



Published in final edited form as:

Injury. 2009 September ; 40(9): 993–998. doi:10.1016/j.injury.2009.03.004.

REVISITING THE VALIDITY OF APACHE II IN THE TRAUMA ICU: IMPROVED RISK STRATIFICATION IN CRITICALLY INJURED ADULTS

Lesly A Dossett, MD MPH^{*}, Leigh Anne Redhage, MD^{*}, Robert G Sawyer, MD[^], and Addison K May, MD^{*}

Lesly A Dossett: lesly.dossett@vanderbilt.edu; Leigh Anne Redhage: leigh.redhage@vanderbilt.edu; Robert G Sawyer: RWS2K@hscmail.mcc.virginia.edu; Addison K May: addison.may@vanderbilt.edu

^{*}From the Division of Trauma & Surgical Critical Care, the Department of Surgery, Vanderbilt University Medical Center; *Nashville, TN*

[^]From the Department of Surgery, University of Virginia Health System; *Charlottesville, VA*

Abstract

Background—Quality and benchmarking initiatives highlight the need for accurate stratified risk adjustment. The stratification of trauma patients has relied on scores specific to trauma populations. While the Acute Physiologic and Chronic Health Evaluation (APACHE) II score has been considered "invalid" in the trauma population, we hypothesized that APACHE II would more accurately predict outcomes in critically injured patients in whom commonly used trauma scores have inherent limitations.

Methods—A prospective cohort of critically injured patients was enrolled. Severity scores and their sub-components were collected, and in-hospital mortality was assessed. The area under the receiver operating characteristic (AUROC) curve was used to determine the predictive value of each score. Logistic regression estimated the odds of death associated with incremental changes in severity scores and their subcomponents.

Results—1,019 patients were available for analysis. APACHE II was the best predictor of mortality (AUROC 0.77 versus AUROC 0.54 for ISS and 0.64 for TRISS). A unit increase in APACHE II was associated with an OR of death of 1.18 (95% CI 1.14 – 1.22). The components of APACHE II that contributed the most to its accuracy included temperature, serum creatinine and the Glasgow Coma Scale (GCS).

Conclusion—Critically injured patients have physiologic derangements not accurately accounted for by commonly used trauma scores. In this subset a more general ICU scoring system is useful for risk adjustment for research, administrative and quality improvement purposes.

© 2009 Elsevier Ltd. All rights reserved.

CORRESPONDING AUTHOR Lesly A Dossett, MD MPH, lesly.dossett@vanderbilt.edu, 404 Medical Arts Building, 1211 21st Avenue South, Nashville, TN 37212, P: (615) 936-0177, F: (615) 936-0185.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

REPRINTS Addison K May, MD, Vanderbilt University Medical Center, 404 Medical Arts Building, 1211 21st Avenue South, Nashville, TN 37212

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to disclose.

INTRODUCTION

A growing focus on health quality and benchmarking has highlighted the need for accurate severity scoring systems for stratified risk adjustment¹². Accurate severity scoring allows for comparison of clinical outcomes across centers and countries, and provides a means for the comparison of interventions and outcomes in a research setting. As the number and complexity of demographic and clinical characteristics impacting clinical outcomes has grown, so too has the number and complexity of severity scoring systems^{1, 10, 16, 17, 20–22, 24, 26, 27}.

The risk stratification of trauma patients has traditionally focused on anatomic or physiologic scores specific to trauma populations. This stems from the belief that the trauma patient population is inherently different from the general patient population. Trauma patients are considered younger, healthier and plagued with unique “disease” patterns; all factors which many believe limit the usefulness of scores addressing broader categories of patients and diseases. Several systems have been developed for stratification of trauma patients and these scores have been previously reviewed⁷. Among the most commonly used are the injury severity score (ISS) and the Trauma and Injury Severity Score (TRISS). ISS is a strictly anatomic scoring system developed in 1974 for the purpose of describing the multiple trauma patient³. The lack of physiologic data led rapidly to revisions of the ISS, as it was shown to be inferior to severity scores which incorporate these parameters⁹. TRISS is one such revision which takes into account the ISS, but also incorporates the Revised Trauma Score (RTS), the patient’s age and the mechanism of injury to determine a patient’s predicted survival⁴. For many years, TRISS has been the premier quantitative measure of injury treatment quality, but a number of recent studies have identified the limitations to this scoring system, particularly among the critically injured.

Despite the known and inherent limitations in trauma specific scores, they continue to dominate trauma registries and the literature since scores used in general critical care populations, such as APACHE II¹⁴, are considered invalid in the trauma patient. Criticisms of the use of APACHE II in the trauma population have been based primarily on the poor correlation between APACHE II and ISS or TRISS, and the inability of APACHE II to accurately predict hospital or intensive care unit length of stay¹⁹. Importantly, the criticism has not been based on an inability to predict death. In fact, when APACHE II has been evaluated as a predictor of clinical outcomes in trauma patients it has proven to be a useful predictor, particularly those who are critically injured^{2, 25}.

