The New Injury Severity Score and the Evaluation of Pediatric Trauma

Enrique Grisoni, MD, Anthony Stallion, MD, Michael L. Nance, MD, Joseph L. Lelli, Jr., MD, Victor F. Garcia, MD, and Eric Marsh, MSN, RN

Background: To compare the effectiveness of the Injury Severity Score (ISS) and New Injury Severity Score (NISS) in predicting mortality in pediatric trauma patients.

Methods: NISS, the sum of the squares of a patient's three highest Abbreviated Injury Scale scores (regardless of body region), were calculated for 9,151 patients treated at four regional pediatric trauma centers and compared with previously calculated ISS values. The power of the two scoring systems to predict mortality was gauged through comparison of misclassification rates, receiver operating characteristic curves, and Hosmer-Lemeshow goodness-of-fit statistics.

Results: Although there were significant differences in mean NISS and ISS values for each hospital, differences in the predictive abilities of the two scoring systems were insignificant, even when analysis was restricted to the subgroup of patients with severe or penetrating injuries.

Conclusion: The significant differences in the predictive abilities of the ISS and NISS reported in studies of adult trauma patients were not seen in this review of pediatric trauma patients.

Key Words: Injury Severity Score, Pediatric trauma, Trauma scoring.

J Trauma. 2001;50:1106–1110.

There has been much discussion in recent years about the limitations of the Injury Severity Score (ISS) in predicting outcome and serving as a measure of the quality of trauma care. ISS, which allows the scorer to consider only one injury per body region, may lead to an underestimation of the patient's overall anatomic injury severity, because the patient's most severe injuries may not be represented. ISS also may be problematic in quality care assessments, since the test does not have a mechanism to differentiate severity of injury from mismanagement of injury before a patient's arrival at a trauma center.[1] Despite these concerns, ISS remains the most widely used injury severity scoring system, largely because an alternate method has not yet been found that both increases the accuracy of mortality predictions and justifies an industry-wide switch to a new system. A number of alternatives have been proposed, one of which involves a modification to the ISS system, rather than a wholesale shift to a new scoring method.

Submitted for publication June 27, 2000.

Accepted for publication January 15, 2001.

Copyright © 2001 by Lippincott Williams & Wilkins, Inc.

From the Department of Surgery (E.R.G., A.S.) and Rainbow Pediatric Trauma Center (E.M.), Rainbow Babies & Children's Hospital and Case Western Reserve University School of Medicine (E.R.G., A.S.), Cleveland, and Division of Pediatric Surgery, Children's Hospital Medical Center and University of Cincinnati College of Medicine (V.F.G.), Cincinnati, Ohio, Department of Surgery, Children's Hospital of Philadelphia and University of Pennsylvania School of Medicine (M.L.N.), Philadelphia, Pennsylvania, Department of Surgery, and C.S. Mott Children's Hospital and University of Michigan Medical School (J.L.L.), Ann Arbor, Michigan.

Address for reprints: Eric Marsh, MSN, RN, Rainbow Pediatric Trauma Center, Rainbow Babies & Children's Hospital, University Hospitals of Cleveland, 11100 Euclid Avenue, Cleveland, OH 44106-6039; email: eric.marsh@uhhs.com.

In 1997, Osler et al. proposed a modification of ISS that they believed would result in more accurate predictions of mortality.[2] Their modifications were simple and logical, intended to correct the limitations of the ISS system. The New Injury Severity Score (NISS) they proposed consists of the sum of the squares of the Abbreviated Injury Scale (AIS) scores of a patient's three most severe injuries, regardless of body region. In theory, NISS would more accurately represent the severity of injury and subsequent risk of mortality. In a retrospective study of 6,585 adult trauma patients, Osler et al. found NISS to be more predictive of outcome than ISS, with a lower misclassification rate and a better fit across the entire range of prediction. They recommended that NISS replace ISS as the standard summary measure of human trauma.

Another alternative to ISS, the *International Classification of Diseases, 9th Revision* (ICD-9) Injury Severity Score (ICISS) was developed by Osler et al. to remove the reliance on AIS and focus instead on ICD-9 injury categories.[3] Several subsequent studies supported the superiority of ICISS over ISS as a predictor of survival.[4–6] A recently published study by Sacco and colleagues, however, found that NISS outperformed both ISS and ICISS as a predictor of mortality.[7] A University of Toronto study comparing the performance of ISS and NISS also found that NISS performed significantly better as a predictor of mortality.[8]

