

# A Trauma Mortality Prediction Model Based on the Anatomic Injury Scale

Turner Osler, MD, MSc,\* Laurent Glance, MD,† Jeffery S. Buzas, PhD,‡ Dana Mukamel, PhD,§  
Jacob Wagner, MD, PhD\* , and Andrew Dick, PhD¶

**Objective:** To develop a statistically rigorous trauma mortality prediction model based on empiric estimates of severity for each injury in the abbreviated injury scale (AIS) and compare the performance of this new model with the injury severity score (ISS).

**Summary Background Data:** Mortality rates at trauma centers should only be compared after adjusting for differences in injury severity, but no reliable measure of injury severity currently exists. The ISS has served as the standard measure of anatomic injury for 30 years. However, it relies on the individual injury severities assigned by experts in the AIS, is nonmonotonic with respect to mortality, and fails to perform even as well as a far simpler model based on the single worst injury a patient has sustained.

**Methods:** This study is based on data from 702,229 injured patients in the National Trauma Data Bank (NTDB 6.1) hospitalized between 2001 and 2005. Sixty percent of the data was used to derive an empiric measure of severity of each of the 1322 injuries in the AIS lexicon by taking the weighted average of coefficients estimated using 2 separate regression models. The remaining 40% of the data was used to create 3 exploratory mortality prediction models and compare their performance with the ISS using measures of discrim-

ination (C statistic), calibration (Hosmer Lemeshow statistic and calibration curves), and the Akaike information criterion.

**Results:** Three new models based on empiric AIS injury severities were developed. All of these new models discriminated survivors from nonsurvivors better than the ISS, but one, the trauma mortality prediction model (TMPM), had both better discrimination [ $ROC_{TMPM} = 0.901$  (0.898–0.905),  $ROC_{ISS} = 0.871$  (0.866–0.877)] and better calibration [ $HL_{TMPM} = 58$  (35–91),  $HL_{ISS} = 296$  (228–357)] than the ISS. The addition of age, gender, and mechanism of injury improved all models, but the augmented TMPM dominated ISS by every measure [ $ROC_{TMPM} = 0.925$  (0.921–0.928),  $ROC_{ISS} = 0.904$  (0.901–0.909),  $HL_{TMPM} = 18$  (12–31),  $HL_{ISS} = 54$  (30–64)].

**Conclusions:** Trauma mortality models based on empirical estimates of individual injury severity better discriminate between survivors and nonsurvivors than does the current standard, ISS. One such model, the TMPM, has both superior discrimination and calibration when compared with the ISS. The TMPM should replace the ISS as the standard measure of overall injury severity.

(*Ann Surg* 2008;247: 1041–1048)

From the \*Department of Surgery, University of Vermont, Burlington, VT; †Department of Anesthesia, University of Rochester, Rochester, NY; ‡Department of Mathematics and Statistics, University of Vermont, Burlington, VT; §School of Medicine, University of California, Irvine, CA; and ¶RAND-University of Pittsburgh Health Institute, Pittsburgh, PA. Supported by a grant from the Agency for Healthcare Research and Quality (RO1 HS 016737).

The National Trauma Data Bank data set (NTDB Version 6.1, 2007) was made available by the American College of Surgeons Committee on Trauma, Chicago, IL.

The views presented in this manuscript are those of the authors and may not reflect those of Agency for Healthcare Research and Quality. The content reproduced from the NTDB data remains the full and exclusive copyrighted property of the American College of Surgeons. The American College of Surgeons is not responsible for any claims arising from works based on the original data.

The 1322 MARC values calculated for this manuscript are available on request from the authors. Instructions for calculating the TMPM from these MARC values are given in Appendix C. An Excel spreadsheet to automate the calculation of the TMPM from any data set is available as shareware from the authors and will be available at <http://www.facs.org/trauma/ntdb.html>.

Reprints: Turner Osler, MD, MSc, Department of Surgery, University of Vermont, 789 Orchard Shore Road, Colchester, VT 05446. E-mail: [tosler@silicondairy.net](mailto:tosler@silicondairy.net).

Copyright © 2008 by Lippincott Williams & Wilkins  
ISSN: 0003-4932/08/24706-1041  
DOI: 10.1097/SLA.0b013e31816ffb3f

Almost 2 million Americans are hospitalized each year due to trauma, and 45,000 (2.4%) of these patients die before discharge.<sup>1</sup> Many of the most severely injured are treated at the 1000 trauma centers across the United States, and, although many lives are saved, variation in outcomes across trauma centers certainly exists and may be substantial.<sup>2</sup> Improving outcomes, whether through private comparisons among trauma centers or public report cards, critically depends on the ability to accurately compare performance across trauma centers. There is risk in making such comparisons, however, because inaccurate or unfair results may lead to poor decisions by researchers, administrators, insurance companies, and government agencies. Indeed, the entire enterprise of trauma outcome assessment might be discredited. Thus it is crucial that accurate statistical adjustment be used if trauma centers are to be compared.