Our goal was to evaluate the ability of commonly used severity scores and their subcomponents to predict death in a large prospective cohort of critically ill trauma patients. We hypothesized that in the trauma ICU, the injury and illness severity scores incorporating physiological data would be superior to anatomic only systems, and that APACHE II would be a useful predictor of death in this select group of trauma patients.

METHODS

Study design and participating centers

This study represents a secondary analysis of a multi-institutional prospective cohort study of critically ill or injured surgical patients. The primary purpose of the study was to determine the role of gender and sex hormones on outcomes in critically ill patients. All patients 18 years of age or older, admitted to the Surgical or Trauma Intensive Care Units (ICUs) of Vanderbilt University Medical Center and the University of Virginia Health Sciences Center were eligible for enrollment. In order to capture a patient population at high risk of death and other secondary complications, patients who died before 48 hours were excluded as were patients discharged from the ICU prior to 48 hours. The minimum stay of 48 hours was intended to exclude

postoperative surgical patients with short observational stays as well as patients who died rapidly from illness beyond the aid of modern critical care. Patient care was at the discretion of the attending physician according to established critical care protocols in the respective ICUs. As a part of this study, illness and injury severity scores were collected prospectively at the time of admission by dedicated research oriented registered nurses at each facility. Information sources included the patients and their families, nursing flow sheets, paper and electronic medical records, and treating healthcare providers. These data were entered into a custom computer database designed specifically for this project.

APACHE II

APACHE II¹³, a modification of the APACHE¹⁵, assigns numerical values (0 to 4 with high scores indicating more severe illness) to 12 clinical and biochemical parameters: temperature, mean arterial blood pressure, heart rate, respiratory rate, oxygenation, arterial ph, serum sodium, potassium and creatinine, WBC and GCS. These combined score from these 12 parameters makes up the Acute Physiology Score (APS) of APACHE II. Points are also assigned for age group and preexisting illness. Combined scores below 10 suggest relatively mild illness while score above 15 indicate moderate to severe illness. According to the APACHE II definition, scores were calculated based on the worst physiologic parameters within the first 24 hours following hospital admission.

ISS

The injury severity score is based on an anatomic grading of injury severity³. Each of six body regions (head & neck, face, chest, abdomen, extremity and external) is assigned an abbreviated injury score (AIS). The three most severely injured regions have their scores squared and summed to produce the ISS. ISS ranges from 0 to 75. If any region is assigned a 6 (un-survivable) the ISS automatically becomes 75.

TRISS

TRISS determines the probability of survival from the ISS and Revised Trauma Score (RTS) using the formula

$$P_s = 1 / (1 + e^{-b})$$

where b is calculated using coefficients from multiple regression of the Major Trauma Outcomes Study. These coefficients differ according to patient age and the mechanism of injury (blunt versus penetrating). The RTS is a physiologic score based on the Glasgow Coma Scale, the systolic blood pressure and the respiratory rate upon first contact with the patient which is heavily weighted toward the GCS^{5, 6}. TRISS scores were determined from data at the time of admission to the hospital.

Statistical analysis

Normally distributed continuous variables were summarized by reporting the mean and 95% confidence intervals and compared using two sample *t*-tests for independent samples. Continuous variables that were not normally distributed were presented by reporting the median and interquartile ranges (IQR) and compared using the Wilcoxon Rank Sum test. Differences in proportions were compared using a chi-square (X^2) or Fischer's exact test. Spearman rank correlation coefficients were used to estimate the correlation between the various severity scores, and with the severity scores and other outcomes such as hospital and ICU length of stay. To estimate the odds ratio (OR) of death associated with incremental

increases in the severity scores and their subcomponents, univariate logistic regression was used.

To compare the predictive ability of the severity scores, a computation of the area under the receiver operating characteristic (AUROC) curve was performed. The ROC curve describes graphically the sensitivity and specificity of a given diagnostic test. As the AUROC curve approaches 1.0, it becomes more accurate; as the area approaches 0.5, it becomes random and of little diagnostic value. The AUROC is reported with the associated 95% confidence interval (CI); ROC curves were compared using the *roc comp* command of STATA which is based on the methods of DeLong *et al*⁸. STATA version 9.2 (STATA Corp., College Station, Texas, USA) was used for analysis. Tests for statistical significance were two sided with an alpha of 0.05.

The study was approved by the Institutional Review Boards of both Vanderbilt University Medical Center and the University of Virginia Health Sciences Center. All data is maintained in a secure, password protected database that is HIPAA-compliant. All patient information is de-identified prior to analysis and reporting.

