An Evaluation of Automatically Generated Briefings of Patient Status

Desmond Jordan[a], Gregory Whalen[b], Blaine Bell[b], Kathleen McKeown[b], Steven Feiner[b]

[a]*Departments of Anaesthesiology and Medical Informatics, Columbia University College of Physicians and Surgeons New York, NY 10027, USA*
[b]*Department of Computer Science, Columbia University, New York, NY 10027, USA*

Abstract

We report on an evaluation of MAGIC, a system that automatically generates briefings of patient status after coronary bypass surgery, completed in the Cardio Thoracic Intensive Care Unit at New York Presbyterian Hospital. Through enhancements in system design, robustness and speed, we compared information obtained by nurses against two briefings, one automatically generated by MAGIC and one provided by physicians upon the patient's arrival to the ICU. Our results show that MAGIC and the physician briefing provide a substantial increase in the amount of information than is available prior to the patient's arrival and that the information MAGIC provides is accurate. In many aspects, MAGIC outperforms the physician briefing; information is reported earlier and is always available. We conclude that MAGIC provides the CTICU staff early on with a better assessment of the patient's status than in current practice and allows them to better prepare for the patient's arrival.

Keywords:

Critical Care, Evaluation Studies, Nursing Administration Research

Introduction

In advance of a patient's arrival at the Cardio Thoracic Intensive Care Unit (CTICU), information about their operative course and clinical status must be available to the CTICU medical team. It is well known that this data provides continuity of operative care and is required for prompt and appropriate therapy should immediate problems arise. Current methods for providing this patient briefing suffer from the time constraints under which caregivers work, as well as the structure, organization, amount and varied importance of the information conveyed by the physician who gives the briefing. As a result, the same information is not consistently available for each patient on arrival. We have been developing MAGIC (Multimedia Abstract Generation for Intensive Care) [1] [2], an experimental system to produce briefings automatically of patient status after CABG (coronary artery bypass graft) surgery. Our goal in developing MAGIC is to provide a system that consistently provides much needed information in real time and prior to the patient's arrival to CTICU so that the medical staff has adequate time to prepare for patient care.

In this paper, we report on an evaluation of a robust version of MAGIC that was extended to run on a much wider variety of input data. The automated system was installed in the CTICU at New York Presbyterian Hospital and produced briefings for patients before they arrived in the unit. This study constitutes our first deployment of the system in the CTICU and contrasts with our previous work, in which we evaluated a text-only version of MAGIC offline [3].

In our study, we hypothesized that MAGIC, when compared to the existing workflow of physician briefings, would provide the early transfer of accurate, appropriate and necessary information. Patients were divided into two groups, control and experimental. In both groups, a briefing was given by telephone by a physician in the operating room (OR) to a nurse in the CTICU, around an hour and forty-five minutes before the completion of surgery. Due to time constraints on OR personnel, who are still busy caring for the patient, this briefing is typically cursory. When the patient arrives in the CTICU (CTICU admission), a second briefing is given orally at bedside by the same physician in the OR to the same nurse and physician in the CTICU. At these times, the type and amount of data available was recorded. In the experimental group, the MAGIC briefing was made available throughout the patient's operative course. MAGIC is intended to provide a full interim report automatically, eliminating the need for the interim pre-arrival telephone call and augmenting the transfer of vital information. The type, amount and accuracy of data available through MAGIC were recorded and the results then compared.

Our study was designed to determine whether (1) MAGIC's briefing increases the amount of information that is available in the time period from the telephone call to CTICU admission; (2) MAGIC provides a similar quantity of information as the physician bedside briefing; and (3) information conveyed by MAGIC is accurate when compared against the bedside briefing and the online record. We also tested whether this enhanced version of MAGIC is robust enough to reliably present information requested at the time of the phone call and provide updates until the actual patient arrival.

