Analysis of Healthcare Quality Indicator using Data Mining and Decision Support System

Young M. Chae[a], Hye S. Kim[a], Kwan C. Tark[b], Hyun J. Park[b], and Seung H. Ho[a]

[a] *Graduate School of Health Science and Management, Yonsei University*
C.P.O. Box 8044, Seoul, Korea
Tel: +82-2-361-5048, Fax: +82-2-392-8996, E-mail: ymchae@yumc.yonsei.ac.kr

[b] *Department of Quality Improvement, Severance Hospital, Yonsei University Medical Center*
C.P.O. Box 8044, Seoul, Korea

Abstract

This study presents an analysis of healthcare quality indicators using data mining for developing quality improvement strategies. Specifically, important factors influencing the inpatient mortality were identified using a decision tree method for data mining based on 8,405 patients who were discharged from the study hospital during the period of December 1, 2000 and January 31, 2001. Important factors for the inpatient mortality were length of stay, disease classes, discharge departments, and age groups. The optimum range of target group in inpatient healthcare quality indicators were identified from the gains chart. In addition, a decision support system was developed to analyze and monitor trends of quality indicators using Visual Basic 6.0. Guidelines and tutorial for quality improvement activities were also included in the system. In the future, other quality indicators should be analyzed to effectively support a hospital-wide continuous quality improvement (CQI) activity and the decision support system should be well integrated with the hospital OCS (Order Communication System) to support concurrent review.

Keywords:

Continuous quality improvement; Data mining; Decision support system

Introduction

An increasing concern with improving the quality of care in various components of the health care system has led to the adoption of quality improvement approaches originally developed for industry. These include 'Total Quality Management' (TQM) [1], an approach which employs process control measures to ensure attainment of defined quality standards, and 'Continuous Quality Improvement' (CQI) [2], a strategy to engage all personnel in an organization in continuously improving quality of products and services.

CQI was originally based on the Quality Assurance (QA) paradigm, which emphasizes monitoring of incidents, mortality and morbidity audits, and hospital infection audits. However, manufacturing industry experiences have shown that QA programs, which focus on end-product evaluation/audit, have little effect on improving quality or decreasing costs [3]. In manufacturing industries, process quality improvement strategies have been proven to be far effective than product oriented quality control programs. At the beginning of the nineties, the emphasis was shifted from the QA paradigm to that of process oriented TQM/CQI, concurrently with the realization of the advantages of the latter throughout the industry.

Process improvement strategies operationalize the Plan-Do-Check-Act (PDCA) process quality management cycle [1]. The outcome targets from the continuous and final quality assurance criteria to be used at quality evaluation-and-improvement checkpoints. The key inputs to the PDCA process are patient assessment/outcome data that are compared to the expected outcome targets and best practice guidelines or protocols. Each set of evaluation results can be used as part of the decision support information for revising the care plan and improving the intervention strategies.

In Korea, QA activity has been launched in 1981 as a part of the Hospital Standardization Project organized by the Korean Hospital Association. Since the Korean Society of Healthcare QA was established in 1994, more comprehensive quality management, evaluation, and research have been implemented. The Hospital Service Evaluation System began in 1995 for an evaluation of CQI activities at the tertiary hospitals initially, but was later expanded to the hospitals with less than 200 beds. Recently, for a more systematic and practical evaluation of hospital quality, researches on re-conceptualization of CQI, development of QI standard and QI indicators, establishment of QI department, and development of QI manual are actively in progress.

But because of inadequate utilization of QI evaluation results and feedback, heavy workloads, and lack of motivation of this endeavor, CQI has not been successfully implemented in most hospitals in Korea. Moreover, a majority of QI activities heavily relied on manual processes such as chart audit. However, manual QI activity without its connection to underlying clinical information produced by hospital information system had been criticized as contributing nothing to quality improvement [4]. Therefore, there is a need for a decision support system that provides patient assessment/outcome information and a clinical pathway to support the PDCA process.

For process quality improvement to be successfully implemented, information on patient care process and the factors influencing quality or treatment outcome must be available at real time for comparison against the desired progress/outcome criteria and development of quality improvement strategies by integrating with the hospital information system. In this study, the factors influencing quality were identified using data mining, and a decision support system for process-oriented CQI based on these factors is another key information for the PDCA process. Data mining is a knowledge discovery method from a large-scale information bank such as a data warehouse. Data mining was used in this study in order to identify pattern or rules about various quality problems or indicators from a large-scale data warehouse. While there were several studies on data mining such as identifying significant factors influencing prenatal care [5] and automatic detection of hereditary syndromes [6], these systems did not explicitly deal with management issues on CQI activities.