RESULTS

Patient Demographic and Clinical Characteristics

1,019 patients met entry criteria and were available for analysis. There were 138 deaths for an overall mortality rate of 14%. The demographic and clinical characteristics of patients by outcome are summarized in Table 1. Non-survivors were older and had higher APACHE II scores (23 ± 5.5) compared to survivors (16 ± 5.5 , $p < 0.001$). Other components of APACHE II which were *statistically* different between the groups included temperature and serum creatinine, but these variables were not impressively clinically different.

The GCS subcomponent was different between the groups (median of 8 for survivors versus 5 for non-survivors, $p < 0.001$). GCS exhibited a bimodal distribution with patients with at the extreme scores (>10 and <5) exhibiting the highest mortality rates, while the lowest mortality was observed in patients with GCS scores in the 7–10 range. A GCS score of 9 was associated with the lowest overall mortality (compared to a GCS of 3, OR 0.17, 95% CI 0.06 – 0.50, $p = 0.001$). This trend was also seen among the AIS head or neck scores with an AIS head score of 2 being associated with the lowest mortality (compared to AIS head 0, OR 0.33 (95% CI 0.01 – 10.2), but this difference was not statistically significant.

ISS was not statistically different between survivors and non-survivors (31 versus 33, $p = \text{NS}$). Of the abbreviated injury scores (AIS) which make up the ISS, AIS head or neck was the only subcomponent score which was different between the groups [AIS Head 3 (2–4) for survivors versus 4(3–5) for nonsurvivors, $p < 0.001$]. The mean TRISS was 0.69 ± 0.30 and the median TRISS was 0.85 (IQR 0.46 – 0.95); the median value corresponded to the observed survival of 86%. The predicted survival as estimated by TRISS was significantly higher for the survivors (0.85 versus 0.48, $p < 0.001$). There were 144 unexpected survivors according to TRISS (TRISS < 0.50 with patient survival), and 91 unexpected deaths (TRISS > 0.50 with patient death).

APACHE II was weakly correlated with both ISS ($r = 0.17$, $p < 0.001$) and TRISS ($r = 0.31$, $p < 0.001$) while ISS and TRISS demonstrated considerable correlation ($r = 0.67$, $p < 0.001$). None of the scores were well correlated with either hospital or ICU length of stay (LOS). The best correlation was between APACHE II and ICU LOS ($r = 0.19$, $p < 0.001$).

Relationship with Mortality and Incremental Changes in Severity Scores

After converting the actual value of individual parameters into the APACHE II subcomponent scores, the OR of death associated with an incremental increase in each sub-component score was estimated. Each unit increase in APACHE II was associated with an OR of death of 1.18 (95% CI 1.14 – 1.22). The serum chemistries were all associated with significantly increased odds of death—serum sodium (per point, OR 1.55, 95% CI 1.16 – 2.07, $p=0.003$), potassium (per point, OR 1.30, 95% CI 1.07 – 1.59, $p=0.009$) and creatinine (per point, OR 1.20, 95% CI 1.01 – 1.44, $p=0.04$). These relationships indicate that as these chemistries deviate from normal, the number of points assigned to the category increases, and this increase is associated with a worse outcome. When forced into a simple logistic regression model, unit increases in GCS were associated with an OR of death of 1.16 (95% CI 1.10 – 1.24, $p<0.01$), but again, the relationship between GCS and subsequent mortality is bimodal with the lowest mortality among those patients with scores between 7 and 10, and forcing GCS into a logistic regression model inadequately captures this relationship. The combined physiologic parameters (the APS component) (OR 1.15, 95% CI 1.11 – 1.19, $p<0.001$) and the age component (OR 1.26, 95% CI 1.16 – 1.37, $p<0.001$) were also significantly associated with subsequent mortality. None of the other components were significantly associated with mortality.

Incremental increases in ISS were weakly associated with increased mortality (per point, OR 1.01, 95% CI 1.00 – 1.03, $p=0.05$). The head or neck sub-component was the only AIS to be independently associated with increased mortality (per point, OR 1.20, 95% CI 1.05 – 1.38, $p=0.009$). Incremental changes in TRISS were also associated with death (per 0.1 point decrease, OR 1.19, 95% CI 1.12 – 1.27, $p<0.001$).

