The Utility of Trauma Scoring Systems in Outcome Prediction of Critical Patients

GAMAL HAMED, M.D.; EMAD OMAR, M.D.; AKRAM ABD EL-BARY, M.D. and SHAMS EL-DEEN SAYED, M.Sc.
The Department of Critical Care, Faculty of Medicine, Cairo University

Abstract

Background: Mortality predictions calculated using scoring scales are often not accurate in populations other than those in which the scales were developed because of differences in case-mix. Scoring systems should be validated to be used in a different population.

Methods: Probabilities of ICU death for critically ill trauma patients were calculated using the anatomic scores: Injury severity score ISS and new injury severity score NISS, the physiologic scores: Revised trauma score RTS and acute physiology and chronic health evaluation II APACHE II score and the combined scores: Trauma and injury severity score TRISS and a severity characterization of trauma ASCOT score and the modification of TRISS to TRNISS. The predicted mortality probabilities were compared to the observed mortality. The scores were assessed for discrimination using receiver operator characteristics area under the curve ROC AUC and for goodness-of-fit using Hosmer-Lemeshow test.

Aim: To study the utility of different trauma systems in outcome prediction and validation of these scores to be used in the Egyptian Hospitals.

The primary outcome under investigation is mortality, but other outcomes were also studied which are the ICU length of stay (LOS) and Glasgow outcome score (GOS).

Results: 80 trauma patients admitted to ICU over 1 year could be analyzed. Non survivors during ICU stay were 29 patients (36.25%). In the study 91.25% of the victim’s injuries were caused by blunt trauma and only 8.75% were caused by penetrating trauma. RTS, APACHE II and NISS showed very good ROC AUC while ISS showed a good one (ROC AUC =0.841, 0.881, 0.847 and 0.779 respectively). The combined scores TRISS, ASCOT and TRNISS showed excellent discrimination (ROC AUC =0.902, 0.937 and 0.911 respectively). Regarding calibration RTS, TRISS and ASCOT were well calibrated in our study (HL statistics; \(p=0.878, 0.781\) and 0.843 respectively) while ISS, NISS, APACHE II and TRNISS were poorly calibrated (HL statistics; \(p=0.152, 0.233, 0.102\) and 0.140 respectively).

Conclusion: Both TRISS and ASCOT performed better than the anatomic and physiologic scores as regard calibration and discrimination. The replacement of ISS with NISS in TRISS (TRNISS) resulted in better discrimination but poor calibration. We recommend the combination of APS and NISS to be applied on a larger scale and validated.

Key Words: Trauma scoring systems – Critical care – ISS – RTS – APACHE II – TRISS – ASCOT.

Introduction

Attempting to summarize the severity of injury in a patient with multiple traumas with a single number is difficult at best; therefore, multiple alternative scoring systems have been proposed, each with its own problems and limitations. An accurate method for quantitatively summarizing injury severity has many potential applications. The ability to predict outcome from trauma (i.e., mortality) is perhaps the most fundamental use of injury severity scoring, a use that arises from the patient's and the family's desires to know the prognosis [1].

More than 50 scoring systems have been published for the classification of trauma patients in the field, emergency room, and intensive care settings. There are three main groups of trauma scores: (1) Anatomical, (2) physiological, (3) comorbidity scores. The three types of scores can be combined in different ways to obtain more accurate information from all possible aspects. Mortality predictions calculated using scoring scales are often not accurate in populations other than those in which the scales were developed because of differences in case-mix. Scoring systems should be validated to be used in a different population [2].

Aim of the work:

The aim of the work is to study the utility of different trauma scoring systems in outcome prediction and validation of these scores to be used in the Egyptian Hospitals.
The primary outcome under investigation is mortality, but other outcomes were also studied which are the ICU length of stay (LOS) and Glasgow outcome score (GOS).

The scoring systems that has been studied are: The anatomical scores (injury severity score ISS, new injury severity score NISS), the physiologic scores (revised trauma score RTS, acute physiology and chronic health evaluation score APACHE II, and its acute component the acute physiology score (APS) and the combined score (trauma and injury severity score TRISS, trauma and new injury severity score TRNISS, a severity characterization of trauma score (ASCOT) [3-8].

Patients and Methods

80 trauma patients admitted to ICU in Ahly Bank Hospital starting July 2008 and through June 2009. All patients were included except:
- Children <2 year.
- Patients with cardio-respiratory arrest before ICU admission.
- Patients with AIS of 6 in any body region (ISS =75).

