

RESEARCH COMMUNICATION

Prediction of Length of ICU Stay Using Data-mining Techniques: an Example of Old Critically Ill Postoperative Gastric Cancer Patients

Xiao-Chun Zhang¹, Zhi-Dan Zhang^{1*}, De-Sheng Huang²

Abstract

Objective: With the background of aging population in China and advances in clinical medicine, the amount of operations on old patients increases correspondingly, which imposes increasing challenges to critical care medicine and geriatrics. The study was designed to describe information on the length of ICU stay from a single institution experience of old critically ill gastric cancer patients after surgery and the framework of incorporating data-mining techniques into the prediction. **Methods:** A retrospective design was adopted to collect the consecutive data about patients aged 60 or over with a gastric cancer diagnosis after surgery in an adult intensive care unit in a medical university hospital in Shenyang, China, from January 2010 to March 2011. Characteristics of patients and the length their ICU stay were gathered for analysis by univariate and multivariate Cox regression to examine the relationship with potential candidate factors. A regression tree was constructed to predict the length of ICU stay and explore the important indicators. **Results:** Multivariate Cox analysis found that shock and nutrition support need were statistically significant risk factors for prolonged length of ICU stay. Altogether, eight variables entered the regression model, including age, APACHE II score, SOFA score, shock, respiratory system dysfunction, circulation system dysfunction, diabetes and nutrition support need. The regression tree indicated comorbidity of two or more kinds of shock as the most important factor for prolonged length of ICU stay in the studied sample. **Conclusions:** Comorbidity of two or more kinds of shock is the most important factor of length of ICU stay in the studied sample. Since there are differences of ICU patient characteristics between wards and hospitals, consideration of the data-mining technique should be given by the intensivists as a length of ICU stay prediction tool.

Keywords: Gastric cancer - intensive care - data mining - China

Asian Pacific J Cancer Prev, 13, 97-101

Introduction

Gastric cancer is one of the most common cancers and a leading cause of cancer death in China and worldwide (Kamangar et al., 2006; Yang, 2006). With the background of aging population in China and advances in clinical medicine, the amount of operations on old patients increases correspondingly. This situation imposes increasing challenges to critical care medicine and geriatrics, due to the fact that more complications and high-risk factors may be prevalent among the old patients.

Prediction for the length of stay in intensive care unit (ICU) has gained more attention in the field of critical care medicine. Severity scores, like the Acute Physiology and Chronic Health Evaluation (APACHE) II has been widely used to predict the length of ICU stay (Abbott et al., 1991; Suistomaa et al., 2002; Dossett et al., 2009). However, they may not be adequate predictors of length of stay, because APACHE II is based on data collected during the first 24h

of ICU treatment, and the most severely ill patients may die after a short length of stay and those who need only post-surgical observation are also transferred to general wards early.

Baseline information of the length of ICU stay can help to make appropriate decisions about the number of ICU beds, it can also result in better communications with patients and relatives because it is a key determinant of the costs of intensive care. Thus, the development of a framework to predicate the length of ICU stay is needed and will be useful in supporting clinical judgment in decision making for individual patients.

In January 2009, critical care medicine was just accredited as an independent subspecialty of clinical medicine by the Ministry of Health of the People's Republic of China. Critical care research in mainland China is still in its infancy (Du et al., 2010). The present study was designed to describe the information of the length of ICU stay from a single institution experience of

¹Department of Intensive Care Unit, The First Affiliated Hospital, ²Department of Mathematics, College of Basic Medical Sciences, China Medical University, Shenyang, China *For correspondence: icuzhangzd@gmail.com

old critically ill gastric cancer patients after surgery and the framework of incorporating data-mining techniques with the prediction of length of stay in ICU.

Materials and Methods

Study unit and data collection

The setting for this study was a 28-bed multidisciplinary adult intensive care unit in a 2250-bed medical university hospital in Shenyang, Liaoning province, northeastern China. For this hospital, annually clinic visits are around 2 million person-visits, and annually hospital admissions are around 80 thousand person-admissions. The intensive care unit accommodates approximately 800 admissions of critically ill patients per year. Almost all patients need fulltime monitoring and mechanical ventilation. This unit is a part of National Key Discipline of General Surgery, it also manages two provincial centers, one is Liaoning Provincial Center of Medical Quality Control and Improvement, the other is Liaoning Provincial Center of Disaster and Critical Care Medicine.