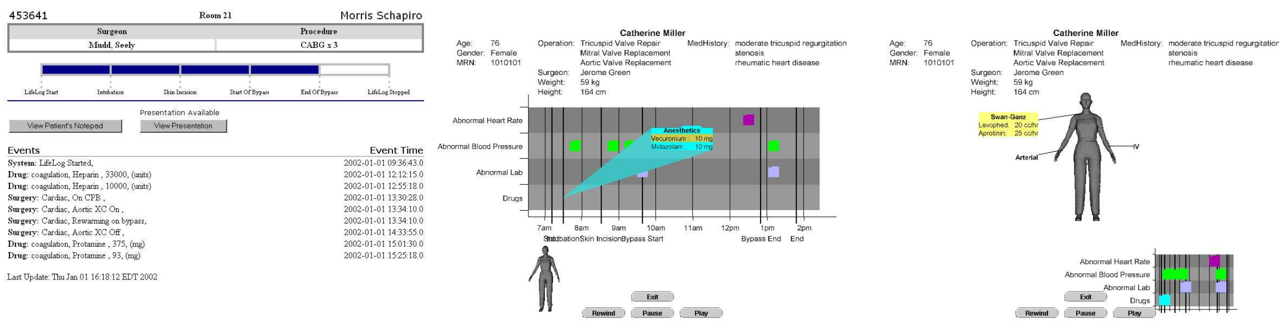

Figure 1: MAGIC Status Display

Methods

Experimental Setup

To record information provided in the two human briefings, as well as by MAGIC, we developed a check sheet that could be easily used by the nurses. This was done jointly with the Nursing Clinical Coordinator and three charge nurses and was based on a modification of the nursing Admission Note, which is an integral component of the daily workflow and regularly used for quality control. The main focus was to make data entry easier by limiting manual data entry to checking boxes and entering values, as opposed to requiring full textual description of any data element. The form included eight categories of data (demographics, lines, procedure specific data, blood products, devices, antibiotics, induction medications and drips). There were a total of 132 individual data items across all categories.

The form was used to record all information communicated in each briefing and use of the form was made part of the nurses' normal routine. Nurses were instructed to write down all information that was communicated in each briefing on a separate form and used them to carry out their normal tasks. Since they needed the information themselves, they were thus motivated to record carefully. For the period of the experiment, all 36 admitting nurses (with 12 years experience on average) in the CTICU participated.

We collected data in a control environment in which MAGIC was not used, as well as in the experimental setting. For the control environment, we collected phone call and bedside forms for each of 90 patients delivered to the CTICU. MAGIC was installed in the CTICU for a two-month period following the initial data collection in the control environment. A phone call, MAGIC, and bedside form was collected for each of 51 patients processed using the MAGIC system.

The nurse was requested to enter values (or check boxes) for each data item present for the patient, leaving blank any items that were not present. For example, 20 possible drips could be given during the operation, but in most cases only a small number of drips are actually used and would be marked on the form.

Once completed, the forms were placed in a return bin and the data entered into a spreadsheet for analysis. First, form data items were grouped into eight clinical categories. The number of form entries for each data item and the mean number of entries for each category were calculated. Data items two standard deviations below the mean for their category were excluded. Second, five physicians who were involved neither in development of MAGIC nor in the data collection effort analyzed the data. For each of the 132 values on the MAGIC form, the physician entered whether it was the same as on the bedside form, different from the bedside, or blank. Different was defined as having conflicting values or incomplete form items. For example, units may be missing on a drips value. We automatically further processed these results comparing MAGIC's values against the online operative record to verify whether these were correct. Error rate was calculated as the number of MAGIC's values that differed both from the bedside and from the online record. In computing the error rate, we did not include data items that had constantly changing values (e.g., drips whose dosage varied over time).

Overview of MAGIC

The new MAGIC system evaluated in this paper consists of two interfaces on adjacent displays positioned at the nurse's station in the CTICU at New York Presbyterian Hospital. For each ongoing operation in the OR (as many as four at once), the Operation Status Monitor, shown on one of the displays (Fig. 1), reveals brief patient summaries, the current operation status, event lists, and some buttons for controlling the presentation. The patient's presentation, shown on the other display, is accessible once an operation has been completed. The presentation consists of graphics and speech that show and explain more detailed information that has been captured from the operating room during the surgery (Fig. 1). Our new version of MAGIC allows the user to pause, stop, rewind, and advance the recitation. Typical presentations last between 1-2 minutes per patient, depending on the quantity of information presented. One of our team was present to help initiate the briefing if needed and control the interface.

During the surgical procedure, information is captured using the LifeLog data acquisition system (Modular Instruments Inc.). Vital signs, inhaled anaesthetics, and ventilation parameters are automatically obtained from medical devices (Hewlett Packard Merlin monitors, Ohmeda anaesthesia machines, and

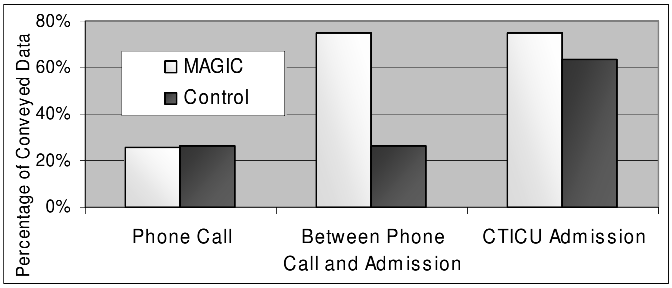

Figure 2: Timeline of Data Availability

Table 1: Differences between Form Categories (P-Values)
C indicates Control Group; E indicates Experimental Group