Foremost among factors important when comparing trauma centers is the dose of injury. Unfortunately, measuring this quantity is difficult because there are many hundreds of possible individual injuries, and patients often have more than a single injury. Because the number of possible injuries is large, the number of possible combinations of injuries is almost limitless.

In practice, quantifying injury requires 3 things: first a lexicon that divides the continuous landscape of human injury into a set of discrete injuries; second, a measure of severity for each injury; and third, a model that summarizes the combined severity of all the injuries that an individual patient has sustained as a single numeric value. Over 30 years ago all 3 tasks were accomplished simultaneously when Baker et al adapted the abbreviated injury scale (AIS)<sup>3</sup> listing of injuries and its attendant expert-assigned severities, and defined the injury severity score (ISS) as the sum of the squares of the severities of the single worst injury in each of the 3 most injured body regions.<sup>4</sup> This score quickly became the standard measure of trauma and was soon incorporated as the dosage of trauma in more comprehensive models of mortality following trauma [eg, trauma and injury severity score (TRISS)<sup>5</sup>]. The ISS has proven remarkably durable. Although alternative measures of injury have been proposed (eg, anatomic profile,<sup>6</sup> new injury severity score<sup>7</sup>), none has displaced the ISS.

Despite its ubiquity, the ISS has important limitations. Because it depends on the expert-assigned severities of the AIS lexicon, the ISS inherits the lack of precision of the AIS. This imprecision arises in part because only 6 severity grades are available in the AIS, and in part because injuries that are assigned the same severity often have observed mortalities that vary systematically across body regions.<sup>8,9</sup> The formula underlying ISS is also problematic. By design at most 3 and often fewer injuries contribute to the ISS, resulting in a loss of information. Additionally, the ad hoc “sum of squares” definition of the ISS score has no obvious justification. The performance of ISS in large data sets has been disappointing.<sup>10</sup> Although scoring systems should be monotonic, higher ISS scores are frequently associated with lower mortality.<sup>11</sup> Surprisingly, a model based on the single worst injury a patient has sustained better predicts mortality than does the ISS.<sup>12</sup>

We propose a new approach in predicting mortality for patients whose injuries are described in the lexicon of AIS codes that addresses both the problem of imprecise AIS severity measures and the ad hoc formula underlying the ISS. Using the National Trauma Data Bank (NTDB) we first derive an empirical severity value for each of 1322 different AIS injury codes that we call the model-averaged regression coefficient (MARC). These values provide a scale to compare the severity of individual injuries and a metric to incorporate individual injuries into mortality prediction models. We then investigate the performance of these MARC values in 3 exploratory mortality models: a single worst injury model, a model based on all injuries, and a final model that incorporates the 5 worst injuries as well as the interaction of the worst 2 injuries and a variable indicating whether the 2 worst injuries are to the same body region. We compare these 3 new models with a probit model based on the current standard, ISS, using measures of discrimination and calibration as well as the AIC.

## METHODS

### Data Source

This study used the January 2007 release (Version 6.1) of NTDB.<sup>13</sup> This data set included a total of 1,466,887 trauma patients hospitalized between 2001 and 2005. Available in-

formation included patient demographics, AIS codes, ISS, mechanism of injury (based on ICD-9-CM Ecodes), encrypted trauma center identifiers, and the outcome “survival to hospital discharge.” Patients without AIS codes (581,874) or with burns or nontraumatic diagnoses (eg, poisoning, drowning, suffocation) (33,210), missing or invalid data (23,928), or age younger than 1 year (96,467) were excluded from our analysis, as were patients who were dead on arrival to the trauma center (2641) or were transferred to another facility (71,273). We also required that patients be cared for at a facility that admitted at least 500 patients in at least 1 year to help ensure that data were collected at trauma centers with robust information systems, and this requirement excluded a further 35,077 patients. Some patients were excluded by more than 1 criterion. For reasons of privacy the NTDB assigns patients older than 89 years to a single age category. We imputed the age of 95 to these 7324 patients. ICD-9-CM Ecodes were mapped to 1 of the 6 mechanisms of injury (gunshot wound, stab wound, low fall, blunt injury, motor vehicle crash, and pedestrian injury) by an experienced clinical trauma surgeon (T.O.). The final data set included 702,229 patients who sustained a total of 2,207,823 instances of 1322 distinct AIS injury codes cared for at 1 of 206 trauma centers.

### Statistical Analysis

Individual severity measures for each of 1322 injuries in the AIS lexicon were derived from 60% of the data set by averaging coefficients obtained from 2 different probit regression mortality models. These MARC values provide an empiric measure of individual injury severities that can be used both to compare the severities of injuries and to incorporate individual injury severities into trauma mortality models. (We explain our preference for the probit model over the very similar logistic model in Appendix A. We elaborate on the 2 models and the averaging technique used to combine their point estimates in Appendix B.) To ensure validity of the mortality models that follow 60% of the total data set used to derive MARC values was excluded from subsequent model development and validation.

