PREDICTIVE MODEL FOR SURVIVAL AT THE CONCLUSION OF A DAMAGE CONTROL LAPAROTOMY

Noriaki Aoki, MD[1,2], Matthew J. Wall, MD[3], Janez Demsar[4], Blaz Zupan, PhD[1,4], Thomas Granchi MD[3], Martin A. Schreiber MD[3], John B. Holcomb MD[3,5], Mike Byrne, MD[3,5], Kathleen R. Liscum, MD[3], Grady Goodwin[3], J. Robert Beck, MD[1], Kenneth L. Mattox MD[3]

1) Information Technology, Baylor College of Medicine, Houston, TX
2) Department of General Medicine and Clinical Epidemiology, Kyoto University Hospital, Kyoto, Japan
3) Department of Surgery, Ben-Taub General Hospital, Baylor College of Medicine, Houston, TX
4) Faculty of Computer and Information Science, University of Ljubljana, Slovenia
5) Joint Trauma Training Center, Ben-Taub General Hospital, Baylor College of Medicine, Houston, TX

Address for correspondence:
Noriaki Aoki
Assistant Professor of Medicine / Information Technology
HAM / TMC Library
1133 M.D. Anderson Blvd., Houston, Texas 77030.
TEL: 713-799-7168; FAX: 713-790-7052
E-mail: naoki@bcm.tmc.edu

Word counts: Abstract: 148 words; body text: 2,062 words

Key Words

1. Damage control surgery
2. Trauma
3. Data mining
4. Logistic regression analysis
5. Prognostic modeling

SUMMARY

This study investigated 174 parameters from 68 trauma patients who underwent damage control maneuvers in an attempt to develop a prognostic model capable of predicting mortality at the end of the first operation. Both statistical methods (univariate and logistic regression analysis) and a data mining approach (feature subset selection and decision tree induction) identified the same prognostic factors: pH at initial ICU admission and worst partial thromboplastin time (PTT) from hospital admission to ICU admission as predictive of mortality. In a triage situation due to resource limitations, these results may be useful to detect severely injured patients who will not survive prolonged or recurrent resuscitation in the intensive care unit and may affect the optimal allocation of limited medical resources.

ABSTRACT

Background: We employed modern statistical and data mining methods to model survival based on preoperative and intra-operative parameters, for patients undergoing damage control surgery.

Methods: One hundred seventy-four parameters were collected from 68 damage control patients in prehospital, emergency center, operating room, and intensive care unit (ICU) settings. Data were analyzed with logistic regression and data mining. Outcomes were survival and death after the initial operation.

Results: Overall mortality was 66.2%. Logistic regression identified pH at initial ICU admission (odds ratio: 4.4) and worst partial thromboplastin time (PTT) from hospital admission to ICU admission (odds ratio: 9.4) as significant. Data mining selected the same factors, and generated a simple algorithm for patient classification. Model accuracy was 83%.

Conclusions: Inability to correct pH at the conclusion of initial damage control laparotomy and the worst PTT can be predictive of death. These factors may be useful to identify patients with a high risk of mortality.

INTRODUCTION

The damage control approach arose from a need to develop effective strategies for salvaging the most critically injured trauma patients. The basic concept of damage control is to avoid extensive procedures on unstable patients, stabilize potentially fatal problems at initial operation, and apply staged surgery after successful initial resuscitation. [1, 2] Several problems remain in the application of the damage control philosophy. One of the most important, but difficult issues is determining eligible patients for damage control. Furthermore, damage control requires a massive investment of personnel, time, and resources in a small group of critically injured patients who carry a mortality rate in excess of 50% even under the best of circumstances. [3] From the viewpoint of resource allocation, it may also be advantageous to predict patient outcome to prevent futile use of limited resources. This is particularly true in a triage situation when resources are by definition limited. Toward this end, several injury-scoring systems have been reported, but they are problematic. [4-7] One of the most important limitations in current trauma scoring systems is that they represent one-time (usually at the end of the hospitalization) evaluation or classification, which eliminates their usefulness for real-time decision support, especially early in the hospital course. Furthermore, general injury severity models do not adequately address the data in severely injured damage control patients.

The purpose of this study is 1) to identify the risk factors associated with mortality from large sets of physiologic and laboratory data at four different treatment phases and 2) to develop corresponding prognostic models for damage control patients.

