Prediction Model for Survival of Trauma Patients

Rahim Hasnani[#1], Syed Zeeshan Arshad[*2], Dr. Hasnain Zafar[#3]

[#] *Agilosoft Company, The Agha Khan University*
Karachi Pakistan
[1]rhasnani@yahoo.com

[*]*SZABIST, Computer Science Department*
Karachi Pakistan
[3]zeeshan@szabist.edu.pk

***Abstract* - One of the most common calculators used for predicting survival of trauma patients is TRISS (Trauma Injury Severity Score) which is based on multiple regression analysis from the Major Trauma Outcome Study (MTOS) database. Based on the data set of 1,009 trauma patients of a tertiary care hospital in Pakistan, data mining was carried out using Logistic Regression Modelling using an all subsets approach. Models were formulated and ranked based on accuracy, Area under Receiver Operating Curve (AUC) and an overall performance measure that include AUC, sensitivity and specificity. Various models were found to perform better than TRISS. Furthermore, reduced models were also formulated that include factors that are simple to determine and models were found to outperform TRISS.**

***Keywords* - Data Mining, Prediction Model, Logistic Regression, Trauma Survival.**

I. Introduction

A trauma registry captures the relevant demographic, treatments, assessments and outcome information of trauma cases which helps in maintaining records of survival rates, analysis of patterns of survival and other analytical and statistical data.

One of the primary uses of trauma registry is prediction of survival of new patients. The most common calculator used for calculating probability of survival is TRISS (Trauma Injury Severity Score).

Several injury severity ratings or scores have been developed and used since 1970s to evaluate trauma patients [1]. The Abbreviated Injury Scale (AIS) determines the severity of injury on a scale of 1-6 (minor to fatal) in five body areas. AIS, however, did not take into account the cumulative effect of different injuries. [2]

The Injury Severity Score (ISS) uses AIS as a base and takes the top three body regions with the most severe injuries and provides a better correlation with mortality than AIS [2]. A Revised Trauma Score (RTS) was developed that only takes into account systolic blood pressure, respiratory rate and Glasgow Coma Scale. This has resulted in more reliable predictions [2].

TRISS or Trauma Injury Severity Score is based on the Injury Severity Score (ISS), Revised Trauma Score (RTS) and age of the patient. This score provides the survival probability of patient. Based on the Major Trauma Outcome Study, outcome norms were obtained using multiple regression analysis. The norms are different for blunt and penetrating injuries [1].

TRISS is still widely used and have been applied in developing countries as well. Using the outcome norms of MTOS, while there have been studies where the TRISS model has worked in developing countries [3], there are other studies [4] where the model did not work. In addition, an Iranian study [5] has used the TRISS model but derived their own coefficients.

A study in 2002 argues that MTOS norms which are based on western cases do not correlate accurately with actual or observed outcome in developing countries [6]

II. PROBLEM DOMAIN

Department of Surgery, The Aga Khan University provided data of 1,009 trauma patients.

The following statements cover the problem domain:

- Formulation of a local trauma prediction model based on the trauma data of 1,009 trauma patients of The Aga Khan University.
- Identification of non-TRISS data variables that relate to trauma outcome prediction.
- Identification of data elements that can be easily and correctly captured by smaller hospitals (district hospitals).
- Formulation of a simple trauma model for prediction of trauma survival that uses minimal set of data variables.

III. RESEARCH METHODOLOGY

The following methodology is followed in conducting this research:

- Review of Raw Data for completeness and accuracy
- Identification of attributes relevant to research. This is done by computing correlation between variables and outcome.
- Formulation of model based on predictive data mining techniques. This is done by creating various candidate logistic regression models and choosing the appropriate model based on accuracy measures.
- Identification of attributes subset that is easy to capture by local hospitals that may not have specialized and trained trauma teams as well as attributes that are identified by domain experts as relevant to research.
- Identification of minimal subset by applying feature reduction techniques keeping in view the attributes identified by local hospitals.