Four mortality models were estimated in 20% of the total data set. Mortality based on the ISS was modeled as a probit model using the ISS as its sole predictor. Three exploratory mortality models were developed using the MARC values calculated above to incorporate individual injuries as their severities: (1) a single worst injury model in which the MARC value for the worst injury (ie, the highest MARC value) was incorporated as the sole predictor in a probit mortality model, (2) an all-injury regression model that used the sum of all MARC values for an individual patient as the sole predictor in a probit mortality model, and (3) the trauma mortality prediction model (TMPM) that incorporated the 5 most severe (highest) MARC values ordered by severity and representing the 5 worst injuries as predictors in a probit model of mortality; the TMPM also included a binary variable indicating whether the 2 worst injuries were in the same body region and a variable reflecting the interaction of the severities of the 2 worst injuries as the product of their MARC values. All 4 models were subsequently estimated with 3 other predictors (age, gender, and mechanism of

injury) added to each model. Robust variance estimators (Huber-White) were used throughout to control for possible correlation of outcomes within individual trauma centers.

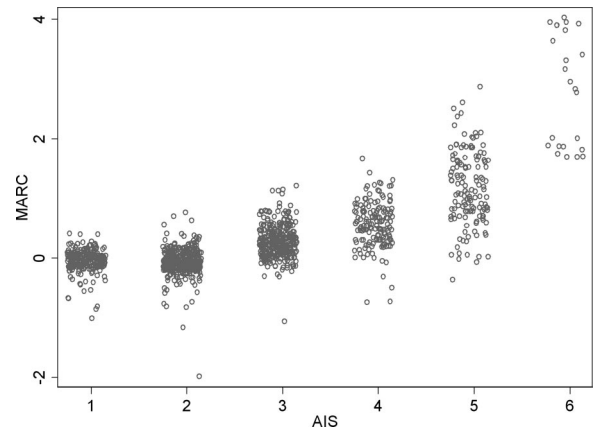
The 20% of the total data set not used for the estimation of MARC values or the development of exploratory mortality models was used to evaluate the performance of all models using measures of discrimination [the receiver operating characteristic curve (ROC)], and calibration (HL statistic and calibration curves). Bias-corrected 95% confidence intervals were calculated for the ROC and HL using 500 repetitions of a bootstrapping algorithm. Data management and statistical analysis were performed using Stata/MP (Version 10, Stata Corp LP, College Station, TX).

## RESULTS

The prevalence of specific injuries was very uneven in this data set. A few injuries were very common (eg, AIS code 210602.1 “minor skin laceration of the face” with 23,913 instances), but many injuries were rare. Twenty-four AIS codes (2%) occurred only once in the entire dataset. One hundred eighty-four codes (14%) occurred 10 or fewer times, and 684 AIS codes (52%) occurred fewer than 100 times. AIS codes with sparse representation were occasionally not life threatening (AIS 530499.1 “vagus nerve injury”) but often were serious [AIS 420608.5 “brachiocephalic (innominate) vein laceration, major, with air embolus right side”]. Most (64%) patients in the data set sustained more than 1 injury, with an average of 3.1 injuries per patient (range, 1–44). Overall mortality for the data set was 4.1%.

We derived an empiric severity (MARC value) for each of the 1322 AIS coded injuries in the data set (Appendix B). These MARC values spanned a range from  $-1.01$  for a trivial injury (AIS 440478.1: “thoracic spine, strain, acute with no fracture or dislocation”) to a value of  $4.03$  for an unsurvivable injury (AIS 113000.6: “Crush, massive destruction of both cranium (skull) and brain”). MARC values are roughly correlated with AIS severity values ( $R^2 = 0.745$ ). Because MARC values are continuous they provide finer granularity than do the 6 integer values available to the AIS severity (Fig. 1). Interestingly, some seemingly “severe” injuries (eg, brachial plexus avulsion) were assigned relatively low MARC values. We believe this is appropriate because, by design, MARC values reflect an injury’s propensity to cause mortality rather than its subjective severity.

The performance of all models is displayed in Tables 1 and 2. The ISS had significantly worse overall discrimination (lower ROC) and a worse (higher) AIC statistic than any MARC-based model, in part because the relationship of mortality to the raw ISS score is markedly nonmonotonic (Fig. 2). Although the calibration of ISS was statistically indistinguishable from that of some MARC-based models, the TPM surpassed ISS by every measure of model performance. Calibration curves underscore the superiority of the TPM over ISS (Fig. 3). The addition of age, gender, and mechanism of injury improved every model, but the TPM continued to dominate all other models (Table 2).



