Journal of Surgical Research **166,** 40–44 (2011)
doi:10.1016/j.jss.2010.04.012

ASSOCIATION FOR ACADEMIC SURGERY

Validating the Injury Severity Score (ISS) in Different Populations: ISS Predicts Mortality Better Among Hispanics and Females

O. B. Bolorunduro, M.D., M.P.H.,* C. Villegas, M.P.H.,† T. A. Oyetunji, M.D., M.P.H.,* E. R. Haut, M.D.,† K. A. Stevens, M.D., M.P.H.,† D. C. Chang, M.P.H., M.B.A., Ph.D.,* E. E. Cornwell III, M.D.,* D. T. Efron, M.D.,† and A. H. Haider, M.D., M.P.H.†,[1]

*Trauma Outcomes Research Group, Department of Surgery Howard University College of Medicine, Washington, DC; and †Center for Surgery Trials and Outcomes Research, Department of Surgery Johns Hopkins University School of Medicine, Baltimore, Maryland

Submitted for publication January 9, 2010

Introduction. **The Injury Severity Score (ISS) is the most commonly used measure of injury severity. The score has been shown to have excellent predictive capability for trauma mortality and has been validated in multiple data sets. However, the score has never been tested to see if its discriminatory ability is affected by differences in race and gender.**

Objective. **This study is aimed at validating the ISS in men and women and in three different race/ethnic groups using a nationwide database.**

Methods. **Retrospective analysis of patients age 18–64 y in the National Trauma Data Bank 7.0 with blunt trauma was performed. ISS was categorized as mild (<9,) moderate (9–15), severe (16–25), and profound (>25). Logistic regression was done to measure the relative odds of mortality associated with a change in ISS categories. The discriminatory ability was compared using the receiver operating characteristics curves (ROC). A *P* value testing the equality of the ROC curves was calculated. Age stratified analyses were also conducted.**

Results. **A total of 872,102 patients had complete data for the analysis on ethnicity, while 763,549 patients were included in the gender analysis. The overall mortality rate was 3.7%. ROC in Whites was 0.8617, in Blacks 0.8586, and in Hispanics 0.8869. Hispanics have a statistically significant higher ROC (*P* value < 0.001). Similar results were observed within each age category. ROC curves were also significantly higher in females than in males.**

Conclusion. **The ISS possesses excellent discriminatory ability in all populations as indicated by the high ROCs.**

Key Words: **trauma; ISS (injury severity score); scoring systems; outcomes.**

INTRODUCTION

Trauma is the leading cause of morbidity and mortality in young adults [1]. In 2006, unintentional injury accounted for over 52,000 deaths among patients aged between one and 45 y in the United States [2]. A detailed understanding of two major factors is essential to predict clinical outcomes post-trauma: the severity of injury [3–5] and the patient's physiologic reserve [6]. A patient's reserve is a composite measure of the influence of age, gender, co-morbidities, insurance, complications, and other unmeasured factors that altogether explain why each patient is different from the other. This paper focuses on the measurement of injury severity in different populations and how this predicts outcomes.

Several scoring methods are used to assess injury severity; they are divided into anatomic, physiologic, and combined scoring systems. Anatomic systems include ISS, New Injury Severity Score (NISS), and Mangled Extremity Severity Score (MESS). Examples of physiologic scores are the Glascow Coma Score (GCS) and Revised Trauma Score (RTS). The ASCOT (A Severity Characterization of Trauma) and Trauma Related Injury Severity Score (TRISS) are examples of combined scoring systems [1, 3. 6–8].The Injury Severity Score (ISS) is the foremost and gold standard test for assessing injury severity [11, 26]. It was developed to measure the severity of injury based on

[1] To whom correspondence and reprint requests should be addressed at Center for Surgery Trials and Outcomes Research (CSTOR), Department of Surgery, Johns Hopkins School of Medicine, 600 N. Wolfe Street, Halsted 610, Baltimore, MD 21287. E-mail: ahaider1@jhmi.edu.

0022-4804/$36.00

injury characteristics [3]. The ISS has been used to account for patient case mix in trauma care evaluation, identifying comparable populations in outcomes research, and could be used in benchmarking and audit purposes [3–5]. Developed in 1974 by Baker *et al.*, the ISS is derived from the Abbreviated Injury Scale (AIS) [3, 4]. The AIS score is on an ordinal scale ranging from 1 to 6; the body is categorized into six distinct regions, and each region is given an AIS score. The ISS is obtained by taking the sum of the squares of the AIS values of the three most severely injured regions of the body. The new methods of scoring listed above were mostly developed to correct the perceived inadequacies of the ISS. These include its failure to account for multiple severe injuries in the same region, underestimation of injury severity in penetrating injuries, and the need for MRI or angiography to give accurate ISS scores in some cases [8–10]. Despite these limitations, the ISS remains the most commonly used scoring system in trauma research [11].

