Comparisons of the Outcome Prediction Performance of Injury Severity Scoring Tools Using the Abbreviated Injury Scale 90 Update 98 (AIS 98) and 2005 Update 2008 (AIS 2008)

Hideo Tohira, MPH, MEng, MD, Ian Jacobs, PhD, David Mountain, MBBS, FACEM,
Nick Gibson, PhD
School of Primary, Aboriginal, and Rural Health Care
The University of Western Australia, Perth, Western Australia, Australia

Allen Yeo, MBBS, FRCS
Sir Charles Gairdner Hospital, Perth, Western Australia, Australia

---

ABSTRACT –The Abbreviated Injury Scale (AIS) was revised in 2005 and updated in 2008 (AIS 2008). We aimed to compare the outcome prediction performance of AIS-based injury severity scoring tools by using AIS 2008 and AIS 98. We used all major trauma patients hospitalized to the Royal Perth Hospital between 1994 and 2008. We selected five AIS-based injury severity scoring tools, including Injury Severity Score (ISS), New Injury Severity Score (NISS), modified Anatomic Profile (mAP), Trauma and Injury Severity Score (TRISS) and A Severity Characterization of Trauma (ASCOT). We selected survival after injury as a target outcome. We used the area under the Receiver Operating Characteristic curve (AUROC) as a performance measure. First, we compared the five tools using all cases whose records included all variables for the TRISS (complete dataset) using a 10-fold cross-validation. Second, we compared the ISS and NISS for AIS 98 and AIS 2008 using all subjects (whole dataset). We identified 1,269 and 4,174 cases for a complete dataset and a whole dataset, respectively. With the 10-fold cross-validation, there were no clear differences in the AUROCs between the AIS 98- and AIS 2008-based scores. With the second comparison, the AIS 98-based ISS performed significantly worse than the AIS 2008-based ISS (p<0.0001), while there was no significant difference between the AIS 98- and AIS 2008-based NISSs. Researchers should be aware of these findings when they select an injury severity scoring tool for their studies.

INTRODUCTION

Many injury severity scoring tools have emerged since the 1970s. Most were designed to predict injured patients' vital outcome. These tools utilize the Abbreviated Injury Scale (AIS) to summarize the overall severity of multiple trauma patients. The AIS was first published in 1971 [AMA, 1971]. After some revisions and updates, the latest AIS 2005 update 2008 (AIS 2008) was released [AAAM, 2008]. Because different AIS versions are not always compatible, injury severity scoring tools using the new AIS should be compared with those using previous versions in terms of score and predictive performance.

Several injury severity scoring tools that utilize the AIS have been developed. The Injury Severity Score (ISS)

CORRESPONDING AUTHOR:
Hideo Tohira, MPH, MEng, MD, School of Primary, Aboriginal, and Rural Health Care, M516 The University of Western Australia, 35 Stirling Highway, Crawley, WA, Australia; soba@tororo.net

was the first AIS-based injury severity measure. The ISS was developed to measure the correlation between sustained injuries and mortality [Baker, O'Neill, Haddon et al., 1974]. The ISS is the sum of squares of the single highest severity level in the three most severely injured body regions. This ISS algorithm ignores the second-most severe injury in one body region, which may be more severe than the most severe injury in other regions. To solve this issue, the New Injury Severity Score (NISS) was introduced [Osler, Baker, and Long, 1997]. The NISS is the sum of squares of the three highest severity levels, regardless of body region. Osler et al. reported that the NISS was better at predicting mortality than the ISS. Unlike the ISS and the NISS, the modified Anatomic Profile (mAP) is a four-number severity characterization [Sacco, MacKenzie, Champion et al., 1999]. The mAP consists of mA, mB, mC and maxAIS. The mA, mB and mC are the square root of the sum of squares of severity levels from 3 to 6 of all injuries in a particular body region. The maxAIS is the highest severity level of any body regions. The mAP can take into account the number, severity, location and the maximum severity level of all injuries. Sacco et al. reported that the mAP displayed better goodness-of-fit to data than

the ISS and the NISS. The most important characteristic of these three tools is that a given patient's scores will not change throughout his or her clinical course. For instance, the ISS of a femoral fracture is always 9, and it will not change, even if the patient dies.

The ISS, the NISS and the mAP are generally used to select patients with a given severity level, characterize study populations and adjust the risk of anatomical injury severity in a multivariate model. For instance, they appear in inclusion criteria for study populations or in definitions of major trauma. One of the definitions of major trauma in the Western Australian trauma registry is an ISS > 15 [Towler, 2007].

Because the predictive performance of the ISS, the NISS and the mAP is limited, the Trauma and Injury Severity Score (TRISS) [Boyd, Tolson, and Copes, 1987] and A Severity Characterization of Trauma (ASCOT) [Champion, Copes, Sacco et al., 1990a] were developed. Both the TRISS and ASCOT are based on a logistic regression model. The TRISS incorporates age, type of injury (blunt/penetrating), ISS and Revised Trauma Score (RTS) [Champion, Sacco, Copes et al., 1989]. ASCOT includes age, mA, mB and mC of the mAP and three components of the RTS as independent variables. These two tools produce a probability of survival and were reported to be better at predicting the outcomes of injured patients than the ISS, the NISS and the mAP [Dillon, Wang, and Bouamra, 2006]. The TRISS and ASCOT were used to evaluate the quality of trauma care by identifying cases that require an audit analysis to determine the appropriateness of provided care [Gabbe, Cameron, Wolfe et al., 2004; Hollis, Yates, Woodford et al., 1995; Rabbani, Moini, Rabbani et al., 2007].