**FIGURE 1.** Model-averaged regression coefficients (MARC) and AIS severity for each of 1322 AIS injury codes. Although AIS severity and MARC values are roughly correlated ( $R^2 = 0.745$ ), MARC values are continuous and therefore allow finer distinctions in injury severity than the 6 severities assigned by the AIS lexicon. MARC values particularly resolve AIS values of 4, 5, and 6 into a broader range of severity. (To facilitate the display of 1322 AIS codes the horizontal axis has been “jittered” to make individual points more visible.)

## DISCUSSION

The prediction of outcome following injury requires that a lexicon of discrete injuries be created, that each injury be assigned a severity, and finally that a rule for combining multiple injuries into a single outcome prediction be defined. The ISS solves the first 2 of these problems by simply adopting the AIS categories and its expert-assigned severities, and address the final problem with a simple “sum of squares” rule. Although it is the current standard for the measurement of injury, theoretical objections to the AIS-ISS approach are numerous,<sup>11</sup> and the performance of ISS in large data sets has been disappointing.<sup>10</sup>

We developed an alternative to the ISS-based mortality model that we call the TPM. This model uses the AIS lexicon to divide the continuous landscape of trauma into discrete injuries, but relies on empirically derived injury severities (MARC values) and a more nuanced statistical model that reflects the clinical interplay of injuries to produce mortality. We find that the discrimination and calibration of the TPM are better than ISS at levels that are highly statistically significant and likely to be important to health care researchers and administrators.

Discrimination is the ability of a predictive model to correctly distinguish survivors from nonsurvivors and is quantified as the area under a plot of the sensitivity versus 1-specificity over all possible cutpoints for a model (ROC). Although this definition seems opaque, the ROC can equivalently be thought of as simply the probability that a randomly selected nonsurvivor will be predicted by the model to be less likely to survive than a randomly selected survivor, that is, the percentage of the time that the model “gets it right.” The ROC is constrained to be between 0.5 and 1, with a value of 1 representing perfect discrimination, and a value

**TABLE 1.** Model Performance: Anatomic Injury Models

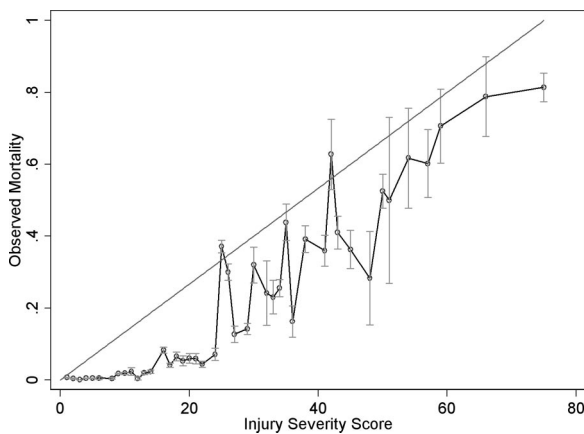
Model Description	ROC (95% CI)	HL Stat (95% CI)	AIC (95% CI)
Injury severity score	0.871 (0.866–0.875)	296 (227–363)	37,225 (36,437–37,908)
Single worst injury model	0.892 (0.888–0.897)	314 (236–387)	34,059 (33,372–34,735)
All injury model	0.900 (0.895–0.904)	325 (247–444)	32,968 (32,282–33,803)
TMPM	0.901 (0.898–0.906)	58 (30–84)	32,003 (31,317–32,662)

The ISS demonstrates the worst discrimination (lowest ROC value) of any model. Although all the remaining MARC-based models have better discrimination than ISS, 1 MARC based model, TMPM, is also much better calibrated (lower HL statistic) than ISS.

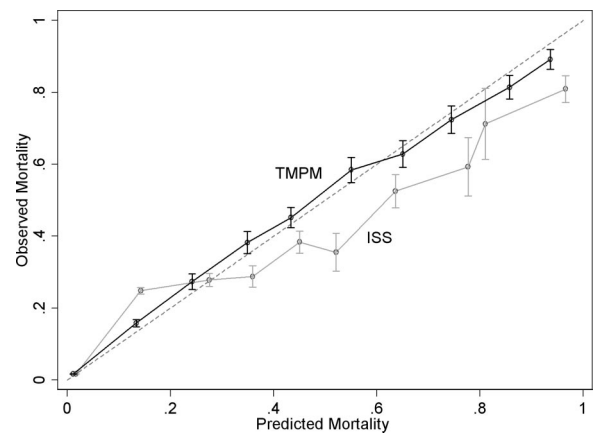
**TABLE 2.** Model Performance: Anatomic Injury Models Augmented With Age, Gender, and Mechanism of Injury

Model Description	ROC (95% CI)	HL Stat (95% CI)	AIC (95% CI)
Injury severity score + age + gender + mechanism	0.904 (0.900–0.908)	54 (23–73)	33,773 (33,117–34,509)
Single worst injury model + age + gender + mechanism	0.915 (0.911–0.918)	128 (75.6–170)	31,770 (31,124–32,432)
All injury model + age + gender + mechanism	0.921 (0.918–0.925)	159 (108–235)	30,546 (29,919–31,336)
TMPM + age + gender + mechanism	0.925 (0.922–0.929)	19 (4–27)	29,645 (29,048–30,290)

Although every models was improved by the addition of more predictors, TMPM continues to dominate all other models.