Methods

Subjects and Scope

The subjects were 8,405 patients who were discharged from the study hospital during the period of December 1, 2000 and January 31, 2001. Of several quality indicators used in the study hospital, this study focused on the inpatient mortality for the decision tree analysis of the influencing factors for quality. Patient characteristics such as age, sex, discharge department, disease classes, and quality indicators were used in the analysis.

Methods

The decision tree was used in the analysis of the factors influencing inpatient mortality. Decision trees are known as effective classifiers in a variety of domains. In our example, the decision tree categorizes the entire subjects according to whether or not they are likely to have hypertension. Most of the decision tree algorithms use a standard top-down approach to building trees. CHAID (Chi-squared Automatic Interaction Detection) and C5.0 are two popular decision tree inducers, based on the ID3 classification algorithm by Quinlan [7].

A CHAID tree is a decision tree that is constructed by splitting subsets of the space into two or more child nodes repeatedly, beginning with the entire data set. To determine the best split at any node, any allowable pair of categories of the predictor variables is merged until there is no statistically significant difference within the pair with respect to the target variable. This process is repeated until no insignificant pair is found. The resulting set of categories of the predictor variable is the best split with respect to that predictor variable. In this paper, the CHAID algorithm with growing criteria of the likelihood ratio chi-square statistic was used for building the tree and evaluating splits because most of our variables were ordinal and discrete continuous variables. To identify nodes of interest (that is, nodes with a relatively high probability), a gains chart was used. The gains chart shows the nodes sorted by the number of cases in the target category for each node.

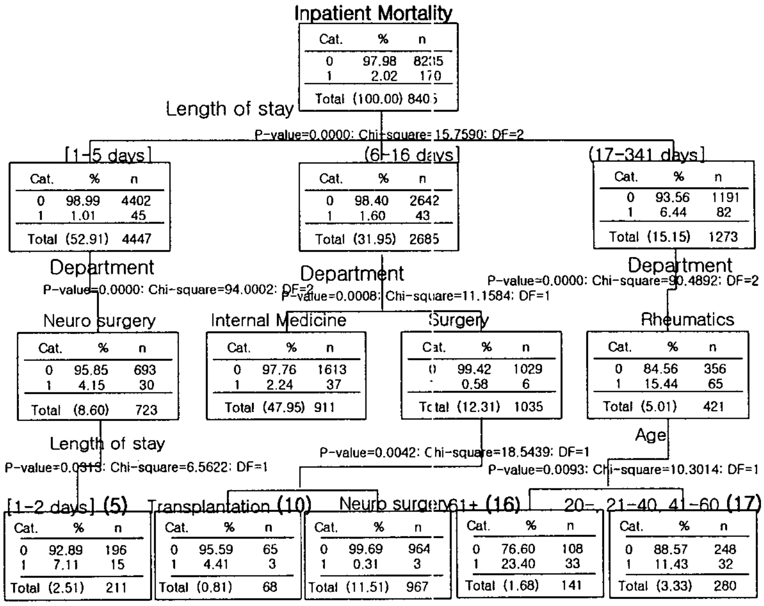
Results

Characteristics of subjects

Among the 8,405 patients, 4,451 (53.0%) were male and 3,954 (47.0%) were female. Patients who were discharged from Internal medicine departments were almost three times (6,109) more than those from the surgery departments (2,296). Patients in the age group of 41-60 had the highest proportion (31.3%). Among all disease classes, neoplasm had the highest proportion (28.8%). Disease classes with the proportion of less than 5% were grouped under miscellaneous. Complete descriptive statistics for the modifiable risk factors are shown in Table 1.