The NISS and ASCOT have been compared with the ISS and the TRISS for their predictive performance using the AIS 98, respectively. Because the NISS was developed to address the limitation of the ISS, the NISS was expected to perform better than the ISS [Osler et al., 1997]. Similarly, because ASCOT incorporated the mAP rather than the ISS and classified age more precisely than the TRISS, ASCOT was expected to perform better than the TRISS [Champion et al., 1990a]. However, comparisons of the NISS to the ISS and ASCOT to the TRISS have not been performed since the AIS 2008 was launched.

The coefficients of the TRISS and ASCOT should have been updated when a new AIS version was launched. The coefficients of the TRISS and ASCOT were originally derived from the AIS 85 and the population of the Major Trauma Outcome Study (MTOS), which

was conducted in the 1980s [Champion, Copes, Sacco et al., 1990b]. Following the introduction of the AIS 90, the coefficients for the TRISS were updated to provide a better estimate of the probability of survival [Champion, Sacco, and Copes, 1995]. In the Western Australian trauma registries, these updated coefficients were used to compute the probability of survival for each case, although the AIS 2005 has been used since 2007. This inconsistency between the AIS on which the coefficients were based and the AIS that was used in the registries can produce an inaccurate estimate of the probability of survival. The updated coefficients should be available using the AIS 2005 (AIS 2008).

Since the AIS 2005 was introduced, differences between calculated severity scores from the AIS 98 and AIS 2005 have been reported [Barnes, Hassan, Cuerden et al., 2009; Palmer, Niggemeyer, and Charman, 2010; Salottolo, Settell, Uribe et al., 2009]. However, the predictive performance of injury severity scoring tools that were based on the new AIS 2008 (AIS 2005-modified) has not yet been investigated. Under these circumstances, we set up the following study questions: Does the performance of injury severity scoring tools differ when a different AIS version is used? Do the NISS and ASCOT perform better than the ISS and the TRISS, respectively, when the AIS 2008 is used? Do the refitted TRISS and ASCOT using the AIS 2008 exhibit better predictive performance than the original TRISS and ASCOT? To address these questions, we aimed to compare the survival prediction performance of injury severity scoring tools that were derived from the AIS 98 with the performance of tools derived from the AIS 2008; to compare the predictive performance of the NISS with that of the ISS using the AIS 98 and the AIS 2008; and to compare the TRISS and ASCOT that were refitted according to the AIS 98 and the AIS 2008 with the original TRISS and ASCOT.

METHODS

Data

We used trauma registry data from the Royal Perth Hospital (RPH) in Western Australia. This data included all adult major trauma patients admitted to the RPH between 1994 and 2008. The RPH is one of four teaching hospitals, and it receives approximately 60% of Western Australian major trauma patients [Towler, 2007]. The RPH trauma registry conforms to the highest data integrity and quality assurance checks. Major trauma patients were defined as those who had one of the following: (1) a fatal or potentially fatal outcome; (2) an ISS of more than 15; (3) acutely disordered cardiovascular, respiratory or neurological function; (4) urgent surgery for intracranial, intrathoracic or intra-abdominal injury or for fixation of

major pelvic or spinal fractures; (5) serious injuries to two or more body regions; or (6) a need for admission to an intensive care unit, including the need for mechanical ventilation [Towler, 2007].

The RPH trauma registry contains three different versions of the AIS for injury description. The AIS 90 was used until 2005, the AIS 98 was used in 2006 only, and the AIS 2005 has been used since 2007. The AIS 98 was a minor upgrade from the AIS 90, and injury scores from these versions has been shown to be comparable [Skaga, Eken, Hestnes et al., 2007]. The AIS 2005 is a major upgrade from the AIS 98. Because more than one hundred injuries have different severity levels between the AIS 98 and AIS 2005, it is inappropriate to simultaneously use injury scores based on two AIS versions. The AIS 2008 is minor update of the AIS 2005 that corrects errors, modifies coding instruction, adds the Functional Capacity Index and changes the severity level for eight codes. These changes were reported to not affect overall injury score based on the AIS 2005 and AIS 2008 [Palmer et al., 2010]. In our data, six cases had one of the modified severity codes. However, the ISS and the NISS values that were computed using the AIS 2008 were the same as those that were computed with the AIS 2005 because the severity-modified codes were not used to compute the scores. Thus, we considered the AIS 2005 and AIS 2008 the same code system in terms of overall severity and use the terms AIS 2005 and AIS 2008 interchangeably.

To resolve the multiplicity of AIS versions in our data, we integrated all AIS codes into two AIS versions. We converted all AIS 90 and AIS 98 codes into AIS 2008 codes (combined with those already coded in AIS 2005 [2008]) to create the AIS 2008 dataset with a mapping table. The same process was applied to AIS 2005 codes to create an AIS 98 dataset. An AIS 2008 dictionary includes a mapping table for these code conversions [AAAM, 2008]. The table basically consists of two columns: a column for AIS 98 codes and a column for AIS 2008 codes. Each row includes a pair of equivalent AIS 98 and AIS 2008 codes. We converted the AIS 98 codes to equivalent AIS 2008 codes by simply matching an AIS 98 code in the dataset with an AIS 98 code in the mapping table to identify an equivalent AIS 2008 code using Predictive Analytics Software (PASW) statistics (Chicago, Illinois). The same process was applied to convert of AIS 2008 codes to AIS 98 codes.

The original mapping table in an AIS 2008 dictionary can theoretically convert all AIS 2008 codes to appropriate AIS 98 codes. However, not all AIS 98 codes can be mapped to AIS 2008 codes with this table

[Cameron and Palmer, 2011]. Of 1,341 AIS 98 codes, 153 codes (11%) cannot be mapped to AIS 2008 codes. To resolve this low efficiency of the original mapping table, three qualified Injury Scaling course instructors independently assigned appropriate AIS 2008 codes to unmappable AIS 98 codes. When the instructors selected different AIS 2008 codes for a given unmappable AIS 98 code, the three instructors discussed the allocation of the code until a consensus was achieved. As a result, we created additional 115 new links from the AIS 98 to the AIS 2008. In this study, we used this modified mapping table [Tohira, Jacobs, Mountain et al., 2010].

