level categories (1-6) and second-level categories (a-d) underwent no further changes after the 45th comment (out of 126) was analyzed. This implies that adding more comments would probably not change them much. Therefore, the categories appear to be qualitatively representative of user comments. Third-level categories ("subcategories") were altered or added through the 112th comment (out of 126). The subcategories of 2b enumerate reasons why a message was not useful. Since the subcategories evolved through the end of the analysis, there are probably other reasons that are not listed; more comments are needed. Other second-level categories were not divided into subcategories because their comments were either homogeneous or too disparate to group logically.

There was no correlation between type of user (physician, nurse, or medical student) and category of comments.

DISCUSSION

The qualitative categories generated by the analysis are representative of user concerns, based upon the fact that categories remained stable after 45 out of 126 comments were analyzed. Nevertheless, the proportion of comments in each category is likely to be severely biased. For example, one would expect to see more comments from users who were very happy or very unhappy with the messages. Since a small number of messages instigate comments, any differential will be amplified in the result. It is interesting, however, that the proportion of users who found messages useful versus those who found messages not useful corroborates Rind's findings (53% found reminders helpful and 31% found reminders annoying) [5].

It is notable that how few users sent comments. To encourage comments one could alter the screen flow so that it becomes inconvenient *not* to send a comment. This would provide a more accurate picture of how many users hold each opinion, and it might uncover additional reasons for MLM failures (more 2b subcategories), but users might find it intrusive. A limitation of the current mechanism is that comments are biased toward users who feel comfortable enough to write text on a screen. Furthermore, users who do not even look at alert or interpretation messages will not enter the screen flow necessary to send a comment.

The reports of benefit (category 1a) and the reports of potential usefulness (1b and 1c) are heartening, but they are tempered by the report of harm (2b). The possibility of the latter outcome must be recognized whenever such systems are studied or installed. Hopefully, such episodes will be minimized by MLM revisions spurred by user comments.

Providing users with the ability to comment freely on the decision-support system has allowed us to analyze their concerns and helped us to adjust MLMs and the clinical event monitor to suit their needs. The open-ended approach has supplied a broad range of information although it may not reflect the proportion of users with each opinion accurately. While they have been very useful, user comments should not be seen as a replacement for a formal evaluation. In the future, the clinical event monitor will be used to implement care plans and practice guidelines, and user comments will be an integral part of the system.

ACKNOWLEDGMENT

This work was supported by the International Business Machines Corporation and by a grant from the National Library of Medicine LM04419 (IAIMS).

REFERENCES

[1] Johnston ME, Langton KB, Haynes RB, Mathieu A. Effects of computer-based clinical decision support systems on clinician performance and patient outcome. Ann Intern Med 1994;120:135-42.