Table 1 – Characteristics of Study Subjects

Characteristics	Value	Frequency	%
Sex	Male	4,451	53.0%
	Female	3,954	47.0%
Discharge Department	1. Internal Medicine	6,109	72.7%
	2. Surgery	2,296	27.3%
Age	1. Under 20	2,103	25.0%
	2. 21-40	1,708	20.3%
	3. 41-60	2,624	31.3%
	4. 61 or older	1,970	23.4%
Disease Class	Neoplasm	2,417	28.8%
	Circulatory	961	11.4%
	Pulmonary	691	8.2%
	Eye and ear	666	7.9%
	Gastrointestinal	486	5.8%
	Muscle & connectivity-tissue	431	5.1%
	Urinary and genital	445	5.3%
	Congenital	426	5.1%
	Miscellaneous	1,883	22.4%
	Total	8,405	100.0%



( ): nodes

Figure 1 - Decision tree for inpatient mortality

Table 1 - Gains chart for inpatient mortality

		Node-by-Node						Cumulative			
Rule	Node	Node: n	Node: (%)	Resp: n	Resp: (%)	Gain (%)	Index (%)	Node: (%)	Resp: (%)	Gain (%)	Index (%)
1	16	141	1.7	33	19.4	23.4	1157.1	1.7	19.4	23.4	1157.1
2	17	280	3.3	32	18.8	11.4	565.0	5.0	38.2	15.4	763.3
3	5	211	2.5	15	8.8	7.1	351.5	7.5	47.1	12.7	625.8
4	10	68	0.8	3	1.8	4.4	218.1	8.3	48.8	11.9	586.2
5	13	413	4.9	17	10.0	4.1	203.5	13.2	58.8	9.0	444.2
6	6	512	6.1	15	8.8	2.9	144.8	19.3	67.7	7.1	349.9
7	8	1650	19.6	37	21.8	2.2	110.9	39.0	89.4	4.6	229.5
8	2	1479	17.6	15	8.8	1.0	50.1	56.6	98.2	3.5	173.7
9	11	967	11.5	3	1.8	0.3	15.3	68.1	100.0	3.0	146.9
10	3	2245	26.7	0	0.0	0.0	0.0	94.8	100.0	2.1	105.5
11	14	439	5.2	0	0.0	0.0	0.0	100.0	100.0	2.0	100.0

Decision tree analysis by CHAID algorithm

The decision tree for inpatient mortality had 17 statistically significant nodes at 5% level (Figure 1). Among 8,405 patients, 170 (2.0%) were inpatient mortality cases. The most significant factor explaining the infant mortality was length of stay (LOS). Mortality rate for the patients with LOS longer than 16 days (6.4%) were almost 6 times higher than the other two LOS groups. Discharge departments were the next significant factors, followed by the age groups.

Each node depicted in the decision tree can be expressed in terms of an 'if-then' rule, as follows:

/*Node 16*/
If (17<LOS<341 and Discharge department=Rheumatics Medicine and Age≥61), then inpatient mortality=23.4%

The gains chart produced by the decision tree can be used for developing quality improvement strategies. As shown in Table 2, there are two parts to the gains chart: node-by-node statistics and cumulative statistics. The gains chart shows the nodes sorted by the percentage of cases in the target category for each node (gain percentage). The first node in the table, node 16 (17≤LOS≤341, age>61, and discharge department was Rheumatics medicine), contains 33 inpatient mortality cases out of 141subjects, or 23.4% inpatient mortality rate (gain %). The Index score or percentage shows how the proportion of inpatient mortality for this particular node compares to the overall proportion of inpatient mortality. For node 16, the Index score was 1157.1%, meaning that the proportion of respondents for this node is about 11 times the inpatient mortality rate for the overall sample.

The cumulative statistics can show us how well we do at finding inpatient mortality cases by taking the best segments of the sample. If we take only the best node (node 16), we reach 19.4% (respondent percentage) of inpatient mortality cases by targeting only 1.7% (node percentage) of the sample. Similarly, if we include the next best node as well (node 17), then we get 38.2% of the inpatient mortality cases from only 5.0% of the sample. Including node 5 increases those values to 47.1% of inpatient mortality cases from 7.5% of the sample. At this stage, we are at the crossover point described above, where we start to see diminishing returns. Note what happens if we include the next node (node 10)--we get 48.8% of inpatient mortality cases, but we must contact 8.3% of the sample to get them.

The gains chart can also provide valuable information about which segments to target and which to avoid. We might base the decision on the number of prospects we want, the desired inpatient mortality rate for the target sample, or the desired proportion of all potential inpatient mortality cases we want to contact. In this example, suppose we want to investigate the cases with an estimated inpatient mortality rate of at least 10%. To achieve this, we would target the first three nodes with a gain percentage greater than 10%, namely, nodes 16, 17, and 5.

