Published in final edited form as:

J Trauma Acute Care Surg. 2018 August ; 85(2): 334–340. doi:10.1097/TA.0000000000001943.

A New Weighted Injury Severity Scoring System: Better Predictive Power for Pediatric Trauma Mortality

Junxin Shi, MD, PhD^{a,b}, Jiabin Shen, PhD^{a,b}, Sarah Caupp, MS^{a,b}, Angela Wang^{a,b}, Kathryn E. Nuss, MD^c, Brian Kenney, MD^{a,d}, Krista K. Wheeler, MS^{a,b}, Bo Lu, PhD^e, and Henry Xiang, MD, MPH, PhD^{a,b,f}

^aCenter for Pediatric Trauma Research, The Research Institute at Nationwide Children's Hospital, Columbus, OH

^bCenter for Injury Research and Policy, The Research Institute at Nationwide Children's Hospital, Columbus, OH

^cDepartment of Emergency Medicine, Nationwide Children's Hospital, Columbus, OH

^dDepartment of Pediatric Surgery, The Ohio State University College of Medicine, Columbus, OH

^eThe Ohio State University College of Public Health, Division of Biostatistics, Columbus, OH

^fThe Ohio State University College of Medicine, Columbus, OH

Abstract

Background—An accurate injury severity measurement is essential for the evaluation of pediatric trauma care and outcome research. The traditional Injury Severity Score (ISS) does not consider the differential risks of the Abbreviated Injury Scale (AIS) from different body regions nor is it pediatric specific. The objective of this study was to develop a weighted injury severity scoring (wISS) system for pediatric blunt trauma patients with better predictive power than ISS.

Methods—Based on the association between mortality and AIS from each of the six ISS body regions, we generated different weights for the component AIS scores used in the calculation of ISS. The weights and wISS were generated using the National Trauma Data Bank (NTDB). The Nationwide Emergency Department Sample (NEDS) was used to validate our main results. Pediatric blunt trauma patients less than 16 years were included, and mortality was the outcome. Discrimination (areas under the receiver operating characteristic curve, sensitivity, specificity, positive predictive value, negative predictive value, concordance) and calibration (Hosmer-Lemeshow statistic) were compared between the wISS and ISS.

CORRESPONDING AUTHOR: Henry Xiang, MD, MPH, PhD, Center for Pediatric Trauma Research, The Research Institute at Nationwide Children's Hospital, 700 Children's Drive, Columbus, OH 43205; Henry.Xiang@NationwideChildrens.org; Phone: 614-355-5893; Fax: 614-355-5897.

Conflicts of Interest: The authors report no conflicts of interest.

Author Contributions

Literature search (Shi and Caupp), study design (Shi and Shen), data analysis (Shi and Shen), data interpretation (all authors), writing (Shi: 1st draft), critical revision (all authors), and final approval (all authors). Angela Wang contributed to the study while conducting her summer research at the Center for Pediatric Trauma Research and the Center for Injury Research and Policy, the Research Institute at Nationwide Children's during the summer of 2017.

Results—The areas under the receiver operating characteristic curves from the wISS and ISS are 0.88 vs. 0.86 in ISS=1–74 and 0.77 vs. 0.64 in ISS=25–74 ($p<0.0001$). The wISS showed higher specificity, positive predictive value, negative predictive value, and concordance when they were compared at similar levels of sensitivity. The wISS had better calibration (smaller Hosmer-Lemeshow statistic) than the ISS (11.6 versus 19.7 for ISS=1–74 and 10.9 versus 12.6 for ISS=25–74). The wISS showed even better discrimination with the NEDS.

Conclusions—By weighting the AIS from different body regions, the wISS had significantly better predictive power for mortality than the ISS, especially in critically injured children.