A. Validation

The validation is done by splitting the sample into development and validation sub-samples. A split of 80:20 was selected. Although, a larger proportion of development sample is desired, a larger split does not give enough validation samples with a false observed value. The split is done such that more recent data is used for validation.

B. 2.2 Comparison Methods

The comparison of predicted versus observed outcomes in logistic regression models is carried out differently as the outcome variable is dichotomous (binary) and the model provides a probability value between 0 and 1. The following comparison methods are used:

1) Accuracy

Using a cut-off of 0.5 a 2 x 2 classification table is created to count number of True positives, True Negatives, False positives, & false negatives as illustrated below:

TABLE 2: CLASSIFICATION TABLE [7]

		Predicted Values	
		True	False
Obser-ved Values	True	True Positive (TP)	False Negative (FN)
	False	False Positive (FP)	True Negative (TN)

Accuracy is then defined as:

Accuracy = (TP + TN) / (TP + FP + FN + TN) [7]

2) Specificity & Sensitivity

In addition to accuracy, specificity and sensitivity measure the accuracy of negative & positive outcomes respectively and are defined as:

Specificity = TN / (TN + FP)
Sensitivity = TP / (TP + FN) [7]

3) Area under Receiver Operating Curve (AUC)

Accuracy and its related measures are based on the cut-off selected and does not distinguish between predictions on the basis of how near or far the prediction is to the cut off value. For example, a false negative with predicted probability 0.48 is treated the same as 0.2. Area under Receiver Operating Curve (AUC) is a better measure of performance evaluation of classification algorithm than accuracy [8] and is given by:

$$AUC = (S_0 - n_0(n_0 + 1)/2) / n_0 n_1 \qquad [8]$$

Where n_0 & n_1 are the number of positive & negative examples,
$S_0 = \sum r_i$, where r_i is the rank of the i^{th} positive example in the ranked list.

4) Overall Performance

An overall performance measure is introduced and is derived from the above measures and is defined as:

Overall Performance =
(AUC + Specificity + Sensitivity) / 3

Accuracy is not taken into consideration as specificity and sensitivity make the overall accuracy.

C. 2.3 Research Methods & Tools

For predictive mining, logistic regression model (LRM) is used. Since TRISS is a logistic regression model and is widely used, results of LRM can be compared with TRISS.

1) Selection of Variables

The independent variables are filtered by first removing the variables with missing or invalid data.

The independent variables are selected by constructing a one variable logistic regression model for each independent variable. The model will be of the form:

$$P_s = 1 / (1 + e^{-(b_0 + b_1 x)})$$

Once the model is constructed, the significance of independent variables is to be computed. This is done by calculating the Wald Statistic (W), which is calculated by:

$$W = \beta / SE(\beta) \qquad [9]$$

Where β is the estimated co-efficient and SE(β) is the standard error of the co-efficient.

The Wald statistic will follow a standard normal distribution and a p value can be calculated. A suitable level of significance (1% or 5%) will help identify significant variables.

IV. MODEL FORMULATION

The data set has a total of 142 different variables and after initial screening to remove variables with no or significantly missing data, or variables not related to study as well as derived variables, and after recoding variables, 36 variables remained as potential independent variables for the study.

A. Significant Independent Variables

A one-variable logistic regression model is created between each candidate independent variable and the dependent variable and the Wald Statistic (W) is computed for the potential independent variable. Wald Statistic shows the significance of the variable and follows a normal distribution.

At the significance level of 0.05 ($p <= 5\%$), twenty five (25) variables show correlation with the dependent variable. However, studies have shown that a traditional cutoff of p value like 5% often fails to identify variables known to be important [10]. Therefore, taking a p value of 0.25 (25%), seven (7) additional variables show correlation with the dependent variables. These thirty two (32) variables are filtered out for further analysis.

B. All Subsets Regression Models

Based on the all subsets approach, all possible two-factor, three-factor, four-factor and five-factor regression models are generated.