METHODS

Subjects

Sixty-eight patients who required damage control surgery at the Ben Taub General Hospital during the period from January 1994 to June 1997 comprised the subject population. A total of 174 variables, including patient demographic and physiologic findings, laboratory results and therapeutic procedures at four different treatment phases (prehospital, emergency center, operating room and intensive care unit (ICU)) were collected. Arterial blood gases were adjusted for patient body temperature. Prediction models were developed using in-hospital survival as the outcome of interest. This retrospective data collection study was performed within the guidelines of the Baylor Institutional Review Board.

Prediction Model Development with Statistical Methods

A univariate analysis was performed for all variables with chi-square or Fisher's exact tests as the first step for data selection. Statistically significant variables and others selected by a panel of experienced clinicians as potentially important variables for trauma outcomes were included in further multivariate analysis. Log-odds ratio of each parameter was used to assess whether a continuous or categorical variable would be preferred in the logistic regression model. For categorical values, one or two cut-offs were set based on pathophysiologic considerations or results of log-odds analysis. Logistic regression models were used to control for confounding factors and to assess interactions between variables. In all tests, p values of 0.05 or less were considered statistically significant. Analyses were done using Excel 98 (Microsoft Inc. 1998) and SPSS statistical package 7.5 (SPSS Inc. 1997).

Prediction Model Development with Data Mining Techniques

An entropy and minimal description length-based algorithm [8] was used that tries to categorize all numerical variables by finding suitable cutoff points. In a case when the algorithm was not able to find any suitable cutoff points, the variable was considered uninformative about the outcome and was removed. Remaining variables were then further filtered using the ReliefF algorithm [9] to exclude irrelevant factors. ReliefF estimates relevance of a target variable “x” by taking patients individually, and for each patient finding a subset of other patients who are most similar in all respects to the index patient except for their result with variable “x”. ReliefF calculates the likelihood that difference in “x” is associated with different outcomes (i.e., survival and death), and eliminates variables stepwise based on the weakness of their association with the outcome of interest.

The selected factors were used to construct a predictive model using standard decision tree induction. [10, 11] Decision tree induction is a recursive partitioning algorithm which classifies patients into two or more subgroups until creating sets with all patients corresponding to the same outcome, survival or death in this analysis. The partition criteria are functions computed from predictor variables. To avoid over-fitting, we used a simple pruning criterion that stops the induction when the sample size for a node falls under five patients or when 90% of a subgroup has the same outcome. The particular data mining approach used in this study is described in detail in our previous paper. [12]

Evaluation of Prognostic Model

We used 1) classification accuracy, 2) sensitivity and specificity and 3) area under the receiver operating characteristic (ROC) curve to evaluate the accuracy of each model. These evaluations were estimated using 10-fold cross validation, [13] which first divides the data into ten independent sets of approximately equal size and distribution of outcomes. Each single set is used for testing the model developed from the remaining nine sets, therefore, ten different models are tested as independent experiments. The statistics for each method are then calculated as an average of ten experiments. For example, if 100 subjects form a data set, 10-fold cross validation would create 10 experimental models, each with the data from 90 subjects used to predict the remaining 10.

Classification accuracy measures the proportion of correctly classified patients and estimates the probability of correct prediction using the model. Sensitivity (true-positive rate) is defined as the proportion of people with the event (e.g. death) predicted to have the event. Specificity (true-negative rate) is the proportion of survivors correctly predicted.

The ROC curve represents the relationship between sensitivity and specificity, by plotting the true-positive rate (sensitivity) against the false-positive rate (1-specificity) as the cutoff level of the model varies. The area under the ROC curve (AUC) is based on a non-parametric statistical sign test to compare the probability of events between pairs of patients who have the event and those who do not. AUC may be interpreted as the probability that given any two subjects, one who dies and one who survives, the model would assign a higher probability of death to the one who dies. It is a measure of overall classification performance of a diagnostic test or prognostic model. AUCs were calculated with the methods described by Hanley et al [14, 15].

RESULTS

Overall mortality for 68 patients was 66.2% (45/68). Median survival time of these patients was 0 days, that is, over half died before, during, or immediately after initial damage control. Table 1 contains mortality of each mechanism and definitive diagnoses in the population. Of 174 variables, 82 were excluded because the data were absent in more than 50% of patients. Therefore, 92 variables were used for further analysis. Univariate analysis found 10 statistically significant variables related to three domains: acidosis, coagulopathy and hypotension/blood loss (Table 2). The most significant risk factor in each domain and clinically relevant factors were used for final logistic regression analysis.