**FIGURE 2.** Injury severity score observed mortality. Observed mortality for each of the 44 possible ISS scores calculated using the 140,446 patients available in the validation data set. The 95% confidence intervals are based on the binomial distribution and are occasionally wide because some ISS values rarely occur.



**FIGURE 3.** Calibration curves for the trauma mortality prediction model (TMPM) and the injury severity score (ISS). Predicted probability of death for the ISS was calculated using a probit model with ISS as the sole predictor. The 44 possible ISS scores have been collapsed into 10 groups, which smoothes the underlying lack of monotonicity. Dotted reference line represents perfect calibration. (95% binomial confidence intervals for both models are based on the same test data set of 140,446 patients.)

of 0.5 representing random discrimination, that is, discrimination no better than a coin toss. Although the improvement in the ROC statistic offered by the TMPM over the ISS (0.90 vs. 0.87) may at first seem slight, because the ROC statistic has an upper bound of 1, this difference is actually a substantial advance toward a model with perfect discrimination.

Calibration is the tendency of a model's predicted outcomes to closely reflect the true outcome experience. For example, if a model with perfect calibration predicts a mortality of 75% for each of 100 patients, 75 of this group will die and 25 will go on to survive. Typically measured as the Hosmer Lemeshow statistic (with lower HL statistics indicating models with better calibration), calibration can also be

displayed as a calibration curve. Both the HL statistic and the calibration curves show the TMPM is also much better calibrated than the ISS (Tables 1 and 2, Fig. 3).

The use of empiric injury severity measures is largely responsible for improving model discrimination. A probit model based on a patient's single worst injury has a much higher ROC statistic when that single injury is incorporated as an MARC value than as the square of its AIS severity (ROC = 0.892 vs. ROC = 0.824, analysis not shown). Allowing more than 1 injury to contribute to outcome prediction further improves discrimination, as seen by compar-

ing the single worst injury model with a model that incorporates the MARC values for every injury a patient sustains (Table 1). Incorporating the relative severities of injuries and a term specifying the interaction of the worst 2 injuries into the model leads to the TPM. Although these last 2 refinements do not affect discrimination, both calibration and the AIC are substantially improved over the all injury model (Table 1). (The TPM considers at most only the 5 most severe injuries because so few patients in our data set had more than 5 important injuries that we were unable to obtain meaningful estimates for coefficients for any additional, less severe injuries.) The addition of age, gender, and mechanism improves the performance of all models, but the TPM continues to dominate all other models (Table 2).

The TPM estimates a coefficient for each of a patient's worst 5 injuries based on the relative severity of these injuries. The probit regression procedure assigns approximately twice the weight to the worst injury when compared with each of the remaining, less severe injuries. This relationship may explain, in part, the success of "single worst injury" trauma outcome models that incorporate only a single injury.<sup>12</sup> (In Table 3 the presence of an interaction term in the model obscures this simple relationship.) Further discounting of predicted mortality results if the 2 worst injuries occur in the same body region. The interaction of the worst 2 injuries (expressed as the product of their MARC values) is also associated with a lower probability of death (Table 3). Although these weightings arise naturally from the statistical regression procedure, they also have clinical face validity: trauma surgeons typically summarize a patient's clinical condition in terms of his or her worst 1 or 2 injuries, reflecting their understanding that death is caused by the few serious injuries, not the many less serious injuries a patient may have sustained. The origin of the discounting applied when the second worst injury is in the same body region as the first injury is more subtle, but the regression suggests that we might think of 2 injuries in a single region as a single injury, more extensive than the worst injury but not quite as severe as the sum of the 2 worst injuries. This observation seems reasonable if we think, for example, of a head injury on computed tomography scan. It may seem artificial to think of the subdural hematoma and the brain laceration that caused it

as 2 different injuries and more natural to think of the whole complex as a single injury, worse than either injury in isolation, but not quite as life threatening as the sum of 2 such separate injuries. Similar reasoning explains why the interaction term for the worst 2 injuries also serves to discount the likelihood of death: Although, a second injury increases the probability of death, if the severity of the second injury approaches that of the primary injury its contribution to the likelihood of death is reduced.