These studies have looked exclusively at adult trauma patients, who traditionally present with a higher incidence of severe injuries—and more multisystem injuries—than pediatric patients. We undertook this study to determine whether these findings regarding the increased predictive abilities of NISS would hold true for pediatric trauma patients. We retrospectively examined records from four American College of Surgeons–verified Level I pediatric trauma centers: Rainbow Babies & Children's Hospital (Cleveland, OH), Children's Hospital Medical Center of Cincinnati (Cincinnati, OH), Children's Hospital of Philadelphia (Philadelphia, PA), and C.S. Mott Children's Hospital (Ann Arbor, MI). To ensure continuity with the initial NISS study and provide a basis for comparison, we replicated the methods used by Osler et al.[2]

PATIENTS AND METHODS

Cleveland, Cincinnati, Ann Arbor, and Philadelphia all have trauma registries maintained by registrars trained in Association for the Advancement of Automotive Medicine methodology. Each of the four trauma registries provides detailed information regarding injury type and mechanism, AIS scores, length of stay, intensive care unit length of stay, treatment received, and outcome. Registrars at each of these trauma centers reviewed patient records and submitted information on pediatric trauma patients under the age of 18. Cleveland's registrar provided records from 1993 through October 1999, and records from 1992 through August 1999 were submitted by Ann Arbor and Cincinnati and records for 1996 through October 1999 were received from Philadelphia.

These data were reviewed for completeness and accuracy before analysis. Those records for which any of the following fields were missing or incomplete were excluded from the study: age, gender, outcome, and AIS or ISS values. Included in the analysis were 2,007 patients from the Cleveland registry, 3,476 patients from the Cincinnati registry, 1,603 patients from the Ann Arbor registry, and 2,065 patients from the Philadelphia registry, for a total of 9,151. NISS values were calculated for each patient on the basis of the three highest AIS scores recorded.

The study compared NISS values calculated from patients' records with previously calculated ISS values. Comparisons included misclassification rates, receiver operating characteristic (ROC) curve analysis, and Hosmer-Lemeshow goodness-of-fit statistics. Data were analyzed using SPSS 10.0 software (SPSS, Inc., Chicago, IL). Data first were examined on a hospital-by-hospital basis to control for any differences in trauma scoring or patient demographics, and were then combined and reanalyzed. Two subsets of combined data also were analyzed: patients with injuries classified as serious (ISS/NISS of 15–24) and/or severe (ISS/NISS ≥ 25), and patients with penetrating injuries.

Misclassification rates were defined as the sum of false positives (survivors for whom death was predicted) and false negatives (nonsurvivors for whom survival was predicted) divided by the total number of cases. After a comparison of misclassification rates across the spectrum of NISS and ISS values, cut points were selected for each data group on the basis of the ISS and NISS values that minimized misclassifications. All patients with scores below the cut points were coded as predicted survivors, and patients with scores at or above the cut points were coded as predicted nonsurvivors. For the combined data set, cut points were established at 34 (ISS) and 44 (NISS).

The Hosmer-Lemeshow goodness-of-fit test provides a tool for measuring the relationship between predicted and observed outcomes. The test was well suited for our needs, since it allows for very small estimated probabilities (i.e., <0.2).[9] A higher p value (>0.05) indicates a significant relationship between the two variables (ISS or NISS and outcome). ROC curves also were used to compare the performance of ISS and NISS in predicting mortality. The area under the curve (C value) provides a measurement of each test's predictive value, with a value of 1 corresponding to a test that functions perfectly.

RESULTS

The survival rate for the Ann Arbor data set was 97%, and the remaining three data sets had survival rates of 98%. The survival rate for the combined data set was 98%. For all data groups, patients were predominantly elementary school-aged (mean ages ranged from 7.29 for Philadelphia to 10.64 for Ann Arbor) and boys outnumbered girls nearly 2:1. Blunt trauma was the primary mechanism of injury for Cleveland (94%), Cincinnati (95%), and Philadelphia (92%); mechanism of injury was not reported for the Ann Arbor data set.

NISS and ISS values differed for a significant number of patients in each group, with the largest proportion of differences occurring in the NISS ≤ 9 group. Mean ISS and NISS values for each data group are contrasted in Table 1. The results of a series of matched pairs Wilcoxon signed-rank tests showed significant differences in the distribution of NISS and ISS values for each data group; similar results were obtained when controlling for outcome. All mean scores, for both ISS and NISS, fell into the mild to moderate injury categories (mild, ISS/NISS ≤ 9; moderate, ISS/NISS of 10–14; serious, ISS/NISS of 15–24; and severe, ISS/NISS ≥ 25).