Level of Evidence and study type—Level IV Prognostic/Epidemiological

Keywords

Injury Severity Score; Weighting; Pediatric; Mortality

Background

An accurate injury severity measurement is essential for the evaluation of pediatric trauma care and outcome research. Over the past four decades, the Injury Severity Score (ISS), created by Susan Baker in 1974, (1) has been the most commonly used injury severity measurement in pediatric literature.(2) ISS is based on the Abbreviated Injury Scale (AIS), which is an anatomically based consensus-derived global severity scoring system that classifies each injury by body region according to its relative severity on a six-point ordinal scale: 1 (Minor), 2 (Moderate), 3 (Serious), 4 (Severe), 5 (Critical), and 6 (Maximal).(3) An ISS (ranging from 1 to 75) is computed as the squared sum of the highest AIS severity scores from the three different most severely injured body regions, with the exception that any AIS score of 6 is automatically assigned an ISS of 75.

Although ISS has been the standard injury severity scale for over 40 years, it has a critical limitation, namely the fact that it is weighted equally across all body regions (Head/neck, Face, Chest, Abdomen or pelvic contents, Extremities or pelvic girdle, and External) without taking into account the differential mortality risks, leading to an inconsistent correlation between ISS and mortality.(4–7) This effect may be more prominent in pediatric trauma; since the AIS was developed for patients of all ages, and there are no AIS scores specifically designed for pediatric patients. Brown and colleagues reported that only the severity of head and chest injuries were associated with mortality among pediatric trauma patients with single-system injury; in pediatric patients with multisystem injury, only head, face, neck, and torso injuries were associated with mortality.(8) Traumatic brain injury as a primary cause of death among pediatric trauma patients has also been noted by other researchers.(9, 10) Each of these studies shows that an injury with the same AIS score may display large discrepancies in mortality risks depending on the patient's age and developmental status. In summary, the conventional ISS score gives equal weighting to different body regions, and an improved severity measurement should also take into account the effect of age differences on the severity in patients.

Since the inception of the ISS scoring system, a number of new injury severity scoring tools have been developed, including the New ISS (NISS),(11) the Trauma and Injury Severity Score (TRISS),(12) and the International Classification of Diseases (ICD)-based Injury Severity Score (ICISS).(4) However, none of these scoring systems fully address the above-mentioned critical limitations of ISS, and none of them were designed or validated for use in the pediatric trauma population. Because ISS is a widely used scale in pediatric trauma, addressing these critical limitations is important. This study aims to bridge this gap using a two-step procedure to develop and validate a new weighted-ISS scoring system for pediatric blunt trauma patients. We used the National Trauma Data Bank (NTDB) to find the best set of weights to calculate a weighted ISS, and then validated these results using another national database (the Nationwide Emergency Department Sample, NEDS).

Methods

Study population

The ISS was originally designed for blunt trauma. We limited our study population to pediatric patients with blunt trauma (age<16 years). Patients were excluded: penetrating trauma; burn patients; patients who were transferred to another hospital; patients who arrived without signs of life; and patients with ISS=75. Those transferred out were excluded because they may not receive thorough or complete evaluations prior to transfer and their survival is unknown. Patients with an ISS=75 were excluded since their probability of death is believed to be close to 1. Death is the main outcome of the present study and is defined as death occurring in the emergency department (ED) or during the hospital stay.

Databases

Two databases were used: the National Trauma Data Bank (NTDB) Research Datasets (design database) and the Nationwide Emergency Department Sample (NEDS) (validation database). The NTDB contains standardized trauma registry data from more than 900 trauma centers in the United States, and is the largest aggregation of U.S. trauma registry data. We used 8 years of NTDB Research Datasets (2007–2014) which contains more than 6 million records.(13) NEDS is the largest all-payer ED database in the United States, containing information from about 30 million ED visits at 950 hospitals annually, which approximates a 20-percent stratified sample of U.S. hospital-based ED visits.(14) NEDS also includes information for the hospital stay if the patient was admitted directly from the ED. We used six years of NEDS data (2009–2014).

Data analysis

All the analyses were done using SAS Enterprise Guide, Version 7.11 HF3 (SAS Institute Inc., Cary, NC, USA). There are four major steps in the data analysis.