In this unit, concerned patient data has been gathered and stored in an ICU patient information system since 2008, and with enrichment from the beginning of 2010. A full-time data entry clerk is in charge of daily data collection, data entry and database maintenance. Thus, retrospective design was adopted to collect the consecutive data about gastric cancer patients 60 years of age or older after surgery recorded in the patient information system from January 1st 2010 to March 31st 2011.

Characteristics of patients and the length their ICU stay were gathered for analysis. Length of ICU stay was measured using the interval between the day of ICU admission and ICU discharge on a trisection scale of calendar days (morning, evening and night).

Data analysis

Cox regression

Univariate Cox regression was performed first to examine the relationship between the potential candidate factors and the length of ICU stay (we treated it as time variable for survival analysis), then the impact of multiple covariates was assessed in the multivariate model. The above analyses were conducted by using the Statistical Product and Service Solutions (SPSS 12.0 for windows, SPSS Inc., Chicago, IL, USA).

Construction of regression tree

The regression tree was constructed to predict the length of ICU stay and explore the important indicators. The main purpose of regression tree was to produce a tree-structured prediction rule. The construction of classification and regression tree model has been well described in previous studies (Guan et al., 2008; Lapinsky et al., 2008; Berney et al., 2011). Major steps were as follows, (1) Binary partitioning. According to whether $X \in A$, where X was a variable and A was a constant, the response to the questions was yes or no, then the prediction space was partitioned into two parts. (2) Goodness-of-split criteria for choosing the best split. (3) Tree Pruning. Minimal cost-complexity pruning was carried out in this

Table 1. Characteristics and Length of ICU Stay of Old Critically Ill Gastric Cancer Patients after Surgery

Variables	N	%
Length of ICU stay, days (median, range)	3.3 (1.3-22.7)	
Age, years (median, range)	78 (60-87)	
Gender		
Male	75	78.1%
Female	21	21.9%
ICU Admission Source		
Operation room	89	92.7%
Emergency room	7	7.3%
APACHE II score (Median, Range)	8 (3-24)	
SOFA score (Median, Range)	2 (0-16)	
Acute comorbidities		
Infection	7	7.3%
Sepsis	4	4.2%
Shock	11	11.5%
Organ dysfunction		
Respiratory system	42	43.8%
Circulation system	72	75.0%
Renal system	8	8.3%
Hematological system	5	5.2%
Invasive manipulation		
Central venous catheterization	86	89.6%
Indwelling urethral catheterization	91	94.8%
Diabetes	15	15.6%
Nutrition support need	45	46.9%

study. This method relied upon a complexity parameter, denoted α , which was gradually increased during the pruning process. (4) Cross validation. The tree was computed from the learning sample, and its predictive accuracy was examined by applying it to predict the class membership in the test sample. Software R 2.3.1 (<http://www.R-project.org>) was adopted to structure the classification and regression trees.

Ethical review

The present study was approved by research institutional review board of The First Affiliated Hospital of China Medical University, in accordance with the standards of the Declaration of Helsinki. All the data collected were treated confidentially, the names and identity codes of patients were removed before data analysis.

Results

From January 1st 2010 to March 31st 2011, there were 979 consecutive admissions in the study unit. More than two thirds of the patients were admitted to ICU from the same hospital, mostly from operation room. Ninety-six old critically ill gastric cancer patients after surgery met the inclusion criteria, with a median age of 78 (60-87) years. Among them, 7 patients were from emergency room due to acute peritonitis or acute intestinal obstruction caused by primary gastric cancer. Median APACHE II was 8 (3-24), and 78.1% of patients were male. The median length of ICU stay was 3.3 days, with the range from 1.3 to 22.7 days (Table 1).