Probabilities of ICU death for critically ill trauma patients were calculated for each score.

The probability of mortality using APACHE II score was adjusted according to diagnostic categories. Measures of model discrimination were based on Receiver Operator Characteristics Area Under Curve (ROC AUC). Measures of model calibration were based on goodness-of-fit Hosmer-Lemshow (HL) statistics. The scores were correlated to LOS and GOS. Logistic regression was used for the different components of the scores to structure a new model for mortality prediction of trauma patients [9].

Results

80 trauma patients admitted to ICU over 1 year could be analyzed 67 were males. Mean age was 31±14 years (ranged from 17 to 47 years). In the study 91.25% of the victims’ injuries were caused by blunt trauma and only 8.75% were caused by penetrating trauma. Table (1) summarizes the study population characteristics. According to the primary outcome we subdivided the patients into two groups:
- **Group I:** Patients who survived through the ICU stay included 51 patients (63.75%).
- **Group II:** Non survivors during ICU stay were 29 patients (36.25%) There was no significant difference between both groups as regards age or gender or type of trauma Table (1).

<table>
<thead>
<tr>
<th>Table (1): Study population characteristics.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characteristic</td>
</tr>
<tr>
<td>-----------------------------------------</td>
</tr>
<tr>
<td>Number</td>
</tr>
<tr>
<td>Gender</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Injury type</td>
</tr>
<tr>
<td>Blunt</td>
</tr>
<tr>
<td>Penetrating</td>
</tr>
<tr>
<td>ISS Pd</td>
</tr>
<tr>
<td>RTSb</td>
</tr>
<tr>
<td>APACHE II Pd</td>
</tr>
<tr>
<td>TRISSb</td>
</tr>
<tr>
<td>ASCOTb</td>
</tr>
<tr>
<td>APSb</td>
</tr>
<tr>
<td>NISSb</td>
</tr>
<tr>
<td>TR+NISSb</td>
</tr>
<tr>
<td>LOSb</td>
</tr>
<tr>
<td>Days of MVb</td>
</tr>
<tr>
<td>Seizuresa</td>
</tr>
<tr>
<td>VTEa</td>
</tr>
<tr>
<td>Sepsisa</td>
</tr>
</tbody>
</table>

a: Count (percent), b: Mean (SD), ***: Significant.
The scores RTS, APACHE II, APS and NISS showed very good ROC AUC while ISS showed a good one. The combined scores TRISS, ASCOT and TRNISS showed excellent discrimination. Regarding calibration RTS, APS, TRISS and ASCOT were well calibrated in our study while ISS, NISS, APACHE II and TRNISS were poorly calibrated. Table (2) shows the scores discrimination and calibration abilities. (Figs. 1, 2) show the scores ROC AUC.

All the scoring systems were correlated to GOS while only APACHE II, APS, ISS and TRNISS were correlated to ICU length of stay Table (3).

A proposed model made up by APS as the physiologic components, NISS as the anatomic component and AGE as the comorbidity component was tested using logistic regression analysis. Age was excluded from the model while APS and NISS formed a significant model for mortality prediction with the following coefficients:

\[
\text{Logit} = -0.41 + 0.494 \times (\text{APS}) + 0.396 \times (\text{NISS})
\]

\[
\text{Predicted death rate} = \frac{1}{1 + e^{-\text{logit}}}
\]

\[e = 2.718282\] (base of Naperian logarithm).

**Table (2): Trauma scores discrimination and calibration abilities.**

<table>
<thead>
<tr>
<th>Score</th>
<th>AUC</th>
<th>SE</th>
<th>95% confidence interval</th>
<th>(\chi^2)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>APACHE II</td>
<td>0.881</td>
<td>0.036</td>
<td>0.810–0.952</td>
<td>13.292</td>
<td>0.102</td>
</tr>
<tr>
<td>APS</td>
<td>0.891</td>
<td>0.036</td>
<td>0.820–0.962</td>
<td>3.515</td>
<td>0.089</td>
</tr>
<tr>
<td>RTS</td>
<td>0.841</td>
<td>0.049</td>
<td>0.745–0.936</td>
<td>6.768</td>
<td>0.878</td>
</tr>
<tr>
<td>ISS</td>
<td>0.779</td>
<td>0.051</td>
<td>0.678–0.879</td>
<td>10.715</td>
<td>0.152</td>
</tr>
<tr>
<td>ASCOT</td>
<td>0.937</td>
<td>0.027</td>
<td>0.858–0.990</td>
<td>4.149</td>
<td>0.843</td>
</tr>
<tr>
<td>TRISS</td>
<td>0.902</td>
<td>0.035</td>
<td>0.833–0.979</td>
<td>3.847</td>
<td>0.871</td>
</tr>
<tr>
<td>NISS</td>
<td>0.847</td>
<td>0.044</td>
<td>0.760–0.933</td>
<td>9.286</td>
<td>0.233</td>
</tr>
<tr>
<td>TRNISS</td>
<td>0.911</td>
<td>0.032</td>
<td>0.850–0.973</td>
<td>12.235</td>
<td>0.140</td>
</tr>
</tbody>
</table>