Step one: Calculation of the Weighted ISS

The weights for the AIS from the six ISS body regions should reflect the strength of the relationship between the AIS and mortality. We explored nine sets of weights which can be divided into three method groups: Group A is based on AUC (the area under the receiver operating characteristic curve); Group B is based on the concordance (among all possible

combinations of one event and one non-event, the concordance is defined as the proportion of pairs where the observation with event has a higher predicted probability than the observation of non-event); and Group C is based on mortality from each body region. To expand the scope of weighting methods, we applied two data transformation algorithms, the logarithm and square root, in Group A and Group B. The details are as follows:

1. By using a Stata package named ICDPIC,(15) we translated ICD-9-CM diagnosis codes (International Classification of Diseases, Ninth Revision, Clinical Modification) into six maximum AIS severity scores (maxAIS) corresponding to the six ISS body regions.
2. We fitted six logistic models with death as the outcome, and each of the six maxAIS as predictors to get the c statistic (AUC) and the concordance. To make sure the current maxAIS was from the principal injury, in each model we chose patients with the current maxAIS as the largest AIS score and maxAIS scores from other body regions that were smaller.
3. We generated nine sets of weights. Specifically,

A1: $100 \times (\text{AUC} - 0.5)$;	A2: $\text{LOG}(100 \times (\text{AUC} - 0.5))$;	A3: $\text{SQRT}(100 \times (\text{AUC} - 0.5))$;
B1: Concordance;	B2: $\text{LOG}(\text{Concordance})$;	B3: $\text{SQRT}(\text{Concordance})$;
C1: Weighted mortality rate;	C2: Highest mortality rate; and	C3: Overall mortality rate.

In A1, A2, and A3, we subtracted 0.5 from the AUC to get a “net” discrimination ($\text{AUC} = 0.5$ is considered without any discrimination). In C1, we calculated a weighted mortality rate for each body region. This method takes into account the fact that, for injuries of a specific body region, the highest mortality rate for an individual AIS value may be generated from a very small number of patients. We did not want these potentially unreliable mortality rates to be overly representative of the injury mortality level for this body region. This weighted mortality rate is the weighted average of the mortality rates of individual AIS values by the proportion of the mortalities in each individual AIS value.

4. We applied (multiplied) the weights to the six maxAIS scores. We then squared the three largest numbers and added them together to get a weighted ISS. We also calculated the traditional ISS in the same way without weighting.

Step two: Discrimination Comparisons

Discrimination in this study was defined as the ability of the score to separate the patients who survived and those who did not. In a Receiver Operating Characteristic (ROC) curve, the true positive rate (sensitivity) is plotted as a function of the false positive rate (1-specificity) for different cut-off points.(16) The area under ROC curve (AUC) equal to 1 represents perfect discrimination; an AUC equal to 0.5 indicates no discriminative power. We compared the AUCs of all nine weighting methods against the traditional ISS among various ranges of ISS by fitting logistic regression models. Ultimately, we chose the weighting method that maximized the AUC in critically injured children (ISS 25–74). Since logistic regression is a monotonic transformation of the severity score, the c statistic from

the logistic regression model is equivalent to the AUC calculated from the raw score. In the following discussions, we will use the acronym “wISS” to refer to our chosen weighted ISS.

We also calculated sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and concordance with each individual score as the cutoff value across the full range of severity scores (ISS 1–74) and in critically injured patients (ISS 25–74). For ease comparison, we reported results from cutoff values at which roughly 50%, 70%, and 90% of deaths could be identified by wISS and ISS.

Step three: Calibration Comparison

Calibration is the ability of predictors to correctly predict an outcome over the entire range of risk. Calibration can be assessed graphically by plotting the actual outcome against the predictors. In our study, the ISS has 42 individual values while the wISS has over 1,000 individual values. The linearity of the mortality rates in correspondence with the severity scores also depends on the number of patients in each score group. If sample sizes for the groups are too small, the line representing the mortality rates will oscillate wildly. If the number of groups is made smaller (i.e., the number of patients in each group will then be larger), the mortality rates tend to be stabilized and are less likely to be affected by the number of patients in each score group. We generated scatter plots of mortality probabilities against wISS for different score groupings (i.e., on the original scale, with unit=1, unit=5, and by 6 groups). For comparison purpose, we generated similar graphs for the ISS, on the original scale and by 6 groups. With the assumption that logistic regression is the correct model to describe the relationship between mortality and severity score, we could compare the calibration indirectly by comparing the goodness-of-fit of the logistic models, most commonly the Hosmer–Lemeshow statistic (HL), where smaller HL values indicates better fit.