Univariate COX analysis indicated that APACHE II score, SOFA score, shock and necessary nutrition support

Table 2. Univariate Cox Analysis of the Length of ICU Stay of Old Critically Ill Gastric Cancer Patients after Surgery

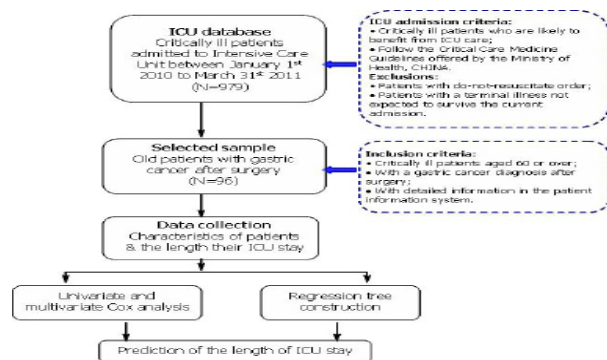
Variables	Coefficients*	Standard Error	Wald χ^2	P value	Exp(B)	95% CI for Exp(B)	
						Lower	Upper
Age	0	0.02	0.01	0.92	1	0.97	1.04
Gender	-0.1	0.25	0.16	0.69	0.9	0.56	1.48
APACHE II score	-0.06	0.03	4.67	0.03	0.94	0.89	0.99
SOFA score	-0.11	0.05	4.67	0.03	0.9	0.82	0.99
Acute comorbidities							
Infection	-0.92	0.48	3.76	0.05	0.4	0.16	1.01
Sepsis	-0.48	0.59	0.66	0.42	0.62	0.2	1.96
Shock	-0.46	0.2	5.31	0.02	0.63	0.43	0.93
Organ dysfunction							
Respiratory system	-0.14	0.21	0.47	0.5	0.87	0.57	1.31
Circulation system	0.44	0.26	2.88	0.09	1.55	0.94	2.55
Renal system	-0.14	0.37	0.15	0.7	0.87	0.42	1.8
Hematological system	0.05	0.42	0.01	0.91	1.05	0.46	2.41
Invasive manipulation							
Central venous catheterization	-0.38	0.34	1.25	0.26	0.69	0.36	1.33
Indwelling urethral catheterization	-0.19	0.46	0.17	0.68	0.83	0.34	2.05
Diabetes	-0.12	0.28	0.17	0.68	0.89	0.51	1.55
Nutrition support need	-0.72	0.22	11.1	<0.01	0.49	0.32	0.74

*The negative coefficient indicates the variable is a risk factor for prolonged length of ICU stay

Table 3. Multivariate Cox Analysis of the Length of ICU Stay of Old Critically Ill Gastric Cancer Patients after Surgery

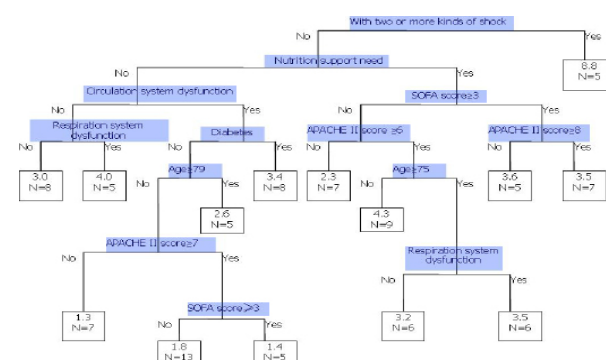
	Coefficients*	Standard Error	Wald χ^2	P value	Exp(B)	95%CI for Exp(B)		
						Lower	Upper	
Step 1	Nutrition support need	-0.72	0.22	11.1	<0.01	0.49	0.32	0.74
Step 2	Circulation system dysfunction	0.64	0.26	5.97	0.02	1.89	1.14	3.16
	Nutrition support need	-0.85	0.22	14.57	<0.01	0.43	0.28	0.66
Step 3	Circulation system dysfunction	0.69	0.27	6.73	0.01	1.99	1.18	3.35
	Shock	-0.44	0.21	4.23	0.04	0.65	0.43	0.98
Step 4	Nutrition support need	-0.76	0.22	11.75	<0.01	0.47	0.3	0.72
	Circulation system dysfunction	0.87	0.28	9.6	<0.01	2.37	1.37	4.1
	Diabetes	-0.57	0.3	3.57	0.06	0.57	0.32	1.02
	Shock	-0.44	0.21	4.37	0.04	0.65	0.43	0.97
	Nutrition support need	-0.87	0.23	13.96	<0.01	0.42	0.27	0.66

*The negative coefficient indicates the variable is a risk factor for prolonged length of ICU stay

**Figure 1. Schematic Illustration of the Proposed Study on the Length of ICU Stay for Old Critically Ill Gastric Cancer Patients after Surgery (Shenyang, China)**

were related to the length of ICU stay in the enrolled patients. Multivariate Cox analysis found that shock and nutrition support need were the risk factors for prolonged length of ICU stay with statistical significance.

