Prediction of Mortality in Pediatric Trauma Patients: New Injury Severity Score Outperforms Injury Severity Score in the Severely Injured

Thomas Sullivan, BS, Adil Haider, MD, MPH, Stephen M. DiRusso, MD, PhD, Peter Nealon, BA, Aasma Shaukat, MD, MPH, and Michel Slim, MD

Background: The Injury Severity Score (ISS) is a widely accepted method of measuring severity of traumatic injury. A modification has been proposed—the New Injury Severity Score (NISS). This has been shown to predict mortality better in adult trauma patients, but it had no predictive benefit in pediatric patients. The aim of this study was to determine whether the NISS outperforms the ISS in a large pediatric trauma population.

Methods: Admissions in the National Pediatric Trauma Registry between April 1996 and September 1999 were included. The ISS and NISS were calculated for each patient. The study endpoints were mortality at hospital discharge, functional outcome in three domains (expression, locomotion, and feeding), and discharge disposition for the survivors. Predictive ability of each score was assessed by area under the receiver operating characteristic curve.

Results: The NISS and ISS performed equally well at predicting mortality in patients with lower injury severity (ISS < 25), but the NISS was significantly better at predicting mortality in the more severely injured patients. Both scores performed equally well at predicting expression and feeding ability. The NISS was superior to the ISS in predicting locomotion ability at discharge. Thirty-seven percent of patients had an NISS that was higher than their ISS. These patients had a significantly higher mortality and suffered worse functional outcomes.

Conclusion: The NISS performs as well as the ISS in pediatric patients with lower injury severity and outperforms the ISS in those with higher injury severity.

Key Words: Pediatric trauma, Survival models, Outcome analysis, New Injury Severity Score (NISS).


Over the past 25 years, the Injury Severity Score (ISS) has become a widely accepted indicator of human injury. During the past few years, a simple modification of this score, the New Injury Severity Score (NISS), has been introduced. Several studies have demonstrated it as a superior predictor of mortality in adults. However, it has been shown recently that the NISS may not outperform the ISS in the pediatric population.

In an effort to explore this dichotomy further, we set out to analyze the predictive capabilities of the NISS and the ISS using a large national pediatric trauma registry. Specifically, we intended to use multiple previously described methods to compare the ability of both scores to accurately predict survivors versus nonsurvivors, and to investigate the ability of anatomic injury scoring systems (ISS and NISS) to predict functional outcomes at discharge.

Submitted for publication February 22, 2002.
Accepted for publication August 19, 2003.
Copyright © 2003 by Lippincott Williams & Wilkins, Inc.
Address for reprints: Stephen C. M. DiRusso, MD, PhD, Department of Surgery, New York Medical College, Munger Pavilion, Valhalla, NY 10595; email: sdirusso@worldnet.att.net.
DOI: 10.1097/01.TA.0000102175.58306.2A

PATIENTS AND METHODS

The patient population was composed of patients entered in phase III of the National Pediatric Trauma Registry (NPTR) with admission dates from April 1996 through September 1999. The NPTR is a national pediatric trauma registry started in 1985 and is designed to provide data for studying outcomes and management of the acutely injured pediatric trauma patient. At the time of our study, 93 pediatric hospitals or trauma centers were contributors to this database, which includes all trauma patients aged less than 20 years admitted with a primary diagnosis of injury. Burns, drowning and near drowning, and poisoning are excluded from the data set. At most hospitals, trauma nurse coordinators collect data, which include demographics, mechanism of injury, injury diagnoses, prehospital care, hospital care and procedures, and outcomes at discharge. The data set includes an ISS calculated by each center and all International Classification of Diseases, Ninth Revision injury diagnosis codes with their associated Abbreviated Injury Scale (AIS) and body system identifiers. We calculated the NISS from these AIS values for all patients exactly as described by Osler et al., by adding the sum of the squares of the three highest AIS values irrespective of body region. During data analysis, discrepancies were noted between our calculated ISS and the ISS in the registry. There were some patients with ISS recorded but no AIS values with their injury diagnoses and some patients with AIS values but no ISS. At this point, we calculated the ISS for each patient on the basis of the same