Step four: External Validation

We used another national database, NEDS, to validate the main conclusion generated by our study database, the NTDB. With NEDS data, we also selected pediatric blunt trauma patients (by excluding cut and firearm related injuries). We excluded those transferred out from either the ED or from the hospital. A weighted ISS was calculated using the set of weights generated from the NTDB data. Discrimination performance was also measured using the AUC.

Results

A total of 545,015 pediatric blunt trauma patients were identified in the 2007–2014 NTDB. Sixty-five percent had only minor injuries (ISS 1–8). There were 14,056 (2.6%) critical injuries (ISS 25–74) that accounted for 50% of all deaths. The overall mortality rate was 0.98%, while in critically injured patients the mortality rate jumped to 18%. Nonlinearity is much more pronounced in the range of ISS 25 (eTable 1).

In patients with ISS 1–74, the AUC from the traditional ISS is 0.86 (95% CI: 0.86–0.87), while the AUC from weighting method C3 is 0.89 (95CI: 0.87–0.88) ($p<0.0001$) (the weights for C3 are: head/neck=1.87; face=0.13; chest=1.52; abdomen/pelvic contents=0.98;

extremities=0.15; external=0.33). The improvement in discrimination is much more substantial for patients with an ISS in the range of 25–74 (the AUC goes from 0.64 to 0.77), with a net discrimination improvement near 100% ($97\%=(0.7732-0.6386)/(0.6386-0.5)$). There were discrimination improvements in all ISS ranges except in the ISS range of 16–24 (eTable 2, and Table 1).

Data presented in Table 2 shows that with similar sensitivity, the specificity, positive predictive value, negative predictive value, and concordance are all higher for the wISS compared with ISS in critically injured patients (ISS 25–74). While the differences across the whole range of severity (ISS 1–74) are not dramatic, improvements are seen for the critically injured. Results generated from all possible cutoff values are shown in the Fig 1.

To compare calibration or how well the severity scores are associated with mortality, we analyzed the data in two ways: (1) by drawing scatter plots for mortality rates against the scores, and (2) by fitting logistic models and comparing the goodness-of-fit of models. Fig 2 shows scatter plots of mortality rates (probabilities) against severity scores. From these graphs, we can see that the trends were clearer with a decrease in number of score groups. The wISS and ISS both produced good trends when there were only six score groups. It is difficult to say, which one is better in predicting mortality just by checking these graphs visually. However, the Hosmer-Lemeshow (HL) statistic is smaller for the wISS model than for ISS (11.6 vs. 19.7 in ISS 1–74, and 10.9 vs. 12.6 in ISS 25–74). In Fig 3, we have depicted observed mortality rate along with the predicted mortality rates from ISS and from the wISS. All three sets of rates are aggregated by individual ISS. The predicted mortality rates from the wISS are much closer to the observed mortality rates compared with ISS, especially in severely injured patients. We saw a non-smooth curve representing the predicted mortality rates from the wISS because a higher ISS does not always correspond with a higher wISS. But the most important thing is to evaluate which one is closer to observed data points; in this case, it is notably the wISS. Assuming the logistic regression models are correct, we could say the wISS has a stronger association with mortality than ISS.

In 2009–2014 NEDS, a total 1,302,348 pediatric blunt trauma patients were selected, 843 died, with overall mortality rate as 0.06%. When we applied the same weighting method to the NEDS data, we found similar results. The AUCs are greater for wISS than for ISS across in the whole ISS range (1–74), and the difference is markedly pronounced in critically injured children (ISS $\geq$ 25). The wISS showed an even better discrimination improvement than it manifested in the NTDB. (Table 1)

Discussion

This study developed a weighted Injury Severity Score (wISS) by applying data-derived weights to consensus-derived AIS scores in the calculation of the ISS (weights: head/neck=1.87; face=0.13; chest=1.52; abdomen/pelvic contents=0.98; extremities=0.15; external=0.33). In a large pediatric blunt trauma sample, we found that the wISS had greater mortality predictive power than the traditional ISS, especially in critically injured children.