**Table (3): Scores correlation to GOS and LOS.**

<table>
<thead>
<tr>
<th>Score</th>
<th>Correlation to GOS</th>
<th>p-value</th>
<th>Correlation to LOS</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>RTS</td>
<td>0.710(**)</td>
<td>0.000</td>
<td>-0.190</td>
<td>0.092</td>
</tr>
<tr>
<td>APACHE II</td>
<td>-0.677(**)</td>
<td>0.000</td>
<td>0.247(*)</td>
<td>0.027</td>
</tr>
<tr>
<td>APACHE pd</td>
<td>-0.511(**)</td>
<td>0.000</td>
<td>0.121</td>
<td>0.284</td>
</tr>
<tr>
<td>APS</td>
<td>-0.692(**)</td>
<td>0.000</td>
<td>0.227(*)</td>
<td>0.043</td>
</tr>
<tr>
<td>ISS</td>
<td>-0.519(**)</td>
<td>0.000</td>
<td>0.273(*)</td>
<td>0.014</td>
</tr>
<tr>
<td>NISS</td>
<td>-0.649(**)</td>
<td>0.000</td>
<td>0.216</td>
<td>0.054</td>
</tr>
<tr>
<td>TRISS pd</td>
<td>-0.684(**)</td>
<td>0.000</td>
<td>0.140</td>
<td>0.214</td>
</tr>
<tr>
<td>ASCOT</td>
<td>-0.733(**)</td>
<td>0.000</td>
<td>0.103</td>
<td>0.362</td>
</tr>
<tr>
<td>TRNISS</td>
<td>-0.780(**)</td>
<td>0.000</td>
<td>0.266(*)</td>
<td>0.017</td>
</tr>
</tbody>
</table>

****: Correlation is significant at the 0.01 level (2-tailed). *: Correlation is significant at the 0.05 level (2-tailed).

Fig. (1): ROC of RTS, ISS and APACHE II Pd, NISS, APS.

Fig. (2): ROC of TRISS,ASCOT,TRNISS.
Discussion

Injury Severity Score (ISS) as the anatomic score is compared to the physiologic scores RTS and APACHE II. Aspects of comparison included ICU mortality and length of stay.

The study showed significantly increased mortality in relation to increased ISS, APACHE II score and to decreased RTS \((p=0.0001)\). The scores predicted mortality was tested for discrimination and calibration.

The study showed that both physiologic scores had very good AUC. On the other hand ISS had good AUC. Also correct classification rates were higher for RTS and APACHE II than ISS. RTS was better calibrated than both ISS and APACHE II.

Despite this trend for the physiologic Scores to show better discrimination than the anatomic scores, there was no statistically significant difference between scores’ AUCs in our study.

Our findings are similar to Hafiz [10] who found that RTS is a reliable predictor of prognosis of polytraumatized patients. Also, Ohaegbulam et al., [11] confirmed that RTS is a good predictor of both severity of head injury (and thus the need for ICU admission) and mortality.

Markgraf [12] and Lee, [13], validated the APACHE II for predicting mortality of trauma patients as well as their length of stay.

Similar to our result Liang et al., [14], found that APACHE II is a better predictor for ICU trauma patients than ISS. Dossett [15], found that APACHE II is better than ISS as well as TRISS.

Also Wang and Cho [16] showed that APACHE II is reliable for predicting mortality and functional outcome in trauma.

Against our results are the results of Eryilmaz [17] who found that there is no difference between physiologic and anatomic scoring systems to predict mortality.

In our study 2 commonly used combined trauma scores were used and compared to the anatomic and physiologic scoring models: TRISS and ASCOT.

In our study the mean TRISS and ASCOT scores. The study showed that there were statistically significant differences in the scores between survivors and non survivors. However, both TRISS and ASCOT were not correlated to length of stay.