Less than 1% of patients in the combined data set and each individual data group received NISS values that resulted in a correct prediction of mortality that had not been predicted by ISS. As further illustrated in the comparison of misclassification rates (Table 2), differences in ISS and NISS values ultimately resulted in little difference in the predictive ability of NISS, as compared with ISS. Misclassification rates for ISS and NISS at the assigned cut points did not differ significantly for the combined data group or any of the four population samples. ISS misclassification rates were slightly

Table 1 Mean ISS and NISS Values

Data Group	Mean ISS	SD	Mean NISS	SD
Combined	8	8	10	10
Cleveland	7	8	8	10
Cincinnati	8	7	9	8
Ann Arbor	10	9	11	10
Philadelphia	9	8	12	13

Table 2 Statistical Comparison of ISS and NISS

Data Group	Test Type	ISS	NISS
Combined	Misclassification rates (%)	2.1	2.1
	ROC	0.934	0.930
	ROC 95% CI	0.907–0.961	0.903–0.957
	Hosmer-Lemeshow	0.43	1.081
	p value	>0.995	>0.995
Cleveland	Misclassification rates (%)	1.4	1.5
	ROC	0.989	0.986
	ROC 95% CI	0.982–0.997	0.976–0.997
	Hosmer-Lemeshow	3.02	2.42
	p Value	0.95	0.975
Cincinnati	Misclassification rates (%)	1.5	1.5
	ROC	0.981	0.978
	ROC 95% CI	0.973–0.990	0.969–0.988
	Hosmer-Lemeshow	2.969	0.022
	p value	0.9	>0.995
Ann Arbor	Misclassification rates (%)	2.5	2.7
	ROC	0.923	0.919
	ROC 95% CI	0.859–0.986	0.855–0.983
	Hosmer-Lemeshow	0.858	0.331
	p value	>0.995	>0.995
Philadelphia	Misclassification rates (%)	3	3.3
	ROC	0.839	0.837
	ROC 95% CI	0.761–0.917	0.758–0.916
	Hosmer-Lemeshow	0.645	0.210
	p value	>0.995	>0.995

CI, confidence interval.

lower than NISS rates for all data groups, with differences ranging from 0.0003 to 0.003.

Graphing frequency distributions also failed to reveal any sharp distinctions in the predictive abilities of ISS and NISS. As Figures 1 and 2 illustrate, both scoring systems provide good separation between survivors and nonsurvivors. The differences between the median ISS or NISS scores for survivors and analogous scores for nonsurvivors were similar in all samples, with NISS reporting slightly greater differences. The separation of values (median value for nonsurvivors minus median value for survivors) in the combined data group was 24 for ISS and 35 for NISS. Although the study of Osler et al.[2] found a doubling of the separation of median ISS and NISS values in adult trauma patients, this study did not reveal a similar pattern for pediatric trauma patients.

ROC curve values were similar for NISS and ISS in all five data groups. In each group, the area under the ROC curve was identical for both ISS and NISS: 84% for Philadelphia, 92% for Ann Arbor, 98% for Cincinnati, 99% for Cleveland, and 93% for the combined data set (Table 2). Philadelphia's lower ROC value resulted from a number of outliers in both directions (nonsurvivors with low ISS/NISS values and survivors with high ISS/NISS values) and had an impact on the ROC value of the combined data group. For each data set, the ISS group had an ROC value that was 0.2% to 0.4% higher than the NISS group. Thus, the predictive value of ISS was slightly higher than NISS, but not to a significant degree.

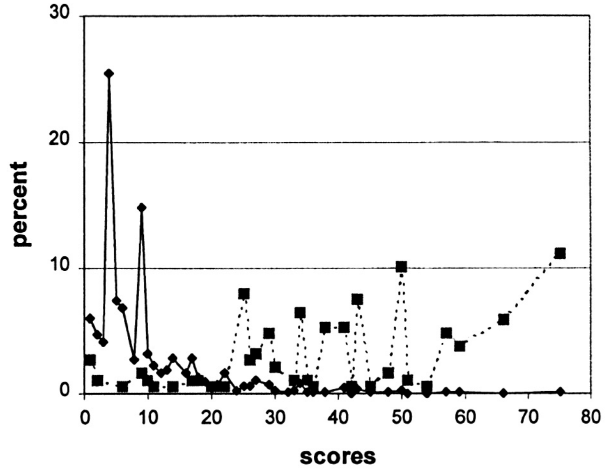

Fig. 1. *Frequency distribution of NISS values for the combined data group;* diamonds, *survivors;* squares, *nonsurvivors.*

The Hosmer-Lemeshow tests reported good fit for all data groups, for both ISS and NISS values. Hosmer-Lemeshow statistics were low and *p* values were high for all samples. Although the Hosmer-Lemeshow statistics reflected slightly better fit for NISS in all hospital data groups, the test revealed slightly better fit for ISS in the combined data group. The Hosmer-Lemeshow test divides data into deciles for analysis; the seeming incongruity between the Hosmer-Lemeshow statistics for the individual hospitals and the combined group springs from the significantly larger number of data points in the combined group and the resultantly larger deciles. Still, differences between Hosmer-Lemeshow values for ISS and NISS were small in every case (<1) and *p* values exceeded 95% for all data groups, indicating good fit for both ISS and NISS.