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.
The endpoints studied were mortality, functional outcome, and discharge disposition. Mortality was taken as survival or death at hospital discharge. Functional outcome at discharge was measured using a modification of the Functional Independence Measure (FIM)\textsuperscript{10} score in each of three major domains: expression, locomotion, and feeding. Each domain had a range from 1 to 4 (1, normal function; 4, poorest function). Differences were analyzed between normal (FIM = 1) and abnormal score (FIM > 1) in each of the three domains. Discharge disposition was defined as whether or not the patient was discharged to a facility for further medical care. Facilities for further care included rehabilitation hospitals, long-term care hospitals, and hospitals closer to home. Discharges without further medical care included home, foster care, child protective services, and correctional facilities.

We further divided the patients into two groups on the basis of whether the ISS and NISS were concordant (ISS = NISS) or discordant (ISS < NISS), and examined differences between these two groups.

The ability to discriminate between these endpoints was assessed by the area under the receiver operating characteristic (ROC) curve\textsuperscript{11,12} for the ISS and the NISS. The ROC curves are constructed by plotting the true-positive fraction versus the false-positive fraction. An ROC of 1 implies perfect discrimination, whereas an ROC of 0.5 is equivalent to a random model. Area under the ROC curves and comparisons between ROC curves were made using the nonparametric method.\textsuperscript{13,14} The concordant and discordant groups were compared using the Student’s $t$ test for continuous variables and the $\chi^2$ test for categorical variables. All statistics were computed using Stata version 7.\textsuperscript{14}

### RESULTS

There were a total of 35,385 patients aged <1 year to 20 years old (average age, 8.1 years), with 64% male patients and 36% female patients. Overall mortality was 3% (1,047 deaths). The average NISS for the entire cohort was 10.8 and the average ISS was 8.1. This cohort suffered predominately (92%) blunt trauma. The NISS was higher in 13,046 (37%) of the cases; when differences occur, the NISS is always higher.

The ability of each score to discriminate survivors from nonsurvivors is shown in Table 1. Data are shown as the area under the ROC curve with the 95% confidence interval (CI). There was no statistically significant difference between the NISS and the ISS in prediction of mortality for the overall population (Fig. 1). Penetrating trauma was found in 2,829 patients and, although the NISS and the ISS are equally good at discriminating survivors in this group, both scores perform significantly better in this subset than in the overall population. No significant differences were noted between the ISS and the NISS when patients were stratified by age. In the most severely injured patients (ISS > 24, n = 1,942), the NISS proved to be significantly superior to the ISS in predicting mortality even though the abilities of both scores were degraded as compared with the overall population. In this severely injured group, the NISS and the ISS differed in 75% of patients as compared with 34% in the nonseverely injured. For those patients in whom the NISS and the ISS differed, the average difference was 15.9 for the severely injured and 4.8 for the nonseverely injured.

Of the patients studied, 1,481 had missing functional outcome data in at least one of three variables studied, yielding 95.8% of patients with complete functional outcome at discharge information. Table 2 shows the ability of the two scores to predict abnormal functional outcome or the need for extended medical care in patients who survive to hospital discharge. Both scores perform equally well in predicting which patients will be discharged to other medical facilities for additional treatment. There was no significant difference in either score’s ability to predict abnormal expression or abnormalities in feeding at discharge. However, the NISS was significantly better than the ISS ($p < 0.01$) at predicting abnormal locomotion ability at discharge. The NISS also significantly outperformed the ISS in predicting the presence of a functional impairment in any one of the three measures studied (Table 2).