The multivariate regression tree model is shown in Figure 2. The importance of variables can be judged according to its location in the tree (the closer to the root, the more important) and the times of appearances

**Figure 2. Multivariate Regression Tree Model for the Length of ICU Stay for Old Critically Ill Gastric Cancer Patients after Surgery (Shenyang, China)**

(the more frequent appearances, the more important) in the model. In the present study, the root of regression tree was divided into two branches according to whether the patient was with the comorbidity of two or more kinds of shock (e.g., Cardiogenic shock, Septic shock, Hypovolemic shock, Anaphylactic shock, Neurogenic shock and etc.). Altogether, eight variables entered the regression model, including age, APACHE II score, SOFA

score, shock, respiratory system dysfunction, circulation system dysfunction, diabetes and nutrition support need. Both age and APACHE II score appeared the most 3 times in the regression tree.

Discussion

In the context that the number of patients after surgery who require intensive care is increasing rapidly in China, there is growing concern about the length of ICU stay. The present study presents a framework for estimating the length of ICU stay for old critically ill gastric cancer patients after surgery. To the best of our knowledge, it is the first study focusing on the length of ICU stay related data-mining in the study area.

The length of ICU stay contributes a large proportion of financial burden during the whole stay in the hospital (Pittoni & Scatto, 2009; Beenen et al., 2011). The gap does exist between the patients families' optimistic expectation and the ICU clinicians' professional judgment and choice of treatment. This study provides the information about the distribution of the length of ICU stay and its related determinants, paying special attention to those elderly patients suffering from multiple co-morbidities. The median age was 78 years in the sample, which indicated that due to aging and the improvement of new technology and various kinds of co-morbidities, more and more old patients need intensive care. Detailed characteristic description of the ICU patients can help the hospital manager to know how to maximize the use of ICU resources and how to better allocate the resources in an optimistic way (Weissman, 1997; Stricker et al., 2003; Terra, 2007).

Univariate survival analysis indicated that APACHE II score, SOFA score, shock and necessary nutrition support were important indicators for the length of ICU stay, while it could not provide the threshold of APACHE II score, SOFA score. There is also no uniform definition for a prolonged ICU stay, the ICU needs to develop a suitable model to predict the length of ICU stay with the available information. The regression tree adopted in our study suggested that among the selected subgroup of ICU patients, it was possible for the ICU clinicians to identify some patients with potentially prolonged ICU stay with several simple variables. It is necessary to broad the collaborative efforts to extend from the patient care arena into the realms of education, research, and administration (Miccolo & Spanier, 1993). Carr DD has also indicated that the collaborative partnerships in critical care, can enhances the organizations' mission to deliver quality patient-centered care with the main focus on. (1) minimization of inpatient transitions, (2) reduction of cost by decreasing the length of stay, (3) promotion of patient and family satisfaction through efforts of advocacy, and (4) enhanced discharge planning (Carr, 2009).

The authors recognize that there are some limitations inherent in this retrospective analysis. First, the generalization was limited because our results were derived from the experience of one ICU in a single hospital, and therefore it is not representative of the region or country as a whole. We do not recommend different

hospitals to use the same model to predict the length of ICU stay, we just want to alert the ICU clinicians to keep these candidate risk factors in mind and show them the potential framework of data-mining which might optimize the resource use. Second, ICU stay was measured using the calendar days (in the three parts of morning, afternoon, night), several studies have proved the superior accuracy of exact ICU stay in minutes compared to measurement using calendar days (Marik & Hedman, 2000; Render et al., 2005). Personal digital assistants (PDAs) (Faulk & Savitz, 2009; Zalon et al., 2010) are being introduced in the studied hospital, it would be available for the doctors and nurses to get the accurate real-time information about the length of ICU stay for future research. The third and also the most important limitation is the complexity of the available and unknown risk factors of the length of ICU stay (Rapoport et al., 2003). With the rapid improvement of the treatment and nursing, more details about the cooperation or interaction of the varying and emerging factors should be explored and thus models should be calibrated in parallel to provide more reference for the selection of candidate factors.