For each model, data set is filtered to include only those data records that do not have any missing values for the variables in question. On the filtered data set, 80% of the data is used for generating the model (i.e. computing the model co-efficients) and 20% is used for internal validation. Measures computed for validation include accuracy, sensitivity, specificity and Area under the Receiver Operating Curve (AUC) and overall performance.

In addition, in order to compare the performance of the models with TRISS, the above measures are also computed for the full data set as well.

V. RESULTS

All possible two-, three-, four-, and five-factor regression models are formulated. Of the possible, 242,792 models, 242,474 were computed by the program.

The maximum accuracy is 98.837% achieved by 11 different models. The maximum AUC is 98.742%. The maximum overall performance is 93.987%.

Twenty Seven (27) models show an overall performance of 93% or higher.

An analysis of these model shows that Maximum AIS in External Region, Age, Delay in Trauma team arrival (in minutes), and Motor scores are prominent in these models.

A. Comparison with TRISS

The performance of TRISS on the data is as follows:

Accuracy: 93.4263%
AUC: 88.3186%
Specificity: 20.00%
Sensitivity: 99.3541%
Overall Performance: 69.2242%

For comparison with TRISS, instead of 20% of data used for internal validation, the whole data set is used as TRISS performance is also calculated on the whole data setup. Accuracy, Specificity, Sensitivity and AUC are computed again for each of the models with the full data set.

The models formulated by this data have superior performance than TRISS when compared on the basis of individual measures of AUC, Accuracy as well as overall performance, as 2,431 models have better AUC than TRISS, as many as 133,611 models have better accuracy. In terms of overall performance, 116,607 models performed better than TRISS.

VI. MODEL REFINEMENT

Based on the discussions with domain experts and local hospitals the following should be considered for trauma outcome prediction:

1) Head injury
2) Delay in reaching the hospital or trauma center

In addition to the above two factors identified above, some other factors are also selected as candidates on the basis of ease with which they can be determined. These factors are selected from the list of significant factors. Here, significant factors are those that appear most in top ranked models, where AUC is 80% or more in internal validation.

From that list, factors that can be calculated easily are taken and added to the factors identified by domain experts as well as TRISS factors that can be calculated easily. This yields a list of ten factors:

1. Sum of AIS in Head & Neck
2. Maximum AIS in Head & Neck
3. Delay in Patient's Arrival to ER (in minutes)
4. Age Index
5. Mechanism
6. Respiratory Assistance Provided
7. Verbal Score
8. Eye Opening Score
9. Motor Score
10. Age

A. Reduced Models

The maximum accuracy of reduced models is 96.296% and the maximum AUC is 94.65%. The maximum overall performance is 72.291%.

Nine (9) models show an overall performance of 72% or higher. These are provided in Table 31.

An analysis of these models shows that Motor Score, Respiratory Assistance Provided, and Age are prominent in all these top models.

Moreover, two of the top 9 models include Delay in Arrival of Patient as a factor which was identified by the domain expert whereas one model each contain Maximum AIS in Head & Neck and Sum of AIS in Head & Neck respectively, which was also identified by the domain expert.

B. Comparison with TRISS

Based on accuracy, nine models of two-factors, 39 models of three-factors, and 95 models of four-factors have better performance than TRISS.

Based on AUC, none of the models compete with TRISS.

Based on overall performance, 12 models of two-factors, 49 models of three-factors, and 112 models of four-factors have superior performance than TRISS.

VII. CONCLUSIONS

Based on the application of logistic regression data mining techniques on the available data, models with two, three, four and five factors are formulated and ranked based on accuracy, Area under ROC (AUC), and overall performance (which is an average of AUC, sensitivity and specificity).

It is demonstrated that models with two, three, four and five factors are identified with greater performance than TRISS.

In addition, simpler models of two, three and four factors are identified based on input from domain experts as well as by using those factors that are simple to compute and that were prominent in the top ranked models. These reduced models also perform better than TRISS.