Decision support system for quality improvement

In order to support quality improvement activities, decision support system (DSS) was developed using 10 quality indicators including inpatient mortality. Figure 2 depicts information flow of 10 quality indicators. This system has four functions: CQI guidelines, quality review, concurrent review, and tutorial.

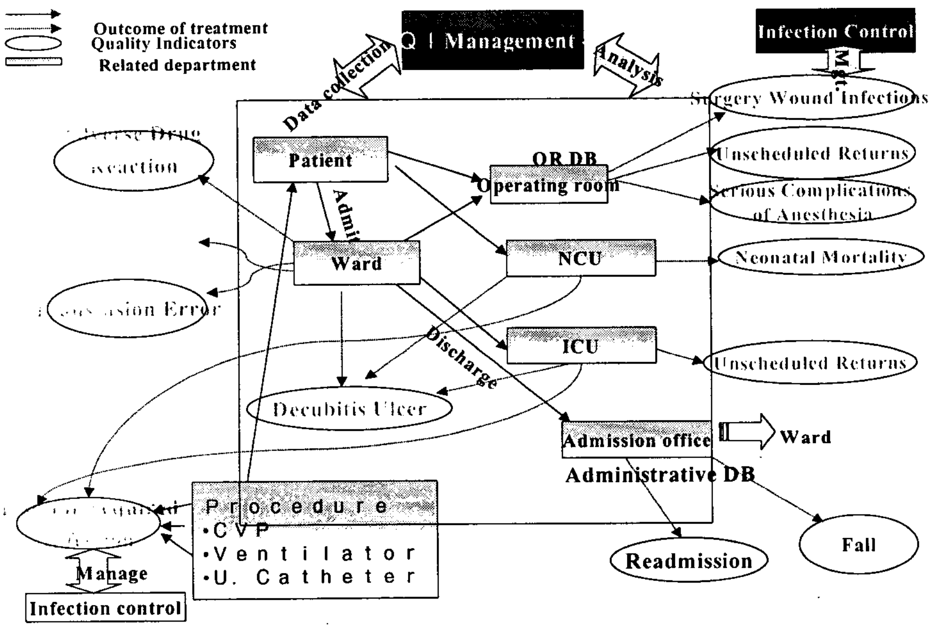


Figure 2 - Flow of quality indicators

CQI guidelines: It provides a definition of each indicator as well as a description of CQI process.
Quality review: It provides information needed for quality improvement activities for the discharged patients such as QI indicators by months, LOS, discharge departments, key clinical information, and patient characteristics. Figure 3 is a sample screen for trend analysis of quality indicators by month.
Concurrent review: It provides information for quality improvement activities while patients are staying at the hospital.
Tutorial: It provides information on clinical practice guideline, clinical pathway, and procedure for CQI activities.

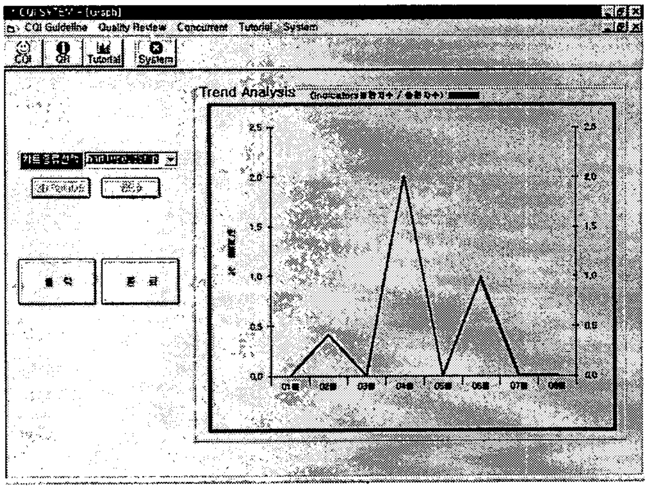


Figure 3 - Screen for trend analysis

Discussion

This study presents an analysis of healthcare quality indicators using data mining for developing quality improvement strategies. Important factors influencing the inpatient mortality were identified using a decision tree method for data mining based on 8,405 patients who were discharged from the study hospital during the period of December 1, 2000 and January 31, 2001. Important factors for the inpatient mortality were length of stay, disease classes, discharge departments, and age groups. The optimum range of target group in inpatient healthcare quality indicators were identified from the gains chart. The cumulative statistics in the gains chart show us how well we do at finding inpatient mortality cases by taking the best segments of the sample.