Only 48 patients were eligible for final analysis who completed damage control surgery and were admitted to the ICU. Seventeen patients died before the end of the first damage control surgery, and thus were not appropriate for a model that would predict success of further resuscitative efforts. Three were excluded due to lack of partial thromboplastin time (PTT) data. pH was used as a continuous variable in logistic regression because the log-odds ratio showed a linear relationship. In contrast, PTT showed best predictability with a cut-off point of 80 seconds (78.7 seconds using data mining). Table 3 shows the final logistic model including two factors: pH at initial ICU admission and worst PTT from hospital admission to ICU admission. The odds ratio of pH is 4.43 for each 0.1 unit decrease (e.g., predicted risk of mortality is 4.43 times higher for patients with pH of 7.2 compared to those with pH of 7.3, and 4.43 times higher for pH 7.3 versus pH 7.4). The odds ratio of patients with a higher PTT (? 80 seconds) is 9.4 relative to those with a lower PTT (< 80 seconds).

The entropy and minimal description length-based algorithm and the ReliefF algorithm determined that only 12 parameters were related to mortality (Table 2). The decision tree built from these 12 parameters selected only two parameters, exactly the same as selected by the logistic regression model (Figure 1). The tree predicts mortality if pH at initial ICU admission is less than or equal to 7.2. Of 14 such patients in the population, all patients died. In contrast, outcome is better if the pH at initial admission to the ICU is greater than 7.33. Of 16 such patients, only 2 patients (12.5%) died. PTT is used for prediction if pH at initial ICU admission is higher than 7.2 and lower than or equal to 7.33. In this subset, patients whose PTT is less than or equal to 78.7 seconds has lower mortality than those with a PTT greater than 78.7 seconds.

Classification accuracy, sensitivity and specificity of each model are listed in Table 4. Both models were statistically significant (p=0.0001). Overall classification accuracy is 83%. The AUCs of the logistic model and decision tree model are 87.7% and 86.7%, respectively.

DISCUSSION

Our results suggest that pH at initial ICU admission and the worst PTT from hospital to ICU admission are significantly correlated with mortality of patients who undergo damage control surgery and survive to the end of the initial procedure. The prediction models developed with both the logistic regression and the decision tree model detected similar prognostic factors and have substantial predictability, although the results were based on a small number of patients.

The statistical univariate method detected 10 significant risk factors and data mining techniques identified 12 informative factors, including all 10 selected by univariate analysis. Of these parameters, 4 were the worst values of certain variables, such as worst blood pressure or pH from hospital admission to initial ICU admission. Some variables might be associated with mortality only in a particular phase of a treatment, such as pH at initial admission to the ICU or PaCO2 in the operating room. This suggests that some risk factors might be appropriately used at particular decision points.

This process can be called "hypothesis development", as compared to "hypothesis testing" in the standard statistical method. Exploratory data analysis is useful to identify important factors and to uncover hidden relationships from numerous data sets. In this analysis, we used data mining to more narrowly explore the relationships between many variables and mortality, and detected the same two prognostic factors with a standard statistical method (logistic regression model). Even though the decision tree modeling has a completely different analytical process to select significant prognostic factors, a resulting set of only two parameters used in the classification tree, proved remarkably similar to the set of variables found by standard statistical methods.

The decision tree model can be viewed as a simple algorithm because of its transparency, enabling physicians to understand the decision process intuitively. In contrast, the logistic regression model provides more information for each factor in terms of odds ratio and its 95% confidence interval, although physicians would have to perform an obscure calculation to obtain an estimated outcome. We believe that the two approaches may be used in combination, where data mining is used to propose a subset of most predictive factors by examining a larger set of data sets. This suggests that data mining could contribute to the development of reliable evidence in terms of determining important relationships from large numbers of variables, or detecting hypotheses for further hypothesis testing with statistical analysis.

Our analyses identified two independent prognostic factors: pH at initial ICU admission and the worst PTT from hospital to ICU admission (Table 3). These two were selected from several indicators of metabolic acidosis and coagulopathy listed in Table 2. Finally, any factors related to hypotension or bleeding were not employed as independent prognostic factors.