Ability to Predict Death of Severity Scores and their Subcomponents

APACHE II (AUROC 0.77, 95% CI 0.72 – 0.81) was superior to both ISS (AUROC 0.54, 95% CI 0.48 – 0.60, $p<0.001$) and TRISS (AUROC 0.64, 95% CI 0.58 – 0.71, $p<0.001$) in predicting death (Table 2, Figure 1). Of the individual variables, age, temperature (AUROC=0.57, 95% CI 0.51 – 0.62), GCS (AUROC=0.64, 95% CI 0.59 – 0.69), and AIS head or neck (AUROC=0.61, 95% CI 0.54 – 0.67) all had some predictive ability. The combined APS predicted death with reasonably good accuracy (AUROC 0.70, 95% CI 0.66–0.75). The ability of APACHE II to predict death was improved in the penetrating trauma patient (AUROC 0.81 95% CI 0.65 – 0.97) as compared to the blunt trauma patient (AUROC 0.76, 95% CI 0.71 – 0.81), but this difference was not statistically significant ($p=0.54$) (Table 3). Both ISS and TRISS trended toward improved predictive ability in blunt trauma patients, but again these differences were not statistically significant. As the original study cohort included trauma and non-trauma patients, the predictive ability of APACHE II was compared. In the trauma population APACHE II had an area under the receiver operating characteristic curve for predicting mortality of (AUROC 0.77, 95% CI 0.73,0.81); this was statistically superior to its predictive ability in the non-trauma population (AUROC 0.68, 95% CI 0.64,0.72, $p<0.01$).

DISCUSSION

National benchmarking initiatives such as the National Surgical Quality Improvement Program (NSQIP) have highlighted the need for accurate risk stratification in both the surgical and trauma patient.¹¹ Historically, the usefulness of severity scores has been in their ability to stratify patients for research and prognostic applications, but they now have implications for stratifying outcomes for both reimbursement and credentialing purposes. In short, the stakes for accurate stratification have been raised. A possible implication for the increased precision needed in risk stratification is the very real likelihood that there is not “one size fits all” score, and this concept may be particularly true for the trauma population. But even within the trauma population, specific subgroups may exist which require specific severity scoring for accurate

risk stratification. Our objective was to evaluate the ability of commonly used trauma scores, along with a more general ICU scoring system (APACHE II), to accurately predict death in critically injured patients requiring extensive care in an ICU.

In this cohort of critically injured patients requiring greater than 48 hours of ICU care, APACHE II was the superior score in predicting mortality. This difference undoubtedly lies in the greater incorporation of physiologic and biochemical data into APACHE II. While TRISS has improved predictive accuracy over the strictly anatomic ISS, it still falls short of the predictive accuracy of APACHE II in this cohort. Others have demonstrated similar findings in this subset of trauma patients. In 691 helicopter-transported injured patients, APACHE II was a good predictor of mortality in acutely injured patients with an AUROC larger than that of ISS or the Trauma Score (TS)²³. This data contradicts previous reports that declared APACHE II to be invalid in the trauma patient¹⁸. McKenea *et al.* have previously reported a very low correlation between APACHE II and length of stay, and also noted that APACHE II did not correlate well with either ISS or TRISS. While these results were interpreted as an invalidation of APACHE II in trauma patients, it is important to note that the primary purpose of this study was to evaluate whether or not APACHE II could accurately predict resource allocation, not clinical outcomes. In fact, the ability of APACHE II to predict clinical outcomes such as death was not reported. It is not surprising that APACHE II does not correlate well with hospital length of stay in a linear model since patients with very low APACHE II (early discharge) and very high APACHE II (early death) will have low hospital LOS, while patients with intermediate APACHE II scores will have longer hospital LOS. The poor correlation between all of the severity scores and length of stay in this study confirms this relationship.

Evaluation of the multiple score sub-components reveals important information regarding the physiologic and biochemical parameters that are useful for severity scoring in this population. Traumatic brain injury (TBI) is clearly an important contributor to death in the trauma patient, and the subcomponent scores which account for head injury (GCS and AIS head or neck) were both independently associated with mortality. The bimodal nature seen in the relationship between GCS and AIS head and subsequent mortality is likely due to the requirement for 48 hours of ICU care for study inclusion. Multiple-trauma patients with GCS >10 scores who required ICU admission for at least 48 hours are likely to have severe chest, abdomen or extremity injury with associated exsanguination. Patients with less severe chest or abdominal trauma in combination with a lower GCS representing even a mild head injury will also require greater than 48 hours of ICU care. In comparing these patient populations, both requiring greater than 48 hours in the ICU, multiple bodily injuries in the absence of TBI may confer a higher risk of death than a mild brain injury and less severe bodily injury accounting for the second peak in the relationship between GCS and AIS head and mortality. It is expected that patients with a severe TBI (GCS<5) will require admission to the ICU regardless of other injuries and will have a higher likelihood of subsequent mortality.

Interestingly the lowest hemoglobin within the first 24 hours was not associated with mortality, but this could be explained by the aggressive use of blood product transfusions. Variables which account for blood product transfusions may improve the accuracy of severity scoring in trauma. Unlike medical ICU patients, trauma patients infrequently present with derangements in common blood chemistries (serum sodium, potassium and creatinine). However, when these derangements are present, they are associated with mortality. Other physiologic parameters that are used as predictors of mortality in trauma patients (lactate, blood glucose, mixed venous oxygen saturation) are notably missing from general ICU scoring systems such as APACHE II.