ISS has been shown to have excellent predictive capability for trauma mortality and has been validated in different datasets [3–5]. But despite its prevalent use, it is unknown if its discriminatory properties are affected by race and gender. This becomes especially important with recent evidence showing ethnic and gender differences in mortality after trauma. Haider *et al.* demonstrated that Black and Hispanic race was independently associated with mortality after trauma [27]. George *et al.* showed that females were associated with reduced risk of mortality after trauma. It is important to delineate if these results are due to actual race and gender differences or if they are due to differential discriminatory ability of the ISS in different populations [28]. This study aims to assess the discriminatory ability of the ISS in different populations, and so validate its use in all ethnic and gender groups.

METHODS

This is a retrospective analysis of trauma patients in the National Trauma Data Bank (NTDB), ver. 7.0, from 2002 to 2006. This version contains data from over 1.8 million patients from over 700 hospitals during this period. Age criteria included patients between 18 and 65 y with blunt trauma. We limited the study to blunt trauma cases because the ISS has been shown to have reduced discriminatory ability in penetrating injuries [12]. The Injury Severity Score (ISS) was categorized as mild (<9), moderate (9–15), severe (16–24), and profound (≥25). Institutional review board approval was obtained from the Johns Hopkins University School of Medicine.

Analysis was first done on all patients, then by gender, and then subsequently on each of three races—Whites, Blacks, and Hispanics. Logistic regressions were done to measure the relative odds of mortality associated with each unit increase in ISS. In order to predict the discriminatory ability of the ISS in these groups, the ISS was used as the only independent variable in each regression [12]. Separate regression analyses were conducted for each gender and race. The predicted probability of mortality for each population was determined, and from these, the discriminatory ability was compared using the receiver operating characteristic curves (ROC)[13–16].

The receiver operating characteristic (ROC) is used to evaluate the ability of a test to discriminate between those who will develop the outcome of interest, in this case mortality. It is derived from calculating the sensitivity and specificity of the test. A ROC curve represents a graphical picture of the tradeoff between the false negative and false positive rates at all possible cut offs. It also shows the tradeoff between specificity and sensitivity [14, 15].

To objectively compare the curves, the null hypothesis of equal "area under the ROC curves (AUC)" was tested, and a P value was calculated. This objectively compared the AUC in males to females and the AUC in different races. Age stratified analyses were also conducted to compare the statistical properties of the ISS in every age group including the elderly.

The AUC was classified as failed if <0.60, poor if 0.60–0.70, fair if between 0.70 and 0.79, good if between 0.80 and 0.89, and excellent if >90 [17]. Calibration was assessed using Hosmer-Lemeshow' goodness of fit statistic, which compares the difference between the models predicted and observed event rates. Good fit was defined as a P value > 0.1. All statistical analyses were performed using STATA MP ver. 10 (StataCorp LP, College Station, TX) [18].

RESULTS

Table 1 shows the demographic distribution of our population. Of the 1.8 million patients in the NTDB ver. 7.0, 1.1 million were between ages 18 and 65, and 906,500 of these suffered injuries from blunt trauma; 872,102 (96%) patients had complete information for our regression analysis on gender, and 763,549 (84%) had complete information for the race analysis. The median age was 37 y, (Interquartile range IQR 25–48), about 23% of them were between age 18 and 24 y. Males

TABLE 1

Baseline Demographics and Injury Severity Characteristics

	n	Percentage
Age		
<25	205,577	22.68%
25–34	195,531	21.57%
35–44	200,595	22.13%
45–54	183,363	20.23%
55–64	121,464	13.40%
Gender		
Male	627,770	69.25%
Female	277,734	30.64%
Race		
White	569,886	62.86%
Black	123,321	13.60%
Hispanic	97,382	10.74%
Asian	13,929	1.54%
Native American/Alaskan Native	6,625	0.73%
Others	42,771	4.72%
ISS categories		
<9	412,739	45.53%
9–14	257,973	28.46%
15–24	120,404	13.28%
>25	91,171	10.06%
Shock	35,950	3.97%
Mortality	27,294	3.01%