These insights into how injuries combine to produce death give a new appreciation of the venerable ISS. The TPM suggests that a patient's 2 or 3 worst injuries largely determine the likelihood of mortality, an insight approximated in the ISS which allows at most 3 injuries in its calculation. Further, TPM suggests that a second, lesser injury to a body region contributes less to mortality than does the primary injury, an insight that ISS incorporates by allowing only the single worst injury from each region to contribute to its calculation. In hindsight, the intuition of Baker et al concerning the relationship between injury and death seem prescient. The insightful approximations of the ISS have doubtless contributed to its popularity and longevity.

The performance of every model examined improved substantially when age, gender, and injury mechanism were added as prediction variables, and the degree of improvement was similar across all models. We believe that, because this additional information is almost always available, injury severity models should routinely make use of these additional predictors if an accurate estimate of the likelihood of survival is the goal. If a summary measure of the dose of trauma is desired, or if other predictors are unavailable, the TPM would be more appropriate, but perhaps expanded to include mechanism of injury as a predictor because mechanism can be thought of as contributing information about an injury's severity.

The addition of more kinds of information almost always improves model predictions, and this is true for the TPM. The addition of a single physiologic predictor, the Glasgow Coma Score motor component, dramatically improved the ROC of the TPM from 0.925 to 0.956 (analysis not shown). Physiologic predictors have been added to the ISS in the past, and the resulting TRISS model has been

**TABLE 3.** TPM Regression Coefficients

Predictor	Coefficient	Robust Std. Error	Z	P >  z	95% CI
MARC1	1.3138	0.0210	62.43	0.000	1.2725 to 1.3550
MARC2	1.5136	0.0498	30.42	0.000	1.4161 to 1.6112
MARC3	0.4435	0.0431	10.29	0.000	0.3591 to 0.5280
MARC4	0.4240	0.0456	9.29	0.000	0.3346 to 0.5134
MARC5	0.6284	0.0707	8.89	0.000	0.4898 to 0.7669
Same region	-0.1377	0.0114	-12.07	0.000	-0.1600 to -0.1153
MARC1 × MARC2	-0.6506	0.0270	-24.08	0.000	-0.7036 to -0.5976
Constant	-2.3281	0.0222	-104.70	0.000	-2.3717 to -2.2845

Coefficients for TPM model were recalculated based on all 280,716 patients not used to calculate MARC values. MARC1 represents a patient's worst injury (highest MARC value), MARC2 the second worst injury and so on. Same region is a binary variable (1 = 2 worst injuries are in the same AIS region, 0 otherwise) and MARC1 × MARC2 captures the correlation of a patient's 2 worst injuries as the product of their MARC values.

widely used. Because the TMPM is superior to the ISS, it is likely that a comprehensive model based on the TMPM as a measure of anatomic injury but augmented with several physiologic predictors (eg, respiratory rate, systolic blood pressure, Glasgow Coma Score) would prove superior to TRISS. Thus, although the basic TMPM is appealing precisely because it can be used when only anatomic injury information is available, TMPM can also serve as the firm basis for more elaborate outcome prediction models that incorporate physiologic parameters.

This study inherits the limitations of the NTDB. Data in the NTDB are contributed by self-selected trauma centers, and have not been validated by abstraction. Because this data set is not population based, TMPM may not be as accurate when applied to other groups of trauma patients, and in particular to patients admitted to nontrauma centers. Additionally, the outcome “death prior to hospital discharge” reported in the NTDB misses patients who survive to discharge but subsequently die of their injuries. Finally, a general objection might be raised to basing mortality predictions on MARC values because these values may change over time, perhaps as the result of new therapies. This objection might also be made to any other trauma mortality model, of course, including the ISS, based as it is on the AIS lexicon severities. An advantage of MARC values, however, is that it is straightforward to periodically recalculate MARC values based on a more recent (or more relevant) data set.

Prediction models of mortality following trauma require estimates of severity for every individual injury. The expert-assigned severities of the AIS were necessary in the past because the calculation of empirical values for each of the 1322 injuries in the AIS lexicon was simply infeasible. Recent increases in computing power coupled with the development of a massive trauma data set, the NTDB, now allow the empirical calculation of these individual severities. We find that such severities coupled with a probit regression model, TMPM, predict mortality following trauma much better than does the current standard, ISS. A further advantage of TMPM is that its predictions directly express the probability of death. The ISS, by contrast, only provides a relative score that requires further statistical manipulation and a reference data set to arrive at a mortality prediction. Because it is both more accurate and directly interpretable, we believe that the TMPM should replace ISS.

Twenty years ago Copes et al reported the erratic performance of the ISS and voiced “. . . concern about the use of the ISS in applications with important consequences.”<sup>9</sup> The use of report cards to compare trauma care providers and centers seems imminent, and may have profound consequences. We must be certain that such comparisons are based on the best risk adjustment available.