The pH indicates metabolic acidosis, usually resulting from inadequate tissue perfusion. [1, 16] Data mining determined two important cut-off points in pH at initial admission to the ICU: 7.2 and 7.33. The outcome was extremely poor if patients had severe acidosis (pH ? 7.2) in spite of massive resuscitative efforts, which might be related to the inability to compensate for overall metabolic acidosis, continued massive hemorrhage, or both. This is consistent with irreversible shock. In contrast, almost all patients whose pH at initial ICU admission were greater than 7.33 survived. Another risk factor detected here, coagulopathy, results from several causes such as massive transfusion and hemodilution. [1, 16]

Three major limitations of this study are sample size, the usefulness of a classification accuracy of 83% in a population with 66% mortality, and the "surprise value" of the results. This is a relatively large clinical study of the utility of data mining, nevertheless only 48 patients were used in the final model. The possible weakness in sample size was overcome by two facts: cross-validation consistently selected the two most relevant parameters, and two essentially different techniques picked the same variables as best predictors, generating two different types of statistically significant models. Also, data mining is relatively new in medicine, and a recent book illustrates that useful results can be obtained from relatively small data sets [17].

In a study population wherein two-thirds have the outcome of interest is an 83% classification accuracy important? Here the AUC helps to interpret the results. In clinical prognosis, an area under the ROC curve, of 87% is quite high (and itself statistically significant). While "statistically significant" is not necessarily mean "clinically important", the clinician must ultimately determine what a medically useful result must be. In triage scenarios where resources are limited, decision support such as these outcomes may assist the clinical decisions required in the midst of patients resuscitation. Perfect prediction is unattainable, but the results of this study suggest that an improvement of classification accuracy above 90% will require the identification of parameters currently unknown.

The known variables identified, metabolic acidosis and coagulopathy, may not surprise the experienced trauma surgeon. These are well-known indicators of poor prognosis, but they are known based on empiric observation. This study, with no a priori bias, selected these two parameters from 92 and showed them to be highly predictive in combination.

In summary, this preliminary work detected two important prognostic factors for severely injured patients undergoing initial damage control surgery who were admitted to the ICU. Two separate methods selected the same factors as most predictive of mortality. As more experience with damage control accumulates, additional parameters may be elucidated that raise the predictability above the clinically useful threshold.

REFERENCES

1. Hirshberg A, Mattox KL. Planned reoperation for severe trauma. Ann Surg 1995; 222:3-8.
2. Mattox KL. Introduction, background, and future projections of damage control surgery. Surg Clin North Am 1997; 77:753-9.
3. Granchi TS, Liscum KR. The logistics of damage control. Surg Clin North Am 1997; 77:921-928.
4. Zoltie N, de Dombal FT. The hit and miss of ISS and TRISS. Yorkshire Trauma Audit Group. Bmj 1993; 307:906-9.
5. Pollock DA. Summary of the discussion: trauma registry data and TRISS evidence. J Trauma 1999; 47:S56-8.
6. Seguin J, Garber BG, Coyle D, Hebert PC. An economic evaluation of trauma care in a Canadian lead trauma hospital. J Trauma 1999; 47:S99-103.
7. Jurkovich GJ, Mock C. Systematic review of trauma system effectiveness based on registry comparisons [see comments]. J Trauma 1999; 47:S46-55.
8. Fayyad UM, Irani KB. Multi-interval discretization of continuous-valued attributes for classification learning, Proc. 13th Int'l Joint Conf. on Artificial Intelligence, 1993. Morgan-Kaufmann.
9. Kononenko I. Estimating attributes: Analysis and extensions of RELIEF., Proc. European Conf. Machine Learning, 1994. Springer-Verlag.
10. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees: Wadsworth International Group, 1984.
11. Quinlan R. Induction of decision trees. Machine learning 1986; 1:81-106.

12. Demsar J, Zupan B, Aoki N, Wall MJ, Granchi TH, Beck JR. Feature mining and predictive model construction from severe trauma patient's data (in press). Int J Med Inform.

13. Machine learning, neural and statistical classification. In: Michie D, Spiegelhalter DJ, Taylor CC, eds: Ellis Horwood, 1994.

14. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982; 143:29-36.

15. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 1983; 148:839-43.

16. Rotondo MF, Zonies DH. The damage control sequence and underlying logic. Surg Clin North Am 1997; 77:761-77.

17. Intelligent Data Analysis in Medicine and Pharmacology. In: Lavrac N, Keravnou E, Zupan B, eds. Boston, 1997.

FIGURE LEGENDS

Figure 1

The decision tree model showed two important predictive factors: pH at initial ICU admission and the worst partial thromboplastin time (PTT) from hospital admission to ICU admission. Box represent predictive outcome in each setting with mortality in our population in each category.