Injury severity scoring tools

We selected five injury severity scoring tools for comparison: the Injury Severity Score (ISS) [Baker et al., 1974], the New Injury Severity Score (NISS) [Osler et al., 1997], the modified Anatomic Profile (mAP) [Sacco et al., 1999], A Severity Characterization Of Trauma (ASCOT) [Champion et al., 1990a] and the Trauma and Injury Severity Score (TRISS) [Boyd et al., 1987]. We chose these five tools because they appear frequently in injury research articles. Although the TRISS and ASCOT have two separate models for blunt and penetrating injuries, we considered only the model for blunt injuries because penetrating injuries were rare in our dataset. We computed each severity score using the AIS 98 and AIS 2008 dataset. In the following section, the number after the name of the injury severity scoring tool denotes the AIS version used. For instance, an ISS based on the AIS 98 is referred as the ISS 98. Hereafter, we refer to the ISS, the NISS and the mAP as stand-alone tools, the TRISS and ASCOT as combined tools and all five tools as target tools.

Study populations

We created two study populations: a whole dataset and a complete dataset. The whole dataset included all adult major trauma patients registered in the RPH trauma registry between 1994 and 2008. The whole dataset was used to compare the performance of the ISS and NISS because these tools do not require physiological parameters and are independent of transfer status and injury type. All injury types (i.e., blunt, penetrating and burn) were included in this dataset. Subjects without an ISS, a NISS and/or outcome data were excluded. We also excluded cases that had an AIS code that was unconvertible with the mapping table.

The complete dataset was used to derive five target tools and compare their predictive performance; thus, the dataset contained only cases that had no missing data for variables that were required to derive the five target tools. The complete dataset was created from the

whole dataset by excluding indirect admissions and cases missing data for AIS codes, physiological parameters (systolic blood pressure, respiratory rate, Glasgow Coma Scale) on arrival and/or age. We excluded indirect admissions because physiological parameters on arrival at the first hospital were not recorded in the RPH trauma registry. We considered that vital signs of a transferred patient measured at the RPH did not reflect their immediate physiological response to injury because of interventions applied and time elapsed [Ridley and Carter, 1989]. These parameters were essential to calculate the TRISS and ASCOT. We further excluded penetrating injuries, burns and isolated hip fractures. We excluded penetrating injuries because they accounted for only 3% of our data, resulting in a low precision of the results. Patients with isolated hip fractures were also excluded because we considered isolated hip fractures as a different cohort from other injuries. Isolated hip fractures were known to have higher mortality rates than those without isolated hip fractures but the same injury severity level. [Bergeron, Lavoie, Belcaid et al., 2005]. Hip fracture was identified using ICD 9-CM (920), ICD 10 (S72.0, S72.1, S72.2), AIS 98 (851808.3 – 851812.3) or AIS 2008 (853111.3-853172.3). We defined any case aged 65 or older with one of the hip fracture codes above and whose NISS was equal to or less than 10 as an isolated hip fracture [Clark, DeLorenzo, Lucas et al., 2004]. A criterion of NISS $\leq$ 10 allows one minor injury other than a hip fracture.

We included dead on hospital arrivals (DOAs) in our analysis. Inclusion of DOAs can affect overall mortality. However, it is uncertain whether DOAs affect the predictive performance of injury severity scoring tools. Gomez et al. investigated the effect of variable case ascertainment of DOA on the assessment of trauma care quality using a logistic regression model [Gomez, Xiong, Haas et al., 2009]. They reported that the variable case ascertainment of DOA did not affect the model's ability to assess the quality of trauma care. This result may indirectly imply that the inclusion of DOAs does not affect the predictive performance of injury severity scoring tools, but uncertainty remains. To be certain, we reported the proportion of DOAs and clearly defined DOA. Our definition of DOA was any patient who had a minimal value for all available physiological parameters (systolic blood pressure, respiratory rate and GCS on hospital arrival). When any parameters were missing, only available parameters were considered to identify DOA. Cases that died at the scene and were not transported to a hospital were not included in our study population because such cases were not included in the trauma registry.

Performance measures

The predictive performance of injury severity scoring tools has been measured by discrimination or discriminative power [Lavoie, Moore, LeSage et al., 2004; Osler et al., 1997]. Discrimination is a tool's ability to predict an event versus a nonevent. We used the area under Receiver Operating Characteristic curve (AUROC) to measure discrimination. The AUROC is equivalent to the probability that a randomly selected subject who experienced the event has a higher predicted risk than a randomly selected person who did not experience the event [Hanley and McNeil, 1982]. Thus, a tool with a large AUROC can accurately select patients with a certain level of injury severity and, in turn, reduce selection bias for missing a target cohort. The highest AUROC is 1.0, meaning that a tool can discriminate events and nonevents completely. The lowest AUROC is 0.5, meaning that a tool predicts events and nonevents by chance.

Comparison

We first compared five target tools by a 10-fold cross-validation using the complete dataset; then we compared the ISS and NISS using the whole dataset.

Ten-fold cross-validation

We selected 10-fold cross-validation to compare target tools. If the same data were used to derive and test a model, the performance measures could be overestimated due to overfitting [Kohavi, 1995]. We could avoid this known issue with different data for deriving and testing. A hold-out method can avoid overfitting by dividing samples into two mutually exclusive subsets: a training and a test set. However, estimates derived from the hold-out method are known to be pessimistic because only a portion of the samples is used in a training set [Kohavi, 1995]. In the 10-fold cross-validation, a test set is different from a training set, and all data can be used in both training and testing. Furthermore, a 10-fold cross-validation estimate is known to be unbiased [Kohavi, 1995].