The strengths of this study include its prospective nature, large sample size, and evaluation of several severity scores and their subcomponents. To our knowledge this represents the largest analysis of scoring systems in critically injured patients using data collected prospectively by dedicated research personnel. Despite these strengths, there are several important limitations. One limitation is that patients who died or were discharged from the ICU prior to 48 hours were not a part of the original study cohort. It is unclear if patients who die within 48 hours or those who are not ill enough to require 48 hours of ICU care differ in a way from those in this cohort as to make our conclusions invalid. In these patients, ISS may be artificially low if not all injuries are discovered prior to death. Regardless of this uncertainty in the ability to generalize these data we believe they are important for two reasons:

1. The difficulty of outcome prediction—and the area where severity scores have the most value—is within the range of patients between those who are neither so severely injured as to die within 48 hours or only admitted to the ICU for short observational stays. For this reason, it is how various scores perform for these “in-between” patients that are of interest. We believe patients remaining in the ICU for at least 48 hours accurately represent this population.
2. Besides outcome prediction, severity scores are used for risk stratification during research. Perhaps, this is their most important utility. As with most research protocols, our original study did not include “all-comers”. Most ICU based research protocols restrict study enrollment to some minimal illness severity and exclude patients who are predicted to die within a short period of time. In other words, this cohort is *exactly* the population to whom these scores are most commonly applied. These data would suggest that studies of critically ill trauma patients should use APACHE II as the most accurate assessment of risk. This suggestion is irrespective of the ability of ISS or TRISS to accurately predict the outcome of “all-comers.”

A second limitation is the variability in the frequency of data collection within our patient population. More frequent parameter checks would lead to more accurate APACHE II scores. Additionally, TRISS and ISS by the definition are determined based only on data collection at admission and do not reflect a patient’s condition over 24 hours.

Using APACHE II as a tool to assess quality of care or physician performance has some noted limitations. By definition, the APACHE II score includes the most deranged physiologic values from the first 24 hours of hospital admission. Physicians or ICUs providing poor care could be inappropriately rewarded if a patient’s condition worsened severely in the first 24 hours leading to a higher APACHE II score. This is a valid grievance against the use of APACHE II as a national quality evaluation tool, but the complaint assumes that poor medical care would be the only cause of a decline in the patient’s health in the first 24 hours. Underlying medical conditions as well as demographic factors can lead to a delayed physiologic response to an anatomic insult. In these circumstances, APACHE II would be a more accurate representation of the patient’s clinical status than either ISS (anatomic only) or TRISS (physiologic variables collected at admission). The differences in TRISS and APACHE II could be mostly explained by this difference in timing of collection, but regardless, the evolving physiologic response to injury in the first 24 hours is noted to be an important determinant of outcome. Therefore, this study supports APACHE II as an example of a physiologic severity score with applicability to critically ill trauma patients, but not as the ideal score for a national benchmarking tool.

In conclusion, this study adds to the available literature regarding severity scoring among the critically injured patient. It demonstrates that general ICU scoring systems with a strong physiologic basis such as APACHE II are beneficial in trauma patients requiring greater than 48 hours of intensive care; in fact, they may be better predictors of clinical outcomes than traditional trauma scores. As benchmarking initiatives become more prevalent, the need for

accurate risk stratification will become more important. These initiatives should not focus on “one size fits all” scores, but instead, search for population and sub-populations specific scores that allow for the most accurate risk stratification. Further research should focus on the development of a scoring system specifically for critically injured patients which would ideally be easy to use and widely adopted.

Acknowledgments

This work was supported by a National Institutes of Health Grant—RO1 AI49989-01 (ClinicalTrials.gov identifier NCT00170560) and an Agency for Healthcare Research and Quality grant—T32 HS 013833. Portions of this data will be presented in poster form at the 2008 Eastern Association for the Surgery of Trauma meeting Amelia Island, FL. The authors have no conflicts of interest to disclose.