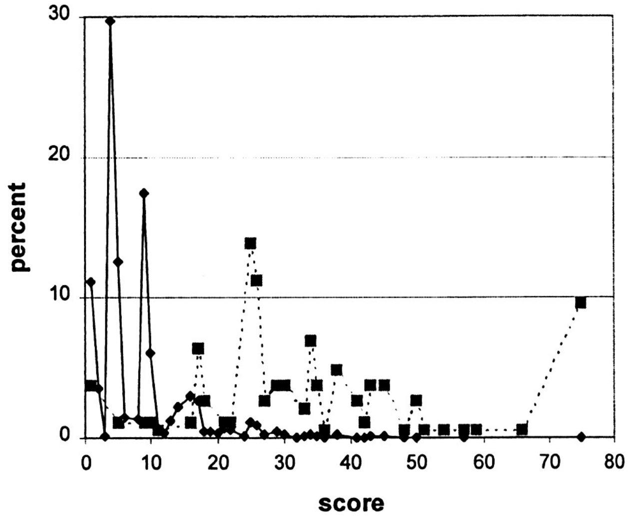

Fig. 2. *Frequency distribution of ISS values for the combined data group;* diamonds, *survivors;* squares, *nonsurvivors.*

Table 3 Statistical Comparison of ISS and NISS in More Injured Subgroups

Test Type	ISS ≥ 15	NISS ≥ 15	ISS ≥ 25	NISS ≥ 25
Cut point	44	60	44	60
Misclassification rate (%)	12.0	12.2	25.8	26.4
ROC	84% (0.824)	83% (0.830)	67% (0.671)	67% (0.667)
ROC 95% CI	0.793–0.855	0.798–0.862	0.616–0.726	0.611–0.722
Hosmer-Lemeshow statistic	21.828	30.145	183.380	105.790
p value	<0.01	<0.001	<0.001	<0.001

CI, confidence interval.

Unlike adult trauma patient demographics, the vast majority of pediatric patients in this study were only mildly or moderately injured (86%) and had nonpenetrating injuries (95%). To determine whether these differences could account for the variation in results for pediatric and adult trauma patients, subgroups limited to seriously and/or severely injured patients were analyzed, as was the subgroup of patients with penetrating trauma. As Table 3 illustrates, both ISS and NISS performed poorly in predicting outcome when the data set was limited to patients with a greater degree of injury, but there remained no significant difference in the performance of the two scoring methods. Similarly, although both ISS and NISS performed well in predicting outcome for patients with penetrating trauma (Table 4), no significant differences were noted in the predictive abilities of the two scoring methods.

DISCUSSION

The fundamental differences between NISS and ISS calculation ensure that many NISS values will differ from the ISS values obtained for the same patient. The question is whether these differences have a significant impact on the test's predictive value. In this study, differences in the predictive power of ISS and NISS were too small to lead to the conclusion that either scoring system outperformed the other.

One difficulty in applying NISS to a pediatric trauma population is that, in comparison with adult trauma patients, mortality predictions are less complex, the majority of patients are less severely injured, and misclassification rates are lower—regardless of the method of prediction. In other words, both ISS and NISS may be more accurate in predicting mortality in pediatric populations than in adult populations. Reflecting this difference between adult and pediatric patients, ROC curves were stronger for the pediatric patient population than for the adults referred to in other studies. Whereas the C value for the combined data set was 93%, Brenneman and colleagues reported C values of 80% for ISS and 85% for NISS, and Osler and colleagues reported C values ranging from 87% to 91%. The results of the Hosmer-Lemeshow goodness-of-fit tests further supported the conclusion that there was little difference in the two scoring systems' predictive abilities; both tests performed strongly in predicting mortality in pediatric trauma patients. Even when controlling for some of the key differences between pediatric and adult trauma patient populations—average severity of injury and frequency of penetrating trauma—no significant differences in the predictive performance of ISS and NISS emerged.