Category	$Phone^C$/ $Phone^E$	$Phone^C$/ MAGIC	$Phone^C$/ $Bedside^C$	MAGIC/ $Bedside^C$
Demographics	0.25	0.58	0.60	0.931
Lines	0.938	0.426	**P < 0.01**	**P < 0.01**
Procedure-Specific Data	0.04	**P < 0.01**	**P < 0.01**	0.379
Blood Prod.	0.797	**P < 0.01**	**P < 0.01**	**P < 0.01**
Antibiotics	0.025	**P < 0.01**	**P < 0.01**	0.181
Anesthetics	0.045	**P < 0.01**	**P < 0.01**	**P < 0.01**
Drips	0.818	0.0628	0.03	0.124

saturation monitors), while other information, such as bolus drugs, postoperative drugs, laboratory results, intravenous lines, data from echocardiograms, and information about devices such as a pacemaker, is manually entered by the anaesthesiologist using the LifeLog interface. At five-minute intervals, MAGIC's inference engine further processed the captured information by applying a set of inference rules to highlight and group information about abnormal events that occur during the operation. The system uses the captured and processed data to generate graphics, speech, and coordination information for each operation presentation (see [3] for an earlier evaluation of a text-only version of MAGIC).

Data Analysis

Data analysis was performed using analysis of variance for differences between groups and t-test for differences between information categories. Data items were counted as highly correlated if values were present on both forms (modified by accuracy) and uncorrelated otherwise. A significance level of $p < 0.01$ was used. The percent of information capture per clinical category on the forms was calculated as the number of correct or verified form items, divided by the number of patients in the group times the total number of form items for the category. A follow-up questionnaire, based on UMD's QUIS 7 (http://lap.umd.edu/QUIS) was done at the completion of the study to evaluate the user interface, to establish required configuration changes for the next version of the system, and to ascertain user acceptance of the system.

Results

Our study illustrates the difference in amount of information that is available at different points along the timeline from phone call to CTICU admission. We first compare quantity of information that has been conveyed at the time of the phone call, the period between the phone call and the CTICU admission, and at the CTICU admission, in the control and the experimental condition. We then provide a more detailed analysis by category of data and establish MAGIC's accuracy through comparison against two standards and the physician evaluation.

At the time of the phone call from the OR, no differences in the quantity or distribution of data were determined between control and experimental groups (Figure 2). Of note, the phone call provides cursory information, which centers on demographics (patient identifiers), lines and drips. On average, the phone call occurs one hour and forty-five minutes prior to patient arrival in the CTICU. This verbal exchange requires the synchronization of effort between the physician in the OR and the nurse in the CTICU. Thus, the time available for this interchange and transfer of information is short. Our data shows that both groups of patients, control and experimental, were similar prior to the introduction of the experimental condition.

After the phone call and prior to the patient arrival in the CTICU, the information gain in the experimental group is quite evident. There is a 200% increase in the experimental over the control group (Figure 2). Importantly, this information gain differed by category of data (Table 1).

In four categories, MAGIC provided substantially more information than the phone call. In the category procedure-specific data, the MAGIC form consistently provided more information for all individual data items except one (Circulatory Arrest Time), which did not reach our inclusion criteria. The category includes preop cardiac output, postop cardiac output, bypass time, and cross clamp time. Typically, MAGIC provided four times as much information as did the phone call. (50.2% of all items for this category were filled in on the MAGIC forms and 7.8% were filled in on the phone form, there is a 6.4 fold difference in data entry) Note that this type of information was rarely available at the time the phone call was made. For the category antibiotics, cefazolin, gentamicin, and vancomycin met our inclusion criteria for reliable comparisons. MAGIC reported on 15 patients which the phone call form did not. (48 fold difference, 24% and .5% for MAGIC and phone, respectively) In the blood products category, reliable comparisons could only be done with cell saver blood units (other blood products such as FFP are not given as often). The MAGIC form provides values an order of magnitude more often as the phone call form (12.5 fold difference, 25% and 2% for MAGIC and phone, respectively). Within the category induction medications, three drugs fentanyl, midazolam, and rocuronium met our criteria for comparison. For these, MAGIC provided information on 27 patients that the phone call did not (48 fold difference, 24% and .5% for MAGIC and phone, respectively).