Table 3 shows the differences between the concordant (ISS = NISS) and discordant (ISS < NISS) groups. The discordant group is slightly older but has the same gender distribution as the concordant group. The discordant group also had a higher mortality and, in the survivors, had a greater need for extended medical care after discharge. This discordant group also had significantly worse functional outcome in all three domains studied. When the analysis was limited to only the discordant group, we found no statistically significant difference in either score’s prediction of mortality. Area under the ROC curve in the discordant group was as follows:

### Table 1  Ability of the ISS and the NISS to Predict Mortality in Pediatric Trauma Patients\textsuperscript{a}

<table>
<thead>
<tr>
<th>Mortality</th>
<th>No.</th>
<th>ISS (95% CI)</th>
<th>NISS (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All patients</td>
<td>35,385</td>
<td>0.888 (0.872–0.904)</td>
<td>0.889 (0.873–0.905)</td>
</tr>
<tr>
<td>Penetrating trauma</td>
<td>2,829</td>
<td>0.946 (0.910–0.983)</td>
<td>0.947 (0.912–0.983)</td>
</tr>
<tr>
<td>Blunt trauma</td>
<td>32,276</td>
<td>0.884 (0.867–0.901)</td>
<td>0.886 (0.869–0.903)</td>
</tr>
<tr>
<td>Severely injured (ISS &gt; 24)</td>
<td>1,942</td>
<td>0.532 (0.506–0.559)</td>
<td>0.643 (0.619–0.666)</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Area under ROC curve.
NISS, 0.929 (95% CI, 0.918–0.939); ISS, 0.926 (95% CI, 0.915–0.936).

**DISCUSSION**

Over the past several years, many different scoring systems have attempted to replace the ISS as the standard for measuring the severity of traumatic injury.\(^7^1,15\)–\(^17\) Although these scores have been shown to have specific advantages, none of them have been shown to be superior enough to warrant a "change of metric," given the considerable cost and effort required to do so.

The NISS, which was developed by a group including the original authors of the ISS, has come to be seen as an update that simplifies scoring and accuracy by improving assessment of patients who have severe and multiple injuries located in the same body region. Most of the studies comparing these two scoring techniques have favored the NISS, but these studies have been confined mainly to adult populations. Recently, Grisoni et al.\(^8\) have shown that in the pediatric age group there is no difference in the predictive capabilities of the two scores. Our research responds to suggestions that the scores be compared using a large cohort of patients, and that the comparisons include the ability to predict parameters of morbidity and functional outcome, to determine which score performs the best overall.\(^3\)

Use of the NPTR database was an effective way of exploring the strengths of the two scores. It is the largest database for pediatric trauma. It also contains variables pertaining to functional status at discharge and discharge disposition, allowing us to test the NISS and the ISS on these parameters as well.

In comparing the ability of the ISS and the NISS to discriminate between survivors and nonsurvivors for the entire cohort of patients, we found that the NISS seemed to perform slightly better, although the 95% confidence intervals of the ROC areas overlapped. This was also seen by

---

**Table 2** Ability of the NISS and the ISS to Predict Abnormal Functional Outcome at Discharge and Discharge Disposition in Pediatric Trauma Survivors\(^a\)

<table>
<thead>
<tr>
<th>Functional outcome</th>
<th>Survivors with Abnormality (%)</th>
<th>ISS (95% CI)</th>
<th>NISS (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expression</td>
<td>4</td>
<td>0.821 (0.807–0.834)</td>
<td>0.823 (0.810–0.837)</td>
</tr>
<tr>
<td>Locomotion</td>
<td>26</td>
<td>0.658 (0.652–0.665)</td>
<td>0.675 (0.669–0.681)</td>
</tr>
<tr>
<td>Feeding</td>
<td>9.6</td>
<td>0.669 (0.659–0.679)</td>
<td>0.663 (0.653–0.674)</td>
</tr>
<tr>
<td>Functional impairment in any one of the three domains</td>
<td>31.3</td>
<td>0.634 (0.628–0.640)</td>
<td>0.619 (0.613–0.625)</td>
</tr>
<tr>
<td>Discharged to extended medical care facility</td>
<td>4.5</td>
<td>0.821 (0.809–0.833)</td>
<td>0.822 (0.810–0.834)</td>
</tr>
</tbody>
</table>