## REFERENCES

1. Ownes PL, Russo CA, Stocks C. Frequency and Cost of Hospital Admissions for Injury, 2004, HCUP Statistical Brief No. 18. Available at: <http://www.hcup-us.ahrq.gov/reports/statbriefs/sb18.pdf>. Accessed August 7, 2007.
2. Shafi S, Friese R, Gentilello L. Moving beyond personnel and process: a case for incorporating outcome measures in the trauma center designation process. *Arch Surg*. 2008;143:115–119.

3. Rating the severity of tissue damage: I. The abbreviated injury scale. *JAMA* 1971;215:277–280.
4. Baker SP, O'Neill B, Haddon W, et al. The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. *J Trauma*. 1974;14:187–196.
5. Boyd CR, Tolson MA, Copes WS. Evaluating trauma care: the TRISS method. *J Trauma*. 1987;27:370–378.
6. Champion HR, Copes WS, Sacco WJ, et al. A new characterization of injury severity. *J Trauma*. 1990;30:539–545.
7. Osler T, Baker SP, Long W. A modification of the injury severity score that both improves accuracy and simplifies scoring. *J Trauma*. 1997;43:922–926.
8. Beverland DE, Rutherford WH. An assessment of the validity of injury severity score when applied to gunshot wounds. *Injury*. 1983;15:19–22.
9. Copes WS, Champion HR, Sacco WJ, et al. The injury severity score revisited. *J Trauma*. 1988;28:69–77.
10. Meredith JW, Gregory E, Kilgo PD, et al. A comparison of the abilities of nine scoring algorithms in predicting mortality. *J Trauma*. 2002;53:621–629.
11. Kilgo PD, Meredith JW, Hensberry R, Osler T. A note on the disjointed nature of the injury severity score. *J Trauma*. 2004;57:479–487.
12. Kilgo PD, Osler T, Meredith W. The worst injury predicts mortality outcome the best: rethinking the role of multiple injuries in trauma outcome scoring. *J Trauma*. 2004;56:928–934.
13. National Trauma Data Bank Web Page. Available at: <http://www.facs.org/trauma/ntdb.html>. Accessed August 11, 2007.
14. Stukel TA. Generalized logistic models. *J Am Stat Assoc*. 1988;83:426–431.
15. Sauerbrei W, Royston P. Building multivariable prognostic and diagnostic models: Transformation of the predictors by using fractional polynomials. *J R Stat Soc Ser A*. 1999;162:71–94.

## APPENDIX A

### A Comparison of the Probit Model and Logistic Models

Originally conceived to model the effect of toxins on insects, the probit (“PROability unit”) model antedates the logistic model by a decade. The logistic model has become far more popular, however, because the coefficients it produces have a natural interpretation and it has a closed mathematical form that simplifies calculations. The 2 models differ in that the underlying distribution of tolerance to injury is assumed by the logistic model to have a logistic distribution, whereas the probit model assumes the underlying tolerance to follow the more familiar normal distribution. Because these 2 distributions are actually very similar, the logistic and probit models usually produce indistinguishable results, even in moderately large data sets. The very large size of the data set used in this study provided sufficient resolution to see that the “thinner tails” of the probit distribution more accurately model the relationship between the dose of trauma and mortality. The superior fit of the probit model is suggested by its significantly lower Hosmer Lemeshow statistic, but was explicitly tested using Stukel’s test<sup>14</sup> (analysis not shown). Because modern computers easily handle the probit model and because we are interested in precise prediction rather than the interpretation of model coefficients, we chose the accuracy of the probit model over the convenience of the logistic model for this analysis.

**APPENDIX B**

**Model Averaging and the Variance of Model 2 (Region-Severity Model)**

One way to arrive at the empirical severity of the 1322 possible AIS injury codes is to create a probit model that predicts mortality using the 1322 possible AIS injury codes as binary predictors. If such a model could be estimated the coefficient of each of the possible injuries would reflect that injury’s propensity to result in death. Unfortunately, this approach fails because many AIS codes have such sparse representation that their coefficients cannot be accurately estimated, or cannot be estimated at all due to perfect correlation or high collinearity.

To address this problem we created 2 separate probit regression models. Model 1 used all 1322 possible AIS codes as binary predictors of mortality. (Although of no intrinsic interest at this stage of modeling, 4 additional variables were included in both models to decrease variance and potential bias: age (as age<sup>3</sup> and log(age) × age<sup>3</sup>, suggested by fractional polynomial analysis<sup>15</sup>), gender, mechanism of injury, and a separate indicator variable for each trauma center.) Unfortunately, 206 of these coefficients could not be estimated because they were uniformly associated with survival or death (204 coefficients), or because of collinearity (2 coefficients). These injuries were assigned MARC values generated by MARC generating Model 2, which was based on 47 region severity predictors all of which could be estimated. Although this substitution was required for 16% of all AIS codes, because these injuries were rare this approximation affected only 1% of patients. For the remaining AIS codes both MARC generating models yielded estimates, and for these codes the 2 estimates were combined as an inverse variance weighted average to produce a single estimate as follows:

$$\text{Coef}_{\text{MARC}} = \frac{\frac{1}{\text{Var}(\text{Model 1})} \text{Coef}_{\text{Model 1}}}{\frac{1}{\text{Var}(\text{Model 1})} + \frac{1}{\text{Var}(\text{Model 2})}} + \frac{\frac{1}{\text{Var}(\text{Model 2})} \text{Coef}_{\text{Model 2}}}{\frac{1}{\text{Var}(\text{Model 1})} + \frac{1}{\text{Var}(\text{Model 2})}}$$