[2] McDonald CJ, Wilson GA, McCabe GP. Physician response to computer reminders. JAMA 1980;244:1579-81.
[3] Barnett GO, Winickoff RN, Morgan MM, Zielstorff RD. A computer-based monitoring system for follow-up of elevated blood pressure. Med Care 1983;21:400-9.
[4] Pestotnik SL, Evans RS, Burke JP, Gardner RM, Classen DC. Therapeutic antibiotic monitoring: surveillance using a computerized expert system. Am J Med 1990;88:43-8.
[5] Rind DM, Safran C. The development and evaluation of computer-generated alerts in an inpatient setting. In: Lun KC, et. al., editors. Proc. 7th World Congress on Medical Informatics (MEDINFO 92). Amsterdam: North Holland, 1992; 249-254.
[6] Teach RL, Shortliffe EH. An analysis of physician attitudes regarding computer-based clinical consultation systems. Comput Biomed Research 1981;14:542-58.
[7] Henderson SE, Crapo RO, East TD, Morris AH, Wallace CJ, Gardner RM. Computerized clinical protocols in an intensive care unit: how well are they followed? In: Miller RA, editor. Proceedings of the Fourteenth Annual Symposium on Computer Applications in Medical Care; 1990 Nov 4-7; Washington, D.C. New York: IEEE Computer Society Press, 1990: 284-8.
[8] Bankowitz RA, McNeil MA, Challinor SM, Parker RC, Kapoor WN, Miller RA. A computer-assisted medical diagnostic consultation service. Ann Intern Med 1989;110:824-32.
[9] van der Lei J, Musen MA, van der Does E, Man in't Veld AJ, van Bemmel JH. Comparison of computer-aided and human review of general practitioners' management of hypertension. Lancet 1991;338:1504-8.
[10] Hripcsak G, Cimino JJ, Johnson SB, Clayton PD. The Columbia-Presbyterian Medical Center decision-support system as a model for implementing the Arden Syntax. In: Clayton PD, editor. Proceedings of the Fifteenth Annual Symposium on Computer Applications in Medical Care; 1991 Nov 17-20; Washington, D.C. New York: McGraw-Hill, Inc., 1992; 248-52.
[11] Flanagan JC. The critical incident technique. Psychological Bulletin 1954;51:327-58.
[12] Lindberg DAB, Siegel ER, Rapp BA, Wallingford KT, Wilson SR. Use of MEDLINE by physicians for clinical problem solving. JAMA 1993;269:3124-9.
[13] Hripcsak G, Ludemann P, Pryor TA, Wigertz OB, Clayton PD. Rationale for the Arden Syntax. Comput Biomed Res, in press.

```
Computer-Generated Alert Display

 Name: PATIENT, TEST              Sex: M  Birthdate: 10/03/934   MRN: 3131313
 MLM: CREATININE_MONITOR                                Date: 93/10/27  05:27
 ------------------------------------------------------------------------------
    This message is computer-generated; your own clinical judgment must
    supersede.  Press F1 for more information.  Press F6 to send comments.

 The patient's serum creatinine level (1.4 mg/dl on 27 Oct 1993 at 04:44) may
 signify new or worsening renal insufficiency.

 This analysis was based upon the following recent creatinines (the time
 intervals between successive creatinines are not necessarily equal):

 (0.9, 0.7, 0.8, 0.9, 1.4)









 Help=F1  Alert List=F3  Comment on Alert=F6  Scroll Up/Down=F7/F8   Signoff=F11
 Print=F10    Prev Alert=F9    Next Alert=ENTER   Census=F2    CIS Main Menu=F1
```

Figure 1. Alert display screen. The user presses F6 to send a comment.

Total (126)

1. Indications that MLMs are useful (65)
    - a. reports of benefits to a specific patient due to an MLM (8)
    - b. statements that MLM advice was followed for a specific patient, without indicating whether actions would have been different without the MLM message (13)
    - c. general expressions of gratitude ("thank you") and general statements that MLMs are useful (44)
2. Indications that MLMs are not useful (28)
    - a. reports of harm to a specific patient due to an MLM (1)
    - b. statements that an MLM message was not pertinent to or useful for a specific patient (21)
        - i) complicating condition (7)
        - ii) medication discontinued (6)
        - iii) duplicate message (2)
        - iv) laboratory error (1)
        - v) missing message that should have been sent, reported under another message (1)
        - vi) not noted (4)
    - c. general statements that MLMs are potentially wasteful or harmful (6)
3. Suggestions and questions about MLMs (16)
    - a. suggestions and requests for new MLMs (8)
    - b. suggestions for additions to existing MLMs (4)
    - c. suggestions for changes to MLM message wording or spelling (3)
    - d. neutral questions about how an MLM works (1)

        (questions that implied that the user thought the MLM was in error were counted in category 2)
4. Comments about the user interface for messages (6)
    - a. suggestions for improvements to display screens (3)
    - b. reports of bugs in display screens (3)
5. Comments about other systems (5)
    - a. clinical information system (1)
    - b. laboratory system (4)
6. Unintelligible responses (6)

Table 1. Comment categories with number of comments in parentheses.