\(^a\) Area under the ROC curve.
Osler et al. in the original work describing the NISS, and was expected because of the paired nature of the data. However, unlike the results of Osler et al., when the ROC curves for the NPTR were compared, there was no statistically significant difference between the two. One explanation for this lack of statistical difference may be that the pediatric patients tended to be less severely injured, with a mean ISS and NISS of 8.1 and 10.8, respectively. Studies that showed differences between the two scoring systems in primarily adult populations had mean ISS and NISS values of 25 and 33, respectively. This is substantiated by the fact that when we compared the two scores in a subgroup of patients with severe injuries, the NISS did significantly better. Therefore, the NISS does predict mortality better in severely injured pediatric patients. This is in contrast to the results of Grisoni et al., who did not observe a significant difference in predictive ability in severely injured patients. However, the NPTR phase III contains four times as many patients as the data set used in their study.

Another explanation is seen in the percentage of discordant scores. Osler et al. demonstrated 60% of primarily adult patients with differing scores. Our results showed that 37% of pediatric patients had differences between the ISS and the NISS. In patients who are not severely injured, 34% had discordant scores, whereas in the severely injured, 75% of the patients had discordant scores. The average difference in patients with discordant scores was 4.8 in those patients not severely injured, compared with 15.9 for the severely injured.

Accurate prediction of mortality in penetrating trauma is important in children as, contrary to popular belief, there was a relatively high incidence (8%) of penetrating trauma in the NPTR. It has also been proposed that the NISS will do better in penetrating trauma. We did not find this to be true. Both scores showed superior performance in predicting mortality in penetrating injury versus blunt injury. This may be because the penetrating trauma subgroup was less severely injured with a mean ISS of 5.3 ± 7.3 and a mean NISS of 7.3 ± 10.9. The mortality in the penetrating trauma group (3.5%) was not statistically different from that in the blunt group (2.9%).

When the analysis was restricted to patients with discordant NISS and ISS, it was noted that the discordant group had a higher mortality rate and worse functional outcome in all domains. A criticism of previous studies has been that dividing patients this way creates two different populations that should not be compared. Perhaps this is the important point. It is not the value of the NISS but the fact that the two scores are different that is important. This identifies patients with multiple injuries in the same body area that outweigh the injuries in other body systems, which ultimately may reflect strongly on survival and other outcome measures.

The main limitation of this study is that it relies on a registry that is collected from many different centers. These centers almost certainly have differences in the way data are gathered and processed, and this adds a source of variability to the data. This was especially true for the functional outcome measures and discharge disposition; as previously mentioned, they were not available for every patient who survived.

CONCLUSION
The NISS performs as well as the ISS in predicting mortality in pediatric trauma patients who are not severely injured (ISS ≤ 24), and performs significantly better than the ISS in predicting mortality for those who are severely injured. The NISS and the ISS are both able to predict functional outcomes and discharge disposition, with the NISS outperforming the ISS at predicting the presence of at least one limitation; specifically, locomotion. The absolute value of the NISS may not be the most important parameter. Perhaps it is the division of concordant and discordant groups that has
NISS vs. ISS in Pediatric Trauma

significant. The discordant group has a significantly higher mortality and worse functional outcome than the group where the ISS and the NISS agree. This will be the focus of our future research in this area.

REFERENCES


DISCUSSION

Dr. Steven Stylianos (New York, New York): The group from Valhalla has applied the NISS to a large, albeit highly select, cohort of patients from the NPTR. Appropriate statistical methodology was applied, and I applaud the authors’ efforts to look beyond survival and to investigate the relationship between the ISS, the NISS, and functional measures.

As you heard, the early enthusiasm for the NISS in adults has not caught on in the pediatric trauma community, and the authors’ conclusions will not alter that tepid reaction. From the more than 35,000 patients studied, a severely injured cohort of less than 2,000 had results worthy of further scrutiny. It is important to note that the majority of these severely injured children had discordant scores and that the NISS was slightly better able to predict mortality and functional outcome.