At CTICU admission, the control and experimental groups are again similar in the quantity of information available, given that in the experimental group information is available both through MAGIC and the bedside briefing. These data demonstrate that MAGIC provides at least as much information as is provided at the control bedside briefing (Figure 2). There is, however, a statistical difference in the distribution of these data (Table 1). In one category (lines), MAGIC provided 30% less information than bedside forms (43% and 59% for MAGIC and bedside, respectively). In two categories (blood products and anesthetics), MAGIC provided more information than bedside forms (four times as much data for blood products and included anesthetic medications for 19 patients that the bedside form did not).

In direct comparison of the bedside forms done in both groups, bedside forms in the experimental group contained fewer parameters than the forms in the control group. We presupposed that the bedside briefings were performed under dissimilar conditions, since in the experimental group, neither participant in the briefing was blinded to the data. And, in fact an attenuation of the information conveyed at the bedside was evident (about 20%). The MAGIC briefing, which is more similar to the control bedside condition, roughly conveys similar amount of information (Figure 3 and Table 1). Figure 3 shows the total of responses on the MAGIC form.

MAGIC's error rate is low. Information provided by the MAGIC form is, in the majority of cases, either the same as information provided in the subsequent bedside form or verified as correct against the online patient record. Physicians first scrutinized the bedside and magic forms looking for exact matches or discrepancies. Incomplete reporting of form items were included as discrepancies. These differences were then matched against the anaesthesia operative record (online database) to ascertain which was correct. For example, a drug may be started in the operating room and thus reported in the MAGIC briefing, but discontinued prior to transport to the CTICU. These data were distributed as correct (form items that exactly matched the subsequent bedside form, 37.6%), verified correct (items that were different from the subsequent bedside form but verified correct when compared against the operative anesthesia record, 36%), and error, (1.9%). In only 1.9% of the cases, on average, is information in the MAGIC form *different* both from the bedside form and the online record (Figure 3).

The group of CTICU admitting nurses completed a follow-up interview and questionnaire. While there were minor interface problems, which centered around training in system operation prior to installation, the critical rating of MAGIC was very high. There was 94% approval of the system, 96% accuracy rating and a unanimous wish to have the unit maintained as is in the unit.

Discussion

Our results show that MAGIC is more informative than the phone call and as informative as the bedside report. Since

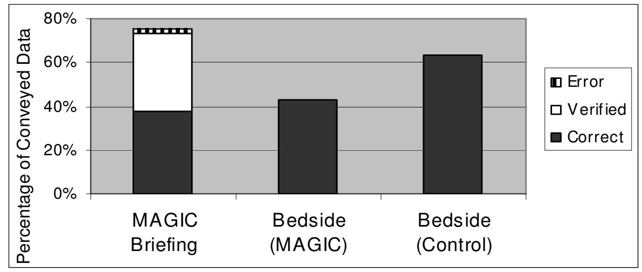

Figure 3: MAGIC/Bedside Form Comparison

MAGIC's report is available earlier than the bedside, it provides advantages over current practice.

First, there is minimal data provided on the operative course and patient status prior to patient arrival without MAGIC. Our present system with all of the additional improvements was robust enough to provide real-time generation of a patient briefing on demand that is accurate and consistent with the bedside report. Clinicians frequently reviewed the MAGIC briefing at shift change and when they were not available at the one time that the bedside briefing was given. We have also shown that MAGIC can be readily integrated into the normal workflow of a highly clinical and medically active environment.

It is important that mandated data items, such as major patient identifiers, are reliably presented in all briefings: the four fields MRN, name, age, weight and height enable accurate identification of the patient. Our results confirm that this is the case. Information about lines is reported in similar quantities across all briefings, but is less accurately presented than demographics. These differences can be explained in part by the imprecise terminology used in the clinical setting. For example, Swan Ganz and Cordis are supposed to be entered separately in the form, but physicians and nurses often will use the two interchangeably and report a value for only one of the two even when both were present. This same phenomenon occurred for IV and peripheral.