External validation showed the robustness and the potential generalizability of the main conclusion of our study.

In the present study, we explored nine methods of weighting. All methods exhibited the same pattern: AIS scores from the head/neck, chest, and abdomen/pelvic contents were weighted more than the AIS scores from the face, extremities, and external body. The results confirmed previous studies.(8, 10, 17) Brown and colleagues reported that maximum head AIS was the strongest predictor of mortality among both single-system and multi-system injury, and chest injuries were the next most deadly injuries, particularly among children with single-system isolated thoracic injury(8) Our study and previous studies suggest that injuries to the head/neck, chest, and abdomen/pelvic contents are more important in predicting survival than injuries to the face, extremities, and external body in pediatric trauma patients. Hence, more weight should be given to the AIS scores from these body regions when calculating injury severity scores, as we did when computing the wISS in this present study. This also foretold that the wISS should have better predictive power than the traditional ISS.

The AUC is the most commonly used measurement of discrimination.(2) Most studies include patients of all levels of injury severity, and the majority will most likely have minor injuries. Given this patient mix, the AUC from many scoring systems tends to be very high. In a study conducted by Meredith and colleagues, they compared nine scoring algorithms in predicting mortality, and their conclusion was that the differences in performance were relatively small. In their study, the AUC for ISS was 0.867. In our study, this number is 0.86 for NTDB data, and in NEDS it is 0.91. The AUC can be intuitively understood in this way: if there is a pair of patients where one is randomly selected from the non-survival group and the other is randomly selected from the survival group, the AUC is equivalent to the probability of correctly identifying the non-surviving patient as the one with a higher severity score and the surviving patient as the one with a lower severity score (correctly classifying the two patients in the random pair). We can see here that if the majority of patients have minor injuries, it is not too difficult to correctly guess which one has a higher severity score and which one has a lower severity score, since the difference in terms of severity between these two patient pools is large. In our study dataset from the NTDB, the median ISS in surviving patients is 7 while in non-surviving patients the median ISS is 25. This is why many researchers throughout the years have tried to develop alternative severity scores but ultimately failed to improve on the already very high AUC of the traditional ISS. Most studies did not have sufficient data to conduct their studies on critically injured patients. Since the traditional ISS has very good discriminative power for mortality in terms of the AUC (almost always approaching 0.9), it is difficult to make improvement upon with this patient case mix. The real challenge and most meaningful one is discriminating among critically injured patients, and our weighted ISS performed much better than the traditional ISS in critically injured children (ISS 25). In this ISS range, the ISS has very limited discrimination (AUC=0.64) while the weighted ISS has an AUC>0.77. A greater difference was observed when we applied the weights to the NEDS data, with the AUC improving from 0.49 to 0.80. The similarities manifested in these two national databases strongly support our statement that the weighted ISS significantly improves upon the predictive capacity of mortality in critically injured children when compared with the traditional ISS.

The results from our study have implications for risk adjustment in trauma outcome research and trauma program evaluation. This weighted ISS is a better alternative to ISS to predict mortality in pediatric trauma patients, especially when studying the most severe patients. The wISS could also be used to risk-adjust mortality in quality improvement efforts.

Study Limitations

Although this study successfully developed a new weighted ISS scoring system that performs better than the traditional ISS, our study has several limitations. First, as mentioned in the methods section, ISS was originally designed for blunt trauma, and the weighted ISS in our study was developed for pediatric blunt trauma. The generalizability of our results to other types of pediatric trauma or adult trauma patients remains unknown. However, the rationale of weighting and the implementation methods can be applied in other trauma populations (e.g., pediatric penetrating trauma or adult trauma patients). Second, the AIS scores used in this study were generated with computer software. Previous researchers reported that there are discrepancies between machine assigned and manually assigned AIS. (18) The impact of this discrepancy on our study findings is unknown. Third, we did not conduct separate analyses for isolated injuries involving only one body system and multi-system injuries. Finally, the calculation of the weighted ISS is more cumbersome than the traditional ISS, and most often will require the use of a computer. However, the computer algorithm is provided here in the supplemental materials and can easily be incorporated into trauma datasets.