We performed 10-fold cross-validation as follows. We randomly divided the complete dataset into 10 subsets. Nine subsets were used to derive a model and one subset was used to test the model (to compute an AUROC). This process was repeated 10 times by changing the deriving and testing subsets. We averaged the 10 AUROCs that were computed at each round of cross-validation to produce an unbiased estimate. Statistical comparisons were not made due to the unavailability of an unbiased estimator of variances [Bengio and Grandvalet, 2004].

In each round of 10-fold cross-validation, we derived a univariate logistic regression model for ISS and NISS and a multivariate logistic regression model for mAP, ASCOT and TRISS to calculate a set of coefficients. After regressing the data, we applied the derived models to the remaining subset and computed the AUROC. With the TRISS, we did not use the Major Trauma Outcome Study (MTOS) coefficients for the RTS [Champion et al., 1989]. Instead, we first refitted the RTS using a deriving dataset and then refitted the TRISS with the refitted RTS. We also refitted the TRISS using not only the ISS but also the NISS because statistical models that incorporated the NISS were reported to perform better than those that incorporated the ISS [Aydin, Bulut, Ozguc et al., 2008; Balogh, Offner, Moore et al., 2000; Frankema, Steyerberg, Edwards et al., 2005].

In this analysis, we compared scores that were generated from the AIS 98 to those generated from the AIS 2008. We also compared ASCOT to the TRISS for each AIS version.

After the 10-fold cross-validation, we refitted the RTS, the TRISS and ASCOT using the entire complete dataset to compute new sets of coefficients, which can be used for future studies. We also refitted the TRISS by incorporating the NISS as I performed in the 10-fold cross-validation.

ISS vs. NISS

We computed the AUROCs of the ISS 98, NISS 98, ISS 2008 and NISS 2008 using the whole dataset. We compared the ISS and the NISS that were derived using the AIS 98 with those derived with the AIS 2008. We also compared the ISS to the NISS for each AIS version. We performed these comparisons separately from the 10-fold cross-validation because we could gain more power to detect differences using the larger dataset than the complete dataset and could perform statistical comparisons.

We used the PASW Ver. 18 (Chicago, Illinois, US) for code conversions and statistical analysis. We calculated AUROCs and standard errors using the ROC command of the PASW. We used a nonparametric method to calculate the standard errors of AUROCs. We compared two AUROCs, as Hanley et al. described [Hanley and McNeil, 1983].

RESULTS

Demographics

We identified 4,174 major trauma patients for the whole dataset (Table 1). The mean age was 41.09, ranging from 13 to 102. Male was dominant (75.2%),

and in-hospital mortality was 14.1%. Most of the patients sustained blunt trauma (89.8%). Direct admission from an event scene accounted for 38.4% of cases. Fifty-eight DOAs were identified. The mean length of hospital stay was 19.38 days. The mean ISS 98 and ISS 2008 were 26.51 and 23.02, respectively. We did not find any isolated hip fractures.

Table 1 - Patient Demographics.

		Whole dataset Mean (SD) or n (%)	Complete dataset Mean (SD) or n (%)
N		4,174	1,269
Age		41.09 (20.83)	44.18 (23.06)
Sex	Female	1,034 (24.8%)	314 (24.7%)
	Male	3,140 (75.2%)	955 (75.3%)
Mortality		587 (14.1%)	209 (16.5%)
Type	Blunt	3,750 (89.8%)	1,269
	Penetrating	135 (3.2%)	0
	Other	289 (6.9%)	0
Admission	Direct	1,603 (38.4%)	1,269
	Indirect	2,571 (61.6%)	0
DOA		58 (1.4%)	39 (3.1%)
LOS		19.38 (27.43)	16.49 (28.49)
AIS 98	ISS	26.51 (10.77)	26.32 (10.82)
	NISS	36.74 (14.99)	37.56 (15.12)
	mA	4.01 (3.46)	4.43 (3.35)
	mB	1.68 (2.36)	1.63 (2.38)
	mC	2.00 (2.34)	1.62 (2.25)
	maxAIS	4.18 (0.72)	4.15 (0.71)
AIS 2008	ISS	23.02 (10.58)	22.89 (10.72)
	NISS	31.45 (15.27)	32.26 (15.70)
	mA	3.54 (3.16)	3.90 (3.11)
	mB	1.14 (1.93)	1.09 (1.92)
	mC	1.57 (2.11)	1.31 (2.03)
	maxAIS	3.93 (0.88)	3.90 (0.89)

Whole dataset: all adult major trauma patients registered in the RPH trauma registry between 1994 and 2008

Complete dataset: cases that had no missing data for variables that were required to derive the five target tools.

SD: standard deviation

DOA: dead on arrival; LOS: length of stay

We found 1,269 major blunt trauma patients in the complete dataset (Table 1). The mean age was 44.18, ranging from 13 to 102. Male was dominant (75.3%), and in-hospital mortality was 16.5%. There were 39 DOAs. The mean ISS 98 and ISS 2008 were 26.32 and 22.89, respectively. Our datasets included older and more severely injured patients and exhibited higher mortality than the MTOS dataset (mean age: 33.1;

mean ISS: 12.8; mortality: 9.0%) [Champion et al., 1990b].

Ten-fold cross-validation

The results of the 10-fold cross-validation are shown in Tables 2 and 3. Each value is the average of 10 AUROCs that were produced at each round of 10-fold cross-validation.

Table 2 - Results of 10-fold cross-validation of combined tools.