Some aspects of our results deserve further investigation. In a large number of cases, information quantity in the experimental bedside form decreased substantially from the MAGIC form (an average decrease of 20%, as shown in Figure 3). However, after the nurse reviewed MAGIC's presentation, the context and content of the bedside briefing was very different from the control group. In the control group, little of the operative information was known prior to the patient's admission to the CTICU. After the bedside briefing, nurses would query the physician for information not reported. All information provided by the physician was assumed to be true and untoward events were not always revealed. In the experimental group, after MAGIC review, nurses questioned information that appeared contrary to the MAGIC briefing and did not query as often for data provided by MAGIC. Thus, bedside briefings in the experimental group were shorter and the verbal interchange appeared limited to clarifying information relayed by the OR physician that conflicted with MAGIC. On several occasions, conflicting information was identified and corrected (e.g., "You said Dobutamine and MAGIC said Dopamine.", "You said no allergies and MAGIC said penicillin."). MAGIC will be correct 98.1% of the time and in future studies, we will examine the impact of reduction in medical error on outcome. Second, while nurses were instructed to record all data, it was particularly difficult to guarantee that nurses would repeat information on the bedside form if that information had already been provided by MAGIC. We also noted that some of the data items missing from the experimental bedside briefing were the ones that must be computed at the time of the briefing (e.g., cardiac output is computed from other data in real-time at the bedside when reported). In most of these cases, the data was actually most useful to the nurses before the patient arrived. Since this data is often less reliable at bedside, in many cases constantly changing over time, we hypothesized that the nurses were not motivated to record this information yet a third time after already recording it on the phone call and MAGIC form. This was verified in informal follow-up interviews; nurses stated that they had already recorded this information. This effect was not pronounced from the phone call to the MAGIC form; since the phone report contained a very limited amount of information, the nurses almost always reentered it.

Related Work

Most evaluations of automated information systems rely on having a reliable "gold" standard to measure the accuracy of the output by comparison [4] [5]. The reliability of the gold standard directly impacts the quality of the evaluation. Expert users in the domain are often used to determine the standard from the system input. Reliability studies are used to find out whether a particular standard qualifies, or whether it needs to be improved by either training or using additional experts [6]. Previous evaluation of MAGIC [2] determined that the system's automatic identification and classification of abnormal surgery events was better in some cases then the expert physicians, and that it could be used as a quality assurance tool to assist the experts. While other researchers study the integration of individual abnormalities to judge the overall severity of patients' conditions [7], our focus is on communicating automatically generated information to the CTICU. In this evaluation, we compare the data MAGIC presents with the amount of information received through current processes. Since the evaluation data is captured by questionnaires and then compared, minimal domain expertise was needed for translating what was written on the forms. Instead, since the system focuses on the presentation of information, the evaluation focused on determining the quantity and quality of what information was presented at specific briefings, including the system's presentation. Similarity between the information in the system and current practice demonstrate the system's benefit.

Conclusions

Our results show that MAGIC provides substantially more information than the phone call (typically 200% more). For blood products and induction medications, MAGIC provided more information than both the phone call and the bedside briefing. With MAGIC, all this information is available before the patient arrives, allowing CTICU staff to make a true assessment of the patient's illness status and prepare for the patient's arrival. In these cases, we show by correlation with the control environment, that MAGIC provides as much information as the bedside briefing. In the experimental setting, the availability of the MAGIC report seemed to influence the bedside briefing report. We suspect that this happened since information was already known.

Acknowledgements

This research is supported in part by NLM Contract R01 LM06593-01 and the Columbia NYSSTF CAT and a gift from Microsoft. Blaine Bell was supported by an IBM Research Graduate Fellowship.

References

[1] Dalal M, Feiner S, McKeown K, Jordan D, Allen B, alSafadi Y. MAGIC: An experimental system for generating multimedia briefings about post-bypass patient status. *Proc AMIA Annu Fall Symp*; 1996 Oct 26–30; Washington, DC. 1996. p. 684–8.

[2] Jordan D, McKeown K, Concepcion K, Feiner S, Hatzivassiloglou V. Generation and evaluation of intraoperative inferences for automated health care briefings on patient status after bypass surgery. J *Am Med Inform Assoc* 2001;8:267–79.

[3] McKeown K, Jordan D, Feiner S, Shaw J, Chen E, Ahmad S, Kushniruk A, Patel V. A study of communication in the cardiac surgery intensive care unit and its implications for automated briefing. *Proc AMIA Symp*; 2000 Nov; Los Angeles, CA. 2000.

[4] Friedman C. *Evaluation methods in medical informatics*. New York: Springer; 1997.

[5] Hripcsak G, Kuperman G, Friedman C, Heitjan D. A reliability study for evaluating information extraction from radiology. *J Am Med Inform Assoc* 1999;37:334–44.

[6] Dunn G. *Design and analysis of reliability studies*. New York: Oxford University Press; 1989.

[7] Knaus W, Wagner D, Draper E. The APACHE III prognostic system: risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* 1991;100:1619–36.

Address for correspondence

Kathleen McKeown
Dept of Computer Science, Columbia University
450 Computer Science Building
1214 Amsterdam Ave., Mailcode: 0401
New York, N.Y. 10027
kathy@cs.columbia.edu
(212) 939-7000