REFERENCES

1. Antonelli M, Moreno R, Vincent JL. Application of SOFA score to trauma patients. Sequential Organ Failure Assessment. *Intensive Care Med* 1999;25:389–394. [PubMed: 10342513]
2. Aslar AK, Kuzu MA, Elhan AH, Tanik A, Hengirmen S. Admission lactate level and the APACHE II score are the most useful predictors of prognosis following torso trauma. *Injury* 2004;35:746–752. [PubMed: 15246796]
3. Baker SP, O'Neill B, Haddon W Jr, Long WB. The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. *J Trauma* 1974;14:187–196. [PubMed: 4814394]
4. Boyd CR, Tolson MA, Copes WS. Evaluating trauma care: the TRISS method. Trauma Score and the Injury Severity Score. *J Trauma* 1987;27:370–378. [PubMed: 3106646]
5. Champion HR, Sacco WJ, Carnazzo AJ, Copes W, Fouty WJ. Trauma score. *Crit Care Med* 1981;9:672–676. [PubMed: 7273818]
6. Champion HR, Sacco WJ, Copes WS, Gann DS, Gennarelli TA, Flanagan ME. A revision of the Trauma Score. *J Trauma* 1989;29:623–629. [PubMed: 2657085]
7. Chawda MN, Hildebrand F, Pape HC, Giannoudis PV. Predicting outcome after multiple trauma: which scoring system? *Injury* 2004;35:347–358. [PubMed: 15037369]
8. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: A Nonparametric approach. *Biometrics* 1988;44:837–845. [PubMed: 3203132]Ref Type: Journal (Full)
9. Guzzo JL, Bochicchio GV, Napolitano LM, Malone DL, Meyer W, Scalea TM. Prediction of outcomes in trauma: anatomic or physiologic parameters? *J Am Coll Surg* 2005;201:891–897. [PubMed: 16310692]
10. Hannan EL, Farrell LS, Gorthy SF, et al. Predictors of mortality in adult patients with blunt injuries in New York State: a comparison of the Trauma and Injury Severity Score (TRISS) and the International Classification of Disease, Ninth Revision-based Injury Severity Score (ICISS). *J Trauma* 1999;47:8–14. [PubMed: 10421179]
11. Khuri SF, Daley J, Henderson W, et al. The Department of Veterans Affairs' NSQIP: the first national, validated, outcome-based, risk-adjusted, and peer-controlled program for the measurement and enhancement of the quality of surgical care. National VA Surgical Quality Improvement Program. *Ann Surg* 1998;228:491–507. [PubMed: 9790339]
12. Khuri SF, Daley J, Henderson W, et al. The Department of Veterans Affairs' NSQIP: the first national, validated, outcome-based, risk-adjusted, and peer-controlled program for the measurement and enhancement of the quality of surgical care. National VA Surgical Quality Improvement Program. *Ann Surg* 1998;228:491–507. [PubMed: 9790339]
13. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med* 1985;13:818–829. [PubMed: 3928249]
14. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med* 1985;13:818–829. [PubMed: 3928249]

15. Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. APACHE-acute physiology and chronic health evaluation: a physiologically based classification system. *Crit Care Med* 1981;9:591–597. [PubMed: 7261642]
16. Kuhls DA, Malone DL, McCarter RJ, Napolitano LM. Predictors of mortality in adult trauma patients: the physiologic trauma score is equivalent to the Trauma and Injury Severity Score. *J Am Coll Surg* 2002;194:695–704. [PubMed: 12081059]
17. Lee JY, Vaccaro AR, Lim MR, et al. Thoracolumbar injury classification and severity score: a new paradigm for the treatment of thoracolumbar spine trauma. *J Orthop Sci* 2005;10:671–675. [PubMed: 16307197]
18. McAnena OJ, Moore FA, Moore EE, Mattox KL, Marx JA, Pepe P. Invalidation of the APACHE II scoring system for patients with acute trauma. *J Trauma* 1992;33:504–506. [PubMed: 1433394]
19. McAnena OJ, Moore FA, Moore EE, Mattox KL, Marx JA, Pepe P. Invalidation of the APACHE II scoring system for patients with acute trauma. *J Trauma* 1992;33:504–506. [PubMed: 1433394]
20. Moore L, Lavoie A, Abdous B, et al. Unification of the revised trauma score. *J Trauma* 2006;61:718–722. [PubMed: 16967013]
21. Napolitano LM, Ferrer T, McCarter RJ Jr, Scalea TM. Systemic inflammatory response syndrome score at admission independently predicts mortality and length of stay in trauma patients. *J Trauma* 2000;49:647–652. [PubMed: 11038081]
22. Reiter A, Mauritz W, Jordan B, et al. Improving risk adjustment in critically ill trauma patients: the TRISS-SAPS Score. *J Trauma* 2004;57:375–380. [PubMed: 15345988]
23. Rhee KJ, Baxt WG, Mackenzie JR, et al. APACHE II scoring in the injured patient. *Crit Care Med* 1990;18:827–830. [PubMed: 2379396]
24. Ruchholtz S, Pehle B, Lewan U, et al. The emergency room transfusion score (ETS): prediction of blood transfusion requirement in initial resuscitation after severe trauma. *Transfus Med* 2006;16:49–56. [PubMed: 16480439]
25. Rutledge R, Fakhry S, Rutherford E, Muakkassa F, Meyer A. Comparison of APACHE II, Trauma Score, and Injury Severity Score as predictors of outcome in critically injured trauma patients. *Am J Surg* 1993;166:244–247. [PubMed: 8368434]
26. Rutledge R, Osler T. The ICD-9-based illness severity score: a new model that outperforms both DRG and APR-DRG as predictors of survival and resource utilization. *J Trauma* 1998;45:791–799. [PubMed: 9783623]
27. Yucel N, Lefering R, Maegele M, et al. Trauma Associated Severe Hemorrhage (TASH)-Score: probability of mass transfusion as surrogate for life threatening hemorrhage after multiple trauma. *J Trauma* 2006;60:1228–1236. [PubMed: 16766965]