A number of non-TRISS factors are identified and prominent among them is Delay in trauma team arrival (in minutes). In addition, factors that form part of overall computation of TRISS are identified in a simplified form that relate directly to the outcome variable. These include Maximum AIS in External region (which is part of overall Injury severity Score calculation in TRISS) and Motor Score (which is part of Glasgow Coma Scale which is in turn part of Revised Trauma Score calculation).

Similarly, in simplified models, the prominent non-TRISS factor is identified as whether Respiratory Assistance is provided. Factors that form part of overall TRISS calculation and are identified in a simplified form in the reduced models include Motor Score.

The factors identified by the domain experts are also present in the top ranked reduced models and these include Delay in Arrival of Patient and Maximum & Sum of AIS in Head & Neck region.

VIII. FUTURE DIRECTIONS

An attempt has been made to analyze the data based on logistic regression data mining techniques. It is suggested that this research & analysis be carried forward in the following directions:

1. External validation of top ranked models identified in the current analysis with the use of new data.
2. Formulation of models using other data mining techniques (like decision tree, Artificial Neural Networks) and doing a comparison of the performance.
3. Use of other internal validation methods, like bootstrapping and k-fold cross validation instead of Split sample.

REFERENCES

[1] Howard R. Champion, Wayne S. Copes, William J. Sacco, Mary M. Lawnick, Susan L. Keast, Lawrance W. Bain, Maureen E. Flanagan, Charles F. Fary, The Major Trauma Outcome Study: Establishing National Norms for Trauma Care, The Journal of Trauma, Vol 30 No. 11, pp. 1356-65

[2] Carl R. Boyd, Mary Ann Tolson, Wayne S. Copes, Evaluating Trauma Care: The TRISS Method, The Journal of Trauma, Vol 27 No 4, April 1987, pp. 370-378 (evaluating trauma care.pdf)

[3] Chaiyut Thanapaisal MD, Narongchai Wongkonkitsin MD, O-Tur Sae Seow MD, Dhanes Rangsrikajee MD, Kriangsak Jenwitheesuk MD, Ake Phugkhem MD, Vajarabhongsa Bhudisawadi MD, Outcome of In-Patient Trauma Cases: Accident and Emergency Unit, Khon Kaen University, Journal of The Medical Association of Thailand, Volume 88 No. 11, November 2005 (Vol88_No11_1540.pdf)

[4] S Hariharan, D Chen, K Parker, A Figari, G Lessey, D Absolom, S James, O Fraser, CT Letsholathebe, Evaluation of trauma care applying TRISS methodology in a Caribbean developing country, The Journal of Emergency Medicine, 2008 June 25

[5] Abbas Rabbani, Majid Moini, Application of “Trauma and Injury Severity Score” and “A Severity Characterization of Trauma” Score to Trauma Patients in A Setting Different from “Major Trauma Outcome Study”, Archives of Iranian Medicine, Volume 10, Number 3, 2007, pp 383 – 386.

[6] H Zafar, R Rehmani, A J Raja, A Ali, M Ahmed, Registry based trauma outcome: perspective of a developing country, Emergency Medical Journal, Volume 19, September 2002, pp. 391-4.

[7] Wikipedia, Sensitivity & specificity – Wikipedia, the free encyclopedia, *Sensitivity & specificity*, [online], Available: http://en.wikipedia.org/wiki/Sensitivity_and_specificity, [Accessed: 19 October 2009]

[8] Charles X Ling, Jin Huang, Harry Zhang, AUC: a Statistically Consistent and more Discriminating Measure than Accuracy, Proceedings of 18th International Joint Conference on Artificial Intelligence, Acapulco, Mexico, August 9-15, 2003, pp. 519-526 (ijcai03.pdf)

[9] David W. Hosmer, Stanley Lemeshow, *Applied Logistic Regression*, Wiley Series in Probability and Mathematical Statistics, pp16-17.

[10] David W. Hosmer, Stanley Lemeshow, *Applied Logistic Regression*, Wiley Series in Probability and Mathematical Statistics, pp86.