The variance of Model 2 in this expression requires elaboration. Although probit regression software provides a “variance” for each coefficient associated with each of the 47 possible region/severity groupings, these values represent the variance for the single overall average mortality for all patients within that region-severity group. Unfortunately, to average coefficients from our 2 models we require not this

overall variance but rather the variance of the individual point estimates for all the individual AIS level coefficients nested within a given region/severity grouping. Because these point estimates are themselves the result of a regression model, they each have variances that must also be accounted for. The following expression is simply the definition of variance, but calculated using the inverse variance weighted average of AIS coefficients as the expected value for the AIS coefficients within region-severity groups while simultaneously inverse variance weighting each term contributing to the resulting summation.

$$\text{Var}(\text{RScoef}_m) = \frac{1}{N_i - 1} \sum_{j=1}^{N_i} N_i \times W_j \times [(\text{AIScoef}_j - E(\text{AIScoef}))^2]$$

where *N* is the number of AIS codes in RS region *m* and *W<sub>j</sub>* is the weight of the contribution of AIScoef<sub>*j*</sub> to the variance of RScoef<sub>*m*</sub>.

$$W_j = \frac{\frac{1}{\text{Var\_AIScoef}_j}}{\sum_{r=1}^{N_i} \frac{1}{\text{Var\_AIScoef}_r}}$$

$$E(\text{AIScoef}) = \sum_{j=1}^{N_i} W_j \times \text{AIScoef}_j$$

**APPENDIX C**

**The Mechanics of Calculating the TMPM**

Although the definition of MARC values involves a somewhat mathematical approach, once MARC values are available the actual calculation of the TMPM mortality prediction for individual patients uses only a simple regression model, a technique that is likely to be familiar to most readers. Calculation of the mortality predicted by TMPM proceeds in 2 stages: First, each AIS code is replaced by a measure of its severity (its MARC value). The TMPM predicted mortality is then simply:

$$P(\text{death}) = \Phi[C_0 + C_1 \times I_1 + C_2 \times I_2 + C_3 \times I_3 + C_4 \times I_4 + C_5 \times I_5 + C_6 \times S + C_7 \times I_1 \times I_2]$$

where *P*(death) is the mortality predicted by the TMPM and  $\Phi$  is the standard normal distribution function (available in statistical software packages and Excel).

*I*<sub>1</sub>, *I*<sub>2</sub>, . . . , *I*<sub>5</sub> are the MARC values for the 5 worst injuries, ordered with the greatest MARC value (worst injury)

first, the second greatest MARC value second, up to the fifth worst injury. (The TMPM considers only the worst 5 injuries that a patient has sustained. Other, less severe, injuries are ignored.) Note that the term  $C_7 \times I_1 \times I_2$  represents the interaction of the worst and second worst injuries that a patient has sustained.

S is an indicator variable set equal to 0 if the worst 2 injuries occur in different body regions and set equal to 1 if the worst 2 injuries occur in the same body region. (The AIS coding scheme ensures that 2 injuries are in the same body region if their AIS codes have the same first digit.)

$C_0, C_2, \dots, C_7$  are coefficients as follow:

$$C_0 = -2.3281$$

$$C_1 = 1.3138$$

$$C_2 = 1.5136$$

$$C_3 = 0.4435$$

$$C_4 = 0.4240$$

$$C_5 = 0.6284$$

$$C_6 = -0.1377$$

$$C_7 = -0.6506$$

Although it is straightforward to calculate the TMPM probability of death, it does require a replacement step (replace AIS codes with their respective MARC values), a sorting step (to arrange MARC values in order from highest to lowest for each patient), a step to determine if the worst 2 injuries are in the same body region, and finally a step to calculate the final prediction based on the previous steps, the provided coefficients ( $C_0, C_1, \dots, C_7$ , above) and the model expression for  $P(\text{death})$ . An Excel spreadsheet (macro) is available from the authors to perform these steps automatically. The TMPM can also be calculated in virtually any computing language, requiring at most only a few dozen lines of computer programming. But, for researchers who wish to be able to quickly compute TMPM values (or perhaps, if they have programmed their own routine to calculate the TMPM, to check their program's output), the TMPM calculator we have created may prove useful.