Indeed both scores showed excellent AUC with no statistically significant difference between both scores’ AUCs. In addition TRISS and ASCOT were well calibrated.

The results of our study give us the support to use both combined scores for mortality prediction in critically ill trauma patients. However, the simplicity of TRISS calculation makes it easier and more applicable that ASCOT.

These results of our study are similar to what Rabbani et al., [18] and Sirirongtaworn et al., [19] confirmed: The accuracy of TRISS methodology in trauma mortality prediction reached.

Hariharan [20] evaluated TRISS in Trinidad hospitals. He agreed with the discrimination ability of TRISS but questioned its calibration and hence its use in a developing country. He attributed the disparity between predicted and observed mortality in his study to the difference in case-mix between Trinidad and USA where the original TRISS was developed.

In our study both TRISS and ASCOT preformed better discrimination than ISS. On the other hand there was no statistically significant difference between either combined scores or the physiologic ones as regard their discrimination ability.

Older study investigated TRISS and ASCOT. Wong et al., [21] in their study showed that both APACHE II and TRISS can accurately predict group mortality (calibration) but neither provides sufficient confidence for prediction of outcome of individual patients (discrimination). They didn’t find statistical difference between the 2 scores.

Champion et al., [22] called for the use of ASCOT as the standard method of trauma outcome prediction. They found improved calibration for ASCOT over TRISS.

On the other hand Markle et al., [23] attempted to validate ASCOT. Their study showed relatively small gain in prediction accuracy by ASCOT over TRISS. However, this gain is offset by ASCOT complexity.

Osterwalder et al., [24] used ISS, TRISS and ASCOT predicted mortality as a tool for quality management. They agreed with the superiority of TRISS and ASCOT over ISS. Similar to our work, they recommended the use of TRISS for easier application.

Although the use of ASCOT score resulted in better mortality prediction, its complexity limited its use and called for the use of the easier to use TRISS [23,24].
In our study we assessed NISS which uses the worst 3 body injuries regardless of their areas. The mean NISS was significantly higher in non survivors than survivors. Using NISS resulted in a good AUC, better CC rate and better calibration than ISS. Yet there was no significant difference between ISS and NISS AUC. These results support the use of NISS instead of ISS for 2 reasons: The improved performance and easier application of NISS over ISS.

Our results are similar to Zhao et al., Tamim et al., and Köksal et al.. All of them encouraged the use of NISS although there was no significant difference in discrimination between ISS and NISS.

In addition Frankema et al. showed that NISS outperformed ISS in prediction of mortality.

Our study tested the use of NISS instead of ISS in TRISS (TRNISS). The study found significant difference in TRNISS between survivors and non survivors. TRNISS had excellent ROC. There was no statistically significant difference in ROC AUC between TRNISS and TRISS. TRNISS was poorly calibrated and was correlated to ICU LOS.

While we suggested that NISS better describes the injury severity than ISS, we didn’t find any improvement in performance for TRNISS over TRISS. A possible explanation is that the coefficient given to ISS in TRISS methodology need to be recalculated if NISS will be used.

That was consistent with what Aydin et al. found. They found no significant difference between TRISS and TRNISS for mortality prediction as regard discrimination. Yet, they recommended TRNISS for its easier application.

Most of the researchers kept RTS as the physiologic index in risk prediction models. In our study we suggested the use of the acute physiologic component in APACHE II which is APS in addition to NISS. The study showed significant difference in APS between survivors and non survivors.

We hypothesized that combining APS and NISS gives the simplicity of applying NISS as a measure of injury severity together with the thorough description of the physiologic derangement given by APS. We included the age as an indicator of co-morbidities and used as continuous variable. Using regression analysis age was excluded from the model \(p=0.068\). The suggested model was statistically significant which allows for further evaluation on larger study population \(p=0.0001\).

Conclusion:

Combining the anatomic and physiologic parameters in TRISS and ASCOT for mortality prediction of trauma patients resulted in improved discrimination and calibration. However, the complexity of ASCOT makes TRISS the preferable scoring system to use. The replacement of ISS with NISS in TRISS methodology didn’t result in significant improvement. Contrarily, TRNISS showed poor calibration. A proposed model made up by APS and NISS is introduced in this study for further evaluation.

References

13- C.K. LEE and T.H. RAINER.: Application of APACHE II in the assessment, classification of severity and predic-
The Utility of Trauma Scoring Systems in Outcome Prediction of Critical Patients


18- RABBANI A. and MONTI M.: 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”. Iranian J. of medicine, 10 (3): 383-6, 2007.