In conclusion, the comorbidity of two or more kinds of shock is the most important factor of length of ICU stay in the studied sample. There are differences of ICU patient characteristics between wards and hospitals, consideration to the data-mining technique should be given by the intensivists as a length of ICU stay prediction tool.

Acknowledgements

This study was partially supported by the National Natural Science Foundation of China (No. 71073175) to DH. The author(s) declare that they have no competing interests.

References

- Abbott RR, Setter M, Chan S, et al (1991). APACHE II: prediction of outcome of 451 ICU oncology admissions in a community hospital. *Ann Oncol*, **2**, 571-4.
- Beenen E, Brown L, Connor S (2011). A comparison of the hospital costs of open vs. minimally invasive surgical management of necrotizing pancreatitis. *HPB (Oxford)*, **13**, 178-84.
- Berney SC, Gordon IR, Opdam HI, et al (2011). A classification and regression tree to assist clinical decision making in airway management for patients with cervical spinal cord injury. *Spinal Cord*, **49**, 244-50.
- Carr DD (2009). Building collaborative partnerships in critical care: the RN case manager/social work dyad in critical care. *Prof Case Manag*, **14**, 121-32, quiz 133-34.
- Dossett LA, Redhage LA, Sawyer RG, et al (2009). Revisiting the validity of APACHE II in the trauma ICU: improved risk stratification in critically injured adults. *Injury*, **40**, 993-8.
- Du B, Xi X, Chen D, et al (2010). Clinical review: critical care medicine in mainland China. *Crit Care*, **14**, 206.
- Faulk JF, Savitz LA (2009). Intensive care nurses' interest in clinical personal digital assistants. *Crit Care Nurse*, **29**, 58-64.
- Guan P, Huang D, Guo J, et al (2008). Bacillary dysentery and meteorological factors in northeastern China: a historical

- review based on classification and regression trees. *Jpn J Infect Dis*, **61**, 356-60.
- Kamangar F, Dores GM, Anderson WF (2006). Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. *J Clin Oncol*, **24**, 2137-50.
- Lapinsky SE, Holt D, Hallett D, et al (2008). Survey of information technology in Intensive Care Units in Ontario, Canada. *BMC Med Inform Decis Mak*, **8**, 5.
- Marik PE, Hedman L (2000). What's in a day? Determining intensive care unit length of stay. *Crit Care Med*, **28**, 2090-3.
- Miccolo MA, Spanier AH (1993). Critical care management in the 1990s. Making collaborative practice work. *Crit Care Clin*, **9**, 443-53.
- Pittoni GM, Scatto A (2009). Economics and outcome in the intensive care unit. *Curr Opin Anaesthesiol*, **22**, 232-6.
- Rapoport J, Teres D, Zhao Y, et al (2003). Length of stay data as a guide to hospital economic performance for ICU patients. *Med Care*, **41**, 386-97.
- Render ML, Kim HM, Deddens J, et al (2005). Variation in outcomes in Veterans Affairs intensive care units with a computerized severity measure. *Crit Care Med*, **33**, 930-9.
- Stricker K, Rothen HU, Takala J (2003). Resource use in the ICU: short- vs. long-term patients. *Acta Anaesthesiol Scand*, **47**, 508-15.
- Suistomaa M, Niskanen M, Kari A, et al (2002). Customized prediction models based on APACHE II and SAPS II scores in patients with prolonged length of stay in the ICU. *Intensive Care Med*, **28**, 479-85.
- Terra SM (2007). An evidence-based approach to case management model selection for an acute care facility: is there really a preferred model? *Prof Case Manag*, **12**, 147-57, quiz 158-59.
- Weissman C (1997). Analyzing intensive care unit length of stay data: problems and possible solutions. *Crit Care Med*, **25**, 1594-600.
- Yang L (2006). Incidence and mortality of gastric cancer in China. *World J Gastroenterol*, **12**, 17-20.
- Zalon ML, Sandhaus S, Valenti D, et al (2010). Using PDAs to detect cognitive change in the hospitalized elderly patient. *Appl Nurs Res*, **23**, e21-7.