Table 4 Statistical Comparison of ISS and NISS with Subgroup of Patients with Penetrating Injuries

Test Type	ISS	NISS
Cut point	34	66
Misclassification rate (%)	2.9	3.5
ROC	99% (0.986)	97% (0.973)
ROC 95% CI	0.976–0.996	0.958–0.988
Hosmer-Lemeshow statistic	0	0.992
p value	>0.995	>0.995

CI, confidence interval.

CONCLUSION

In evaluating sample pediatric trauma populations, differences between ISS and NISS were minimal. Both tests had low misclassification rates and low Hosmer-Lemeshow statistics, generated high ROC values, and showed a strong correlation between score and outcome, all of which indicate that both scoring systems perform well in predicting mortality for pediatric trauma patients. Although NISS calculations led to a large number of values that differed from previously calculated ISS values, the usefulness of those differences in more accurately predicting outcome was not proven. The results of this study indicate that NISS does not provide a significant improvement over ISS when evaluating pediatric trauma patients. In light of these findings, it may be worthwhile to compare the predictive values of ISS and ICISS in a study of pediatric trauma patients, to determine whether this proposed alternative to ISS, which has been tested only on adult patients, also performs differently in a pediatric population.

REFERENCES

1. Rutledge R. The Injury Severity Score is unable to differentiate between poor care and severe injury. *J Trauma.* 1996;40:944–950.
2. Osler T, Baker SP, Long W. A modification of the injury severity score that both improves accuracy and simplifies scoring. *J Trauma.* 1997;43:922–925.
3. Osler T, Rutledge R, Deis J, Bedrick E. ICISS: an international classification of disease-9 based injury severity score. *J Trauma.* 1996;41:380–386.
4. Rutledge R, Osler T, Emery S, Kromhout-Schiro S. The end of the Injury Severity Score (ISS) and the Trauma and Injury Severity Score (TRISS): ICISS, an International Classification of Diseases,

ninth revision-based prediction tool, outperforms both ISS and TRISS as predictors of trauma patient survival, hospital charges, and hospital length of stay. *J Trauma.* 1998;44:41–47.
5. Rutledge R, Osler T. The ICD-9-based illness severity score: a new model that outperforms both DRG and APR-DRG as predictors of survival and resource utilization. *J Trauma.* 1998;45:791–799.
6. Hannan EL, Farrell LS, Gorthy SF, et al. Predictors of mortality in adult patients with blunt injuries in New York State. *J Trauma.* 1999;47:8–14.
7. Sacco WJ, MacKenzie EJ, Champion HR, Davis EG, Buckman RF. Comparison of alternative methods for assessing injury severity based on anatomic descriptors. *J Trauma.* 1999;47:441–446.
8. Brenneman FD, Boulanger BR, McLellan BA, Redelmeier DA. Measuring injury severity: time for change? *J Trauma.* 1998; 44:580–582.
9. Hosmer DW, Lemeshow S. *Applied Logistic Regression.* New York, NY: John Wiley & Sons; 1989:142–143.

EDITORIAL COMMENT

The Injury Severity Score (ISS) has been in use for almost 30 years. Although it is a system of injury stratification based primarily on anatomic diagnosis, it has been demonstrated to be an accurate predictor of both survival and postinjury impairment. It has been used alone and in combination with various measures to identify cases that warrant audit as potentially preventable deaths. Its applicability to the pediatric population has, likewise, been widely accepted as a reasonable method for categorization of injury severity. The New Injury Severity Score (NISS) seems like a logical modification that simply sums the squares of the three worst injuries, regardless of location. This should be especially applicable to childhood injury, which, in its most severe form, usually involves combinations of brain injury and cranial fractures. On the basis of their data from four pediatric institutions, Dr. Grisoni and his colleagues make a reasonably cogent case that when used to predict mortality, the NISS offers nothing beyond the established capability of the ISS. Should the NISS, therefore, be summarily dismissed? Unfortunately, the data presented in this report cannot answer that question completely. By limiting their analysis to mortality, the authors ignore any potential benefit of the NISS as a predictor of other, increasingly relevant indicators of outcome. Impairment status at discharge, incidence of nosocomial infection, ventilator days, and length of stay in intensive care are just a few outcome measures that may be more relevant from a therapeutic perspective and should be considered in comparing ISS to NISS. We hope that the authors will reexamine their data to address these or similar measures. If, at that point, no benefit to the NISS can be demonstrated, it should indeed be relegated to the list of concepts that sounded better than they worked.

Joseph J. Tepas III, MD
Professor of Surgery
University of Florida
Health Science Center Jacksonville
Jacksonville, Florida