Conclusion

In summary, by weighting AIS scores from different body regions, we developed a weighted Injury Severity Score (wISS) system for pediatric blunt trauma patients. The weighted ISS has a significantly better predictive power for mortality than the traditional ISS, especially in critically injured children. The wISS can be used as a better alternative in evaluation of pediatric trauma care and outcome research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Sources of Funding: Sources of funding include: Agency for Healthcare Research and Quality (grant #1R01 HS24263-01) and Health Resources and Services Administration (grant #R40MC30759).

References

1. Baker SP, O'Neill B, Haddon W Jr, Long WB. The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. *J Trauma*. 1974; 14(3):187–96. [PubMed: 4814394]
2. Tohira H, Jacobs I, Mountain D, Gibson N, Yeo A. Systematic review of predictive performance of injury severity scoring tools. *Scand J Trauma Resus*. 2012:20.
3. Rating the severity of tissue damage. I. The abbreviated scale. *JAMA*. 1971; 215(2):277–80. [PubMed: 5107365]

4. Osler T, Rutledge R, Deis J, Bedrick E. ICISS: an international classification of disease-9 based injury severity score. *J Trauma*. 1996; 41(3):380–6. discussion 6–8. [PubMed: 8810953]
5. Aharonson-Daniel L, Givon A, Stein M, Israel Trauma G, Peleg K. Different AIS triplets: Different mortality predictions in identical ISS and NISS. *J Trauma*. 2006; 61(3):711–7. [PubMed: 16967012]
6. Furnival RA, Schunk JE. ABCs of scoring systems for pediatric trauma. *Pediatr Emerg Care*. 1999; 15(3):215–23. [PubMed: 10389962]
7. Copes WS, Champion HR, Sacco WJ, Lawnick MM, Keast SL, Bain LW. The Injury Severity Score revisited. *J Trauma*. 1988; 28(1):69–77. [PubMed: 3123707]
8. Brown JB, Gestring ML, Leeper CM, Sperry JL, Peitzman AB, Billiar TR, Gaines BA. The value of the injury severity score in pediatric trauma: Time for a new definition of severe injury? *J Trauma Acute Care Surg*. 2017; 82(6):995–1001. [PubMed: 28328674]
9. Heron M, Sutton PD, Xu J, Ventura SJ, Strobino DM, Guyer B. Annual summary of vital statistics: 2007. *J Pediatr*. 2010; 125(1):4–15.
10. Schneier AJ, Shields BJ, Hostetler SG, Xiang H, Smith GA. Incidence of pediatric traumatic brain injury and associated hospital resource utilization in the United States. *J Pediatr*. 2006; 118(2): 483–92.
11. Osler T, Baker SP, Long W. A modification of the injury severity score that both improves accuracy and simplifies scoring. *J Trauma Inj Infect Crit Care*. 1997; 43(6):922–5.
12. Champion HR, Sacco WJ, Hunt TK. Trauma severity scoring to predict mortality. *World J Surg*. 1983; 7(1):4–11. [PubMed: 6837064]
13. [Accessed date: December-15-2017] NTDB Research Data Set [Internet]. 2007–2014. Available from: <https://www.facs.org/quality-programs/trauma/ntdb/datasets>
14. The Healthcare Cost and Utilization Project (HCUP). The Nationwide Emergency Department Sample (NEDS). 2006–2014.
15. Clark DE, Osler TM, Hahn DR. ICDPIC: Stata module to provide methods for translating International Classification of Diseases (Ninth Revision) diagnosis codes into standard injury categories and/or scores. 2009
16. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem*. 1993; 39(4):561–77. [PubMed: 8472349]
17. Cooper A, Barlow B, DiScala C, String D. Mortality and truncal injury: the pediatric perspective. *J Pediatr Surg*. 1994; 29(1):33–8. [PubMed: 8120758]
18. Di Bartolomeo S, Tillati S, Valent F, Zanier L, Barbone F. ISS mapped from ICD-9-CM by a novel freeware versus traditional coding: a comparative study. *Scand J Trauma Resusc Emerg Med*. 2010; 18:17. [PubMed: 20356359]

Key Points**Question**

Will a weighted injury severity score (wISS) have better predictive power in pediatric blunt trauma patients when compared with the traditional injury severity score (ISS)?