With this flicker of usefulness suggested for the NISS, I have two questions. First, in your current study and in the previously published data from Grisoni et al., the ISS and the NISS both predict mortality quite poorly in the severely injured children, as determined by area under the ROC curve. What explains this poor performance, and are the predicted values of these two severity scores worth pursuing in severely injured children?

Second, the FIM has been validated only for children 7 years of age and older. Did you limit your analysis of function to children at this appropriate age and, if not, how could this have biased your conclusions? I applaud the authors for a nice article and the program committee for inviting me to discuss it. Thank you.

Dr. J. Wayne Meredith (Winston-Salem, North Carolina): I want to rise also to applaud this group for looking beyond mortality. I think resource use and outcomes are the next layer of data that we need to start applying science to in terms of outcome measures and trauma.

I do have some methodologic concerns about your study. One is that I did not hear any description of your calibration, although I did hear some good descriptions of your calibration comparing these studies, and I thought that was good. The other is, if you think back on where the AIS scores come from, they are all mortality derived. They are looking at every injury and looking at their possible contribution to mortality.

Therefore, if you think about a median nerve laceration, that is a very low AIS score. It’s going to have a very high functional outcomes score. Thus, I think we’re going to need to develop a whole new approach to looking at outcomes and resource use that’s separate from the mortality risk factors. It’s just not going to work to use AIS scores to get to that. Thank you.

Dr. David P. Mooney (Boston, Massachusetts): Because mortality and morbidity in the world of pediatric trauma turn so heavily on brain injury, did you find any significant impact of brain injury on those variants in the analyses?

Dr. Adil Haider (closing): Thank you very much. I’ll take the second question from Dr Stylianos first. Yes, it is true that at this time the FIM is validated for only those above 7 yrs of age. As 49% of our patients were 7 years old or less we were also concerned about this and did two things. Firstly
we utilized a variable in the NPTR that indicates “age-appropriate functional outcome at discharge”. Secondly instead of using the Likert type scaling of the FIM and analyzing the different gradations of functional impairment’s, we re-categorized all patients into two groups, those with any functional limitation and those with none. By using this binary variable in the analysis we were able to avoid or limit the above bias and also limit possible user/rater bias from different centers.

As to Dr. Stylianos’ first question, if we plot a histogram of ISS on the X axis and percentage of survivors and non-survivors with a given ISS on the Y axis we would find that when looking at the entire cohort we would see almost all the survivors grouped with low ISS. The non-survivors however are more spread out. Thus the scores predict surviving well but are not so good at grouping the non-survivors. However if we plot the same curve for the severely injured, as you would expect the X axis plot would not have such a peaked curve as the values are more evenly distributed between 25 and 75 the highest possible value for ISS. This diminishes it’s predictive capability; that is there is more overlap between the two groups. Given this, the reason why the NISS does better is that in the severely injured there are many patients with higher or discordant scores. Therefore when the same graph is plotted there are many more subjects grouped at the higher end of NISS values, allowing it better prediction.

In regards to Dr. Meredith’s questions: really from a purely mathematical point of view, calibration really applies to the measure of a scores ability to predict accurately across all deciles of probability from zero to one. With the ISS and NISS we are really looking at a scoring system and not a probability prediction, and therefore we chose not to use this measure. I wholeheartedly agree with the comment on how AIS are derived from mortality and may not accurately reflect functional limitations. That is why our group has been working on deriving systems based on functional outcome, which we hope to present in the near future.

I agree with Dr. Mooney’s suggestion about analyzing the interaction of TBI and outcomes prediction. Again as AIS scores are derived from mortality, we have been looking at different ways to grade head injury while predicting outcomes. To this effort we have been working on a Relative Head Injury Severity Score or RHISS and we may use that to follow up with Dr. Mooney’s suggestion.

I would like to thank the association for the privilege of the podium and the opportunity to present this work.