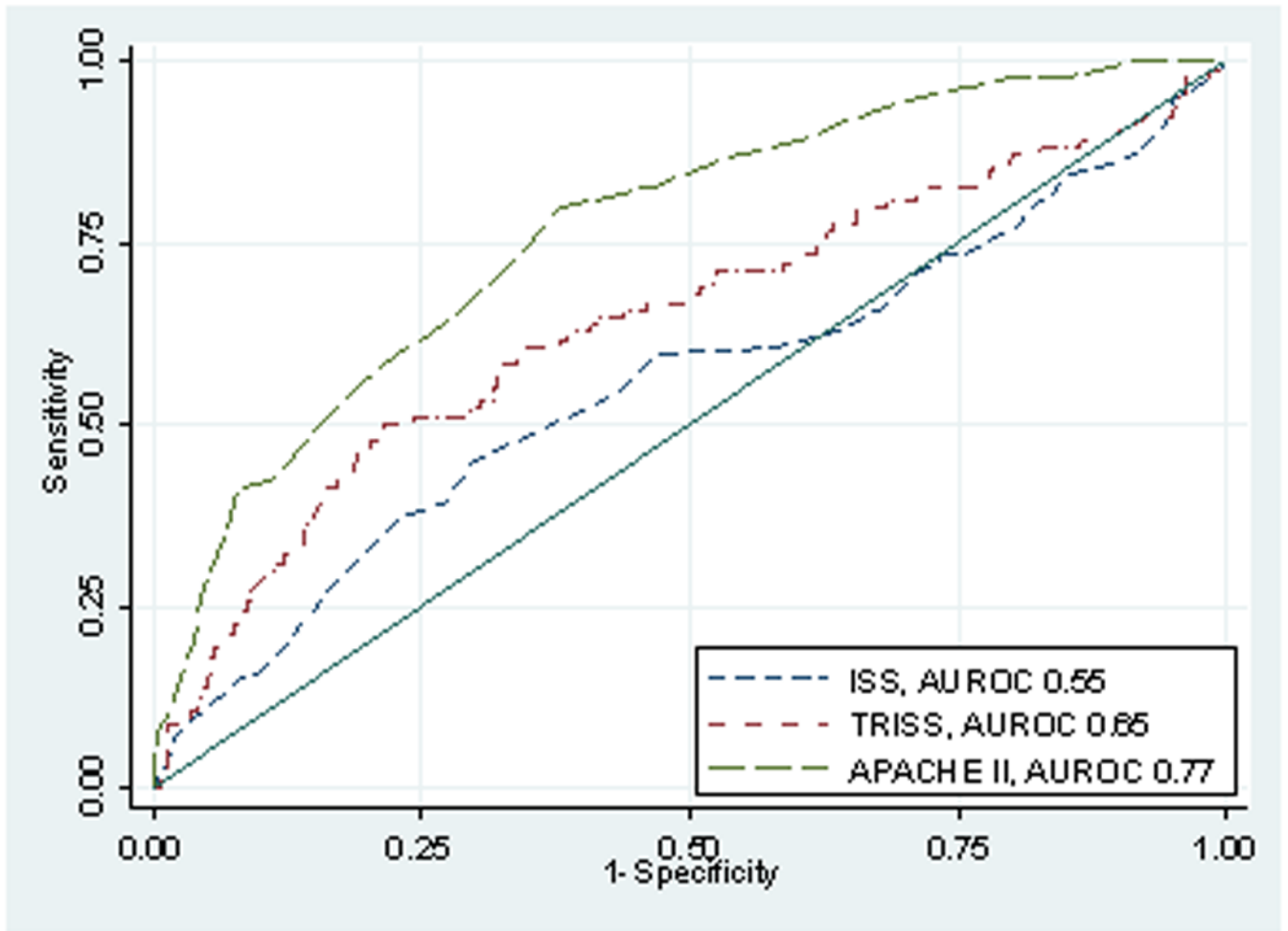


Figure 1. Receiver Operating Characteristic Curves for ISS, TRISS and APACHE II for predicting in-hospital mortality in a subset of critically ill trauma patients.

Table 1
Demographic and Clinical Characteristics By Outcome.