	Averaged AUROC	
	AIS 98	AIS 2008
TRISS (ISS)	0.910	0.911
TRISS (NISS)	0.914	0.911
ASCOT	0.915	0.917
TRISS (MTOS)	0.901	*
ASCOT (MTOS)	0.911	*

AUROC: area under the Receiver Operating Characteristic curve

MTOS: Major Trauma Outcome Study

*: The coefficients based on the AIS 2008 are not available for both the TRISS and ASCOT.

Table 3 - Results of 10-fold cross-validation of stand-alone tools.

	Averaged AUROC	
	AIS 98	AIS 2008
mAP	0.804	0.794
ISS	0.739	0.762
NISS	0.796	0.782

AUROC: area under the Receiver Operating Characteristic curve

Combined tools

The highest AUROC was 0.917 for ASCOT 2008, followed by ASCOT 98 (AUROC=0.915) and the TRISS with NISS 98 (AUROC=0.914). The TRISS with ISS 98 had the lowest AUROC (0.910). No AIS version was clearly superior for the combined tools. The TRISS with ISS and ASCOT had better discrimination for the AIS 98 than the AIS 2008, but the TRISS with NISS had better discrimination for the AIS 2008 than the AIS 98 (Table 2).

ASCOT showed larger AUROCs than the TRISS for both the AIS 98 and the AIS 2008 (Table 2). The TRISS and ASCOT that were refitted with the AIS 98 and the AIS 2008 outperformed the original TRISS and ASCOT, respectively (Table 2).

Stand-alone tools

The mAP 98 demonstrated the highest AUROC (0.804), followed by the NISS 98 (0.796) and mAP 2008 (0.794) (Table 3). The lowest AUROC was 0.739 with the ISS 98 (Table 3). The superior AIS version for stand-alone tools was also inconclusive. The AIS 98 was the better version for the NISS and the mAP, whereas the AIS 2008 was the better version for the ISS (Table 3). The NISS exhibited larger AUROCs for both the AIS 98 and the AIS 2008 than the ISS. The difference in AUROC between the NISS and the ISS was larger for the AIS 98 than the AIS 2008 (Table 3).

Comparison of ISS and NISS

NISS 2008 performed best among the four tools (AUROC=0.7884), whereas ISS 98 performed worst (AUROC=0.735) (Table 4). The ISS 98 performed significantly worse than the ISS 2008 ($p < 0.0001$), while there was no significant difference between the NISS 98 and the NISS 2008. With the NISS versus the ISS, the NISS demonstrated a significantly higher AUROC than the ISS when the AIS 98 was used ($p < 0.0001$), whereas there was no significant difference between the ISS and the NISS when the AIS 2008 was used.

Table 4 - AUROCs and 95% CIs of ISS and NISS

	AUROC	SE	95% CI	
ISS 98	0.735	0.011	0.713-0.757	$p < 0.0001$
ISS 2008	0.769	0.011	0.748-0.789	
NISS 98	0.788	0.012	0.765-0.811	
NISS 2008	0.788	0.011	0.767-0.810	

AUROC: area under the Receiver Operating Characteristic curve.

SE: standard error; CI: confidence interval

One might consider that the AUROCs of the ISS 98 and the ISS 2008 were not significantly different because their 95% CIs overlap. However, overlapping 95% CIs do not always mean that there is no significant difference; the AUROCs of the ISS 98 and the ISS 2008 were derived using the same samples, and one must take a correlation between two AUROCs into account when performing statistical comparisons [Hanley et al., 1983]. In this case, the standard error of difference between two related AUROCs will be smaller than that between two unrelated AUROCs.

Updated coefficients for RTS, TRISS and ASCOT

New coefficients for the RTS, the TRISS and ASCOT are shown in Figure 1. With TRISS, there are four sets of coefficients. Each set was obtained from a model with different injury severity scores (the ISS 98, ISS 2008, NISS 98 and NISS 2008). The TRISS models used new coefficients for the RTS. With ASCOT, there

are two sets of coefficients using the AIS 98 or AIS 2008.

Figure 1 - New coefficients for the Revised Trauma Score (RTS), Trauma and Injury Severity Score (TRISS) and A Severity Characterization Of Trauma (ASCOT).

(a) RTS

$RTS = 0.7115 \times cGCS + 0.6872 \times cSBP + 0.1661 \times cRR$

A subscript c indicates coded value of each parameter.

Coding table for the RTS

Coded value	GCS	SBP	RR
4	13-15	>89	10-29
3	9-12	76-89	>29
2	6-8	50-75	6-9
1	4,5	1-49	1-5
0	3	0	0

(b) TRISS

$Ps = 1/(1 + \exp(-\beta))$

$\beta = \beta_0 + \beta_1 \times cAge + \beta_2 \times RTS + \beta_3 \times ISS$ (or NISS)

$cAge = 1$ if age ≥ 55; otherwise, $cAge = 0$.

Tool	β_0	β_1	β_2	β_3	AUROC
ISS 98	-1.2947	-2.6066	1.1104	-0.0542	0.9076
ISS 2008	-1.2529	-2.5838	1.1121	-0.0637	0.9111
NISS 98	-0.1723	-2.4148	1.0045	-0.0515	0.9119
NISS 2008	-0.8939	-2.3848	1.0415	-0.0454	0.9106

The numbers after ISS or NISS indicate the version of the Abbreviated Injury Scale.

(c) ASCOT

$Ps = 1/(1 + \exp(-\beta))$

$\beta = \beta_0 + \beta_1 \times cGCS + \beta_2 \times cSBP + \beta_3 \times cRR$
$+ \beta_4 \times mA + \beta_5 \times mB + \beta_6 \times mC + \beta_7 \times cAge$

cGCS, cSBP and cRR: components of the RTS

mA, mB and mC: components of the modified Anatomic Profile

cAge: 0 for 0-54 years old; 1 for 55-64 years old; 2 for 65-74 years old; 3 for 75-84 years old; and 4 for 85 years old or older.