Findings

By weighting the abbreviated injury scores (AIS) from different body regions, we produced the wISS. The wISS showed better mortality discrimination and more accurate calibration than the ISS, especially in children with ISS=25–74.

Meaning

The wISS has a significantly better predictive power for mortality than the traditional ISS in critically injured children.

Figure 1.

Comparing specificity, positive predictive value (PPV), negative predictive value (NPV), and concordance against sensitivity for the weighted injury severity score (wISS) and the injury severity score (ISS), in whole range of severity (ISS:1–74) and in critically injured patients (ISS=25–74), NTDB 2007–2014.

Figure 2.

Observed mortalities by wISS and by ISS on different scales, NTDB 2007–2014.

Abbreviations: wISS: weighted injury severity score; ISS: injury severity score; NTDB: National Trauma Data Bank.

Figure 3.

Comparison of goodness-of-fit between wISS and ISS, NTDB 2007–2014. Abbreviations: wISS: weighted injury severity score; ISS: injury severity score; NTDB: National Trauma Data Bank. Observed mortality and predicted by ISS and wISS are all grouped by individual ISS values. Hosmer-Lemeshow statistic: 11.6 vs. 19.7 in ISS 1–74 and 10.9 vs. 12.6 in ISS 25–74, for wISS and ISS respectively.

AUC by ISS and wISS

Table 1

		# patients	# death	% death	ISS		wISS		p value
					AUC	95% CI	AUC	95% CI	
NTDB 2007–2014									
	ISS: 1–74	545,015	5,335	0.98	0.86	(0.86–0.87)	0.88	(0.87–0.88)	<.0001
	ISS: 25–74	14,056	2,608	18.55	0.64	(0.63–0.65)	0.77	(0.76–0.78)	<.0001
NEDS 2009–2014									
	ISS: 1–74	1,302,348	843	0.06	0.91	(0.90–0.92)	0.93	(0.93–0.94)	<.0001
	ISS: 25–74	1,498	261	17.42	0.53	(0.49–0.57)	0.80	(0.77–0.82)	<.0001

Table 2

Comparisons of diagnostic statistics between ISS and wISS, NTDB 2007–2014

		ISS 1–74		ISS 25–74	
		ISS	wISS	ISS	wISS
Scenario A	Sensitivity, %	48.9	49.8	51.9	54.9
	Specificity, %	97.9	97.9	71.7	82.5
	Positive predictive value, %	18.6	18.8	29.5	41.7
	Negative predictive value, %	99.5	99.5	86.7	88.9
	Concordance, %	97.4	97.4	68.0	77.4
Scenario B	Sensitivity, %	71.1	70.1	67.1	68.8
	Specificity, %	89.0	90.7	47.4	70.3
	Positive predictive value, %	6.0	6.9	22.5	34.6
	Negative predictive value, %	99.7	99.7	86.3	90.8
	Concordance, %	88.8	90.5	51.0	70.0
Scenario C	Sensitivity, %	90.3	90.2	86.3	90.1
	Specificity, %	50.7	59.1	22.8	41.6
	Positive predictive value, %	1.8	2.1	20.3	26.0
	Negative predictive value, %	99.8	99.8	88.0	94.9
	Concordance, %	51.1	59.4	34.6	50.6

ISS: injury severity score; wISS: weighted injury severity score; NTDB: National Trauma Data Bank.

All statistics were calculated by setting a cutoff value with which the sensitivities are approaching 50% (Scenario A), 70% (Scenario B), and 90% (Scenario C).