Variable	Survivors (n=881)	Non-Survivors (n=138)	P Value
Age (years) [^]	41±17	51±20	<0.001
Males n (%) ⁺⁺	27.6	19.6	0.05
Penetrating Trauma (%)	13.4 (11.1 – 15.6)	7.8 (3.5 – 12.5)	0.09
Hospital LOS (days)	17 (12 – 25)	8 (6 – 12)	<0.001
ICU LOS (days)	9 (6 – 14)	7 (5 – 11)	0.001
APACHE II [^]	16.3 ± 5.5	22.3 ± 6.1	<0.001
ISS [^]	31.1 ± 11.6	33.2 ± 13.9	0.43
TRISS [*]	0.85 (0.53 – 0.95)	0.48 (0.08 – 0.92)	<0.001
Temperature rectal (°C)	36.3 (35.2 – 38.5)	35.9 (34.8 – 38.3)	0.02
Mean Arterial Pressure (mmHg)	66 (59 – 112)	67 (59 – 120)	0.48
Heart Rate (ventricular response, bpm)	110 ± 32	109 ± 31	0.87
Respiratory Rate (bpm)	21 ± 9	22 ± 9	0.78
Oxygenation			
PaO2 (mmHg)	93 (74 – 133)	93 (70 – 140)	0.41
Arterial pH	7.27 ± 0.11	7.27 ± 0.13	0.88
Serum Sodium (mmol/L)	140 ± 6	140 ± 8	0.15
Serum Potassium (mmol/L)	3.9 ± 0.8	3.9 ± 1.0	0.91
Serum Creatinine (mg/dl)	0.8 (0.6 – 1.0)	0.8 (0.6 – 1.2)	0.03
Hematocrit (%)	28 ± 7	28 ± 8	0.78
WBC (10 ⁹ /L)	16.6 ± 8.3	16.1 ± 7.5	0.43
GCS	8 (5 – 10)	5 (3–8)	<0.001
Total APS Score	14.9 ± 5.2	19.6 ± 6.6	<0.001

[^] mean ± standard deviation; compared using two sample *t*-test

^{*} median (interquartile range); compared using Wilcoxon Rank Sum test

⁺⁺ χ^2 test corrected for repeated measure

[†] The body mass index is the weight in kilograms divided by the square of the height in meters. APACHE (Acute Physiology and Chronic Health Evaluation) II; ISS (Injury Severity Score); TRISS (Trauma Revised Injury Severity Score); GCS (Glasgow Coma Scale); CVP (Central Venous Pressure); AIS (Abbreviated Injury Score)

Table 2
Area Under the Receiver Operating Characteristic Curve for Various Scoring Variables.

Variable	AUROC (95% Confidence Interval)
Age (years)	0.64 (0.59 – 0.69)
APACHE II	0.77 (0.72 – 0.81)
Temperature rectal (°C)	0.57 (0.51 – 0.62)
Mean Arterial Pressure (mmHg)	0.53 (0.48 – 0.59)
Heart Rate (ventricular response, bpm)	0.52 (0.48 – 0.56)
Respiratory Rate (bpm)	0.53 (0.48 – 0.59)
Oxygenation	
PaO2 (mmHg)	0.53 (0.47 – 0.58)
Arterial pH	0.52 (0.46 – 0.57)
Serum Sodium (mmol/L)	0.53 (0.48 – 0.60)
Serum Potassium (mmol/L)	0.52 (0.46 – 0.57)
Serum Creatinine (mg/dl)	0.56 (0.50 – 0.61)
Hematocrit (%)	0.51 (0.46 – 0.56)
WBC (10 ⁹ /L)	0.50 (0.46 – 0.56)
GCS	0.64 (0.59 – 0.69)
Total APS Score	0.70 (0.66 – 0.75)
ISS	0.54 (0.48 – 0.60)
AIS Head or Neck	0.61 (0.54 – 0.67)
AIS Face	0.52 (0.45 – 0.58)
AIS Chest	0.50 (0.45 – 0.56)
AIS Abdomen	0.52 (0.45 – 0.59)
AIS Extremity	0.50 (0.44 – 0.56)
AIS External	0.53 (0.47 – 0.60)
TRISS	0.64 (0.58 – 0.71)

APACHE (Acute Physiology and Chronic Health Evaluation) II; ISS (Injury Severity Score); TRISS (Trauma Revised Injury Severity Score); GCS (Glasgow Coma Scale); CVP (Central Venous Pressure); AIS (Abbreviated Injury Score)

Table 3

Comparison of Predictive Ability of Severity Scores by Mechanism of Injury.

Variable	Blunt Trauma (n=889)	Penetrating Trauma (n=129)
APAHCE II	0.76 (0.71 – 0.80)	0.81 (0.65 – 0.97)
ISS	0.53 (0.47 – 0.60)	0.47 (0.22 – 0.71)
TRISS	0.65 (0.58 – 0.72)	0.60 (0.36 – 0.83)

APACHE (Acute Physiology and Chronic Health Evaluation) II; ISS (Injury Severity Score); TRISS (Trauma Revised Injury Severity Score); GCS (Glasgow Coma Scale)