AIS version	β_0	β_1	β_2	β_3
AIS 98	-0.9412	0.7208	0.5856	0.3241
AIS 2008	-0.9610	0.7604	0.5819	0.2643

β_4	β_5	β_6	β_7	AUROC
-0.2209	-0.1614	-0.1235	-0.8846	0.9159
-0.2296	-0.2238	-0.1430	-0.8734	0.9163

DISCUSSION

This study is the first to compare the discriminative powers of five injury severity scoring tools based on the AIS 98 and AIS 2008. We found that the predictive performance of the ISS 2008 was superior to that of the ISS 98, but the other tools exhibited similar performances whether they were based on the AIS 98 or the AIS 2008. The ISS also performed significantly

worse than the NISS when the AIS 98 was used, whereas there was no significant difference between the ISS and the NISS when the AIS 2008 was used. ASCOT exhibited larger AUROCs than the TRISS for both the AIS 98 and AIS 2008; however, the differences seemed to be too small to be clinically significant. Because the ISS is widely used for selecting target populations with a given severity level, the use of the ISS 98 may impose a larger selection bias in a research than the ISS 2008. On the other hand, either AIS version might be used to derive stand-alone tools, if the tools are to be included in the combined tools for risk adjustment.

Stand-alone tools

The AIS 2008 was superior to the AIS 98 in terms of discriminative power when using the ISS, although the superiority of either the AIS 2008 or AIS 98 was inconclusive for the NISS and mAP. Based on these findings, researchers should not use the ISS when only AIS 98 codes are available in their study data. Instead, the NISS might be an alternative. When AIS 98 and AIS 2008 codes coexist in study data, it might be better to convert AIS 98 codes into AIS 2008 codes so any injury severity scoring tool can be used. When only AIS 2008 codes are available, any tool can be used.

Although the mAP outperformed the ISS and the NISS, the use of mAP is limited because it is not a single number measure but a four-number measure. The mAP might be used in a multivariate model as ASCOT or might be used as a reference measure with high performance when the predictive performance of injury severity scoring tools is studied.

Combined tools

We could not determine a superior AIS version for combined tools because there were small differences in AUROCs between the AIS 98 and the AIS 2008. Although we found that the ISS 98's performance was significantly inferior to that of the ISS 2008 by itself, the ISS 98 might be used for risk adjustment in conjunction with other covariates in a multivariate regression model as the TRISS.

We confirmed that the refitted TRISS and ASCOT outperformed those using the MTOS coefficients. This result does not conflict with other reports. The important point of our methodology is that we also refitted the RTS when deriving the TRISS. Most studies refitted the TRISS with the MTOS RTS. We found only one paper that compared the TRISS using the refitted RTS with one using the MTOS RTS [Garber, Hebert, Wells et al., 1996]. However, this research compared the goodness-of-fit, sensitivity and

specificity, not AUROCs. Thus, we could compare the discriminative powers of TRISSs more appropriately than previous reports.

This study is the first to compare the AUROCs of refitted TRISSs and ASCOTs. We could not identify a study that computed a refitted ASCOT's AUROC by an electric database search. There were six articles that compared the TRISS and ASCOT for discrimination [Champion, Copes, Sacco et al., 1996; Dillon et al., 2006; Frankema, Edwards, Steyerberg et al., 2002; Frankema et al., 2005; Gabbe, Cameron, Wolfe et al., 2005; Rabbani et al., 2007]. All these articles used the MTOS coefficients for ASCOT based on the AIS 85. Some studies used the AIS 85-based coefficients for ASCOT, even though they used the AIS 90 for their study population [Frankema et al., 2002; Frankema et al., 2005; Rabbani et al., 2007]. Hannan et al. derived ASCOT using their local data; however, they computed goodness-of-fit rather than AUROC [Hannan, Mendeloff, Farrell et al., 1995]. Thus, we have computed ASCOT's discriminative power more precisely than previous reports.

ASCOT's superiority to the TRISS is still inconclusive. In our results, ASCOT outperformed the TRISS for both the AIS 98 and AIS 2008. However, the difference might be clinically small. Further, other reports did not always achieve the same result as ours. Of the six reports cited above, two reported that ASCOT outperformed the TRISS [Dillon et al., 2006; Frankema et al., 2005], three reported that the TRISS outperformed ASCOT [Dillon et al., 2006; Frankema et al., 2002; Gabbe et al., 2005], and one reported that the TRISS and ASCOT performed equally [Rabbani et al., 2007]. As mentioned above, these comparisons were not accurate because ASCOTs and RTSs were not refitted. To determine the superiority of ASCOT or the TRISS, further research will be required.

New coefficients for the RTS, the TRISS and ASCOT

We computed new coefficients for the RTS, the TRISS and ASCOT using our local data. Several authors reported new coefficients for the TRISS using their local data [Dillon et al., 2006; DiRusso, Sullivan, Holly et al., 2000; Hunter, Kennedy, Henry et al., 2000; Kilgo, Meredith, Osler et al., 2006; Kroezen, Bijlsma, Liem et al., 2007; Millham, LaMorte, Millham et al., 2004; Moore, Lavoie, Turgeon et al., 2009; Schluter, Nathens, Neal et al., 2010]. However, most authors, except Schluter et al., derived TRISS coefficients, but not those for the RTS. They calculated the RTS using the MTOS coefficient derived in the 1980s and refitted the TRISS. We believe there is a better chance of improving the TRISS's performance if the RTS is also refitted.