We demonstrated how the decision tree could be used in developing CQI strategies. While statistical methods such as logistic regression can also be used for identifying important factors influencing quality, it does not provide information about the segment characteristics of such factors that may be useful for CQI activity. The CHAID algorithm provided cumulative statistics that show how well we do at finding the inpatient mortality cases by taking the best segments of the sample. The gains chart also provided valuable information about which segments to target and which to avoid.

In addition, a decision support system was developed to analyze and monitor trends of quality indicators using Visual Basic 6.0. Guidelines and tutorial for quality improvement activities were included in the system. This system has the potential to prevent adverse medical events, improve the quality of care and produce significant savings on healthcare costs. Such system can also provide nurses with valuable process quality and decision support for them to function as effective nurses as well as CQI staff.

There were several literatures related to this study. In the study of the factors influencing unscheduled readmission, which is another quality indicator, Oh [8] found that when age is older, the readmission rate was 1.03 times higher, and when the disease was in the lower risk group, the rate was 0.36 times higher. Furthermore, it was reported that the higher the insufficient discharge schedule points are, the unscheduled readmission rate could be up to 10 times higher. Solberg [9] used CQI program to improve quality of clinical prevention services for chronically ill patients, especially those with diabetes and reported a reduction in unscheduled readmission. Chu [10] reported that computerized clinical pathway and decision support system could improve the clinical process.

In order to apply the DSS for healthcare quality enhancement, the following are recommended:
First, in order for the decision support system to be successful in improving healthcare quality, there must be strong top management support, active participation and effort of the clinical department to obtain system structure and resources, and monitoring and continuous development according to the amount of task process must occur. Second, for a more effective decision support system, hospitals must build a hospital-wide information infrastructure to obtain quality information from various sources and then convert them into useful information to be applied in the decision making process. Third, the DSS should be actively applied to aberrance monitoring, goal achievement monitoring, status progress monitoring, cohort pursuit monitoring, and test monitoring as a continuous clinical work monitoring tool.

There are several limitations of this study. First, the data collection period was only one month, from December 2000 to January 2001, and so it is insufficient to support the decision for the entire quality improvement of healthcare. Second, healthcare quality improvement must be accomplished through prospective method rather than retrospective method, but the data in this study retrospectively used the discharge summary database.

In the future, other quality indicators should be analyzed to effectively support a hospital-wide CQI activity based on comprehensive database. In addition, the decision support system should be well integrated with the hospital OCS to support concurrent review.

References

[1] Deming W.E. 1986. *Out of the Crisis.* Cambridge, MA: MIT Press.

[2] Juran J.M. ed. 1988. *Juran's Quality Control Handbook.* 4th edition. New York, NY: McGraw-Hill.

[3] Sakofsky S. 1996. Solving the Corrective Action Dilemma. Journal for Quality & Participation 19(6): 56-58.

[4] Luttman R.J. 1993. The Critical Path Method Alone Does Nothing to Improve Performance. *Quality Review Bulletin*: 142.

[5] Prather J.C.; Lobach D.F.; Goodwin L.K.; Hales J.W.; Hage M.L.; and Hammond W.E. 1997. Medical data mining: knowledge discovery in c clinical data warehouse. *Proceedings of Annual Conference of AMIA:* 101-105.

[6] Evans S.; Lemon S.; Deters C.A.; Fusaro R.M.; and Lynch H.T. 1997. Automated Detection of Hereditary Syndromes using Data Mining. *Computers and Biomedical Research* 30: 337-348.

[7] Quinlan J.R. 1993. *C4.5: Programs for Machine Learning.* Morgan Kaufmann Publishers, San Mateo, CA.

[8] Oh H.J. 1996. *Factors Influencing Unscheduled Readmission.* Doctoral Dissertation, Yonsei University.

[9] Solberg L.I.; Reger L.A.; Pearson T.L. et al. 1997. Using Continuous Quality Improvement to Improve Diabetes Care In Populations: *The Ideal Model Journal on Quality Improvement* 23(11): 581-592.

[10] Chu S. 2001. Computerized Clinical Pathway as Process Quality Improvement tool. *Proceedings of World Congress on Medical Informatics (MEDINFO):* 1135-1139.