We also derived the TRISS and ASCOT by changing the AIS version and injury severity scores (ISS, NISS). This variability of models will provide researchers with more alternatives for selecting appropriate injury severity scoring tools.

Limitations

The number of cases that could be used to derive TRISS or ASCOT was limited. The RPH is one of four teaching hospitals/trauma centers in the Perth Metropolitan area. The RPH has the major role in accepting patients who have sustained severe injuries in rural and regional areas. The RPH also accepts referred patients who were initially admitted to other hospitals when high level of care is required. As a result, direct admission accounted for only 38.4% of patients from whom we could refit the TRISS and ASCOT. Indirect admission cases could not be used to derive these models because the initial physiological parameters measured at the first hospital were not recorded in most cases. Thus, the precision of our result was not high enough to show small but potentially important differences. The use of a large database (i.e., a state or national trauma registry) may resolve this issue.

Because the RPH trauma registry used three different AIS versions, we converted all AIS 90 and 98 codes into AIS 2008 codes and all AIS 2005 codes into AIS 98 codes using a mapping table. This conversion may have resulted in inaccurate calculation of injury scores. Palmer et al. compared ISSs derived from manually selected AIS 2005 codes and mapped AIS 2005 codes using the original mapping table [Palmer et al., 2010]. They reported that ISS agreement between manual AIS 2005 and mapped AIS 2005 was only 57%, which suggests that mapped ISS was fairly inaccurate in 43% of subjects. However, they reported that the mean difference of ISSs between the manual and mapped AIS 2005 was 2.5. This difference seems to be clinically acceptable. Moreover, our modification on the original table reduced the number of exclusions. If we had used the original mapping table, we have had to exclude 1,912 cases because of unmappable AIS 98 codes, including 999 rib fractures with pneumothorax and 463 pelvic fractures. On the other hand, our modified mapping table allowed us to convert all rib fracture and pelvic fracture codes to AIS 2008 and reduce the number of exclusions to 612 cases, which contained 570 codes for loss of consciousness. Because manual recoding of all cases with the AIS 98 and AIS 2008 is impractical and unfeasible, we believe that the use of the mapping table is the best available alternative to integrating an AIS version. The validation of our modified mapping table is currently underway.

Future study

A new multivariate model that includes both direct and indirect admissions should be developed to predict the outcomes of major trauma patients in Western Australia. Both the TRISS and ASCOT require data related to the physiological parameters on admission to the first hospital. Cases that are transferred from another hospital and do not have data for these parameters must be excluded. We cannot use physiological parameters measured at the referred hospital because such parameters do not accurately reflect patients' physiological responses to injury due to interventions given at the first hospital and time elapsed after the event [Ridley et al., 1989]. To resolve this issue, a multivariate model that incorporates both the physiological parameters at the transferring hospital and the time elapsed since the event probably needs to be developed. Another solution for missing physiological parameters is to use sophisticated statistical techniques, such as multiple imputation, to impute unavailable physiological parameters. Schluter et al. reported the feasibility of multiple imputation for refitting the TRISS [Schluter et al., 2010]. This technique may be feasible when the proportion of missing data is small.

CONCLUSION

The ISS derived from the AIS 98 was significantly worse for predicting the outcome of injured patients than that derived from the AIS 2008. The NISS, the mAP, ASCOT and the TRISS performed similarly for the AIS 98 and the AIS 2008. The NISS should be used rather than the ISS when the AIS 98 is used. ASCOT constantly produced slightly larger AUROCs than TRISS; however, the difference might not be clinically important. Researchers should be aware of these findings when they select an injury severity scoring tool for their studies.

The AIS coding approach in this manuscript has not been evaluated or endorsed by the AAAM AIS Committee.

ACKNOWLEDGMENTS

REFERENCES

AAAM: Abbreviated Injury Scale 2005 update 2008. Barrington Illinois: Association for the Advancement of Automotive Medicine, 2008.

AMA: Rating the severity of tissue damage. I. The abbreviated scale. JAMA 215:277-280, 1971.

Aydin SA, Bulut M, Ozguc H, et al.: Should the New Injury Severity Score replace the Injury Severity Score in the Trauma and Injury Severity Score? Ulus Travma Acil Cerrahi Derg 14:308-312, 2008.

Baker SP, O'Neill B, Haddon W, Jr., et al.: The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. J Trauma 14:187-196, 1974.

Balogh Z, Offner PJ, Moore EE, et al.: NISS predicts postinjury multiple organ failure better than the ISS. J Trauma 48:624-627, 2000.

Barnes J, Hassan A, Cuerden R, et al.: Comparison of injury severity between AIS 2005 and AIS 1990 in a large injury database. Annual proceedings/Association for the Advancement of Automotive Medicine Association for the Advancement of Automotive Medicine 53:83-89, 2009.

Bengio Y and Grandvalet Y: No unbiased estimator of the variance of k-fold cross-validation. The Journal of Machine Learning Research 5:1089-1105, 2004.

Bergeron E, Lavoie A, Belcaid A, et al.: Should patients with isolated hip fractures be included in trauma registries? J Trauma 58:793-797, 2005.

Boyd CR, Tolson MA and Copes WS: Evaluating trauma care: the TRISS method. Trauma Score and the Injury Severity Score. J Trauma 27:370-378, 1987.

Cameron PA and Palmer C: Developing consensus on injury coding. Injury 42:10-11, 2011.

Champion HR, Sacco WJ, Copes WS, et al.: A revision of the Trauma Score. J Trauma 29:623-629, 1989.

Champion HR, Copes WS, Sacco WJ, et al.: A new characterization of injury severity. J Trauma 30:539-545, 1990a.

Champion HR, Copes WS, Sacco WJ, et al.: The Major Trauma Outcome Study: establishing national norms for trauma care. J Trauma 30:1356-1365, 1990b.

Champion HR, Sacco WJ and Copes WS: Injury Severity Scoring Again. J Trauma 38:94-95, 1995.

Champion HR, Copes WS, Sacco WJ, et al.: Improved predictions from a severity characterization of trauma (ASCOT) over Trauma and Injury Severity Score (TRISS): results of an independent evaluation. J Trauma 40:42-48, 1996.

Clark DE, DeLorenzo MA, Lucas FL, et al.: Epidemiology and short-term outcomes of injured medicare patients. J Am Geriatr Soc 52:2023-2030, 2004.

Dillon B, Wang W and Bouamra O: A comparison study of the injury score models. European Journal of Trauma 32:538-547, 2006.

DiRusso SM, Sullivan T, Holly C, et al.: An artificial neural network as a model for prediction of survival in trauma patients: validation for a regional trauma area. J Trauma 49:212-220, 2000.

Frankema SP, Edwards MJR, Steyerberg EW, et al.: Predicting survival after trauma: A comparison of TRISS and ASCOT in the Netherlands. European Journal of Trauma 28:355-364, 2002.

Frankema SP, Steyerberg EW, Edwards MJR, et al.: Comparison of current injury scales for survival chance estimation: an evaluation comparing the predictive performance of the ISS, NISS, and AP scores in a Dutch local trauma registration. J Trauma 58:596-604, 2005.

Gabbe BJ, Cameron PA, Wolfe R, et al.: TRISS: does it get better than this? Academic emergency medicine 11:181-186, 2004.

Gabbe BJ, Cameron PA, Wolfe R, et al.: Predictors of mortality, length of stay and discharge destination in blunt trauma. ANZ J Surg 75:650-656, 2005.

Garber BG, Hebert PC, Wells G, et al.: Validation of trauma and injury severity score in blunt trauma patients by using a Canadian trauma registry. J Trauma 40:733-737, 1996.

Gomez D, Xiong W, Haas B, et al.: The Missing Dead: The Problem of Case Ascertainment in the Assessment of Trauma Center Performance. J Trauma 66:1218-1225, 2009.

Hanley JA and McNeil BJ: The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143:29-36, 1982.

Hanley JA and McNeil BJ: A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 148:839-843, 1983.

Hannan EL, Mendeloff J, Farrell LS, et al.: Validation of TRISS and ASCOT using a non-MTOS trauma registry. J Trauma 38:83-88, 1995.

Hollis S, Yates DW, Woodford M, et al.: Standardized comparison of performance indicators in trauma: a new approach to case-mix variation. J Trauma 38:763-766, 1995.

Hunter A, Kennedy L, Henry J, et al.: Application of neural networks and sensitivity analysis to improved prediction of trauma survival. Computer Methods & Programs in Biomedicine 62:11-19, 2000.

Kilgo PD, Meredith JW, Osler TM, et al.: Incorporating recent advances to make the TRISS approach universally available. J Trauma 60:1002-1008, 2006.

Kohavi R: A study of cross-validation and bootstrap for accuracy estimation and model selection, in International Joint Conference on Artificial Intelligence, Montreal, Quebec, Canada1995, pp. 1137-1145.

Kroezen F, Bijlsma TS, Liem MS, et al.: Base deficit-based predictive modeling of outcome in trauma patients admitted to intensive care units in Dutch trauma centers. J Trauma 63:908-913, 2007.

Millham FH, LaMorte WW, Millham FH, et al.: Factors associated with mortality in trauma: re-evaluation of the TRISS method using the National Trauma Data Bank. J Trauma 56:1090-1096, 2004.

Moore L, Lavoie A, Turgeon AF, et al.: The trauma risk adjustment model: a new model for

evaluating trauma care. <u>Annals of Surgery</u> 249:1040-1046, 2009.

Osler T, Baker SP and Long W: A modification of the injury severity score that both improves accuracy and simplifies scoring. <u>J Trauma</u> 43:922-925, 1997.

Palmer CS, Niggemeyer LE and Charman D: Double coding and mapping using Abbreviated Injury Scale 1998 and 2005: Identifying issues for trauma data. <u>Injury</u> 41:948-954, 2010.

Rabbani A, Moini M, Rabbani A, et al.: 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". <u>Archives of Iranian Medicine</u> 10:383-386, 2007.

Ridley S and Carter R: The effects of secondary transport on critically ill patients. <u>Anaesthesia</u> 44:822-827, 1989.

Sacco WJ, MacKenzie EJ, Champion HR, et al.: Comparison of alternative methods for assessing injury severity based on anatomic descriptors. <u>J Trauma</u> 47:441-446, 1999.

Salottolo K, Settell A, Uribe P, et al.: The impact of the AIS 2005 revision on injury severity scores and clinical outcome measures. <u>Injury</u> 40:999-1003, 2009.

Schluter PJ, Nathens A, Neal ML, et al.: Trauma and Injury Severity Score (TRISS) coefficients 2009 revision. <u>J Trauma</u> 68:761-770, 2010.

Skaga NO, Eken T, Hestnes M, et al.: Scoring of anatomic injury after trauma: AIS 98 versus AIS 90-do the changes affect overall severity assessment? <u>Injury</u> 38:84-90, 2007.

Tohira H, Jacobs I, Mountain D, et al.: A modified mapping table for the Abbreviated Injury Scale 2005 Updated in 2008 (AIS 2008) made better use of the existing injury data than the original table. <u>Emergency Medicine Australasia</u> 23:36, 2010.

Towler S: <u>Trauma system and services: Report of the Trauma Working Group</u>. Perth, Department of Health Western Australia, 2007.