



Published in final edited form as:

J Trauma. 2006 April ; 60(4): 792–801. doi:10.1097/01.ta.0000214589.02515.dd.

Predicting Hospital Mortality Among Injured Children Using a National Trauma Database

Randall S. Burd, MD, PhD,

Department of Surgery, Division of Pediatric Surgery, Columbia, Missouri.

Tai S. Jang, MS, and

UMDNJ-Robert Wood Johnson Medical School, New Brunswick, New Jersey, Columbia, Missouri,
Department of Mechanical and Aerospace Engineering, University of Missouri-Columbia,
Columbia, Missouri.

Satish S. Nair, PhD

UMDNJ-Robert Wood Johnson Medical School, New Brunswick, New Jersey, Columbia, Missouri,
Department of Mechanical and Aerospace Engineering, University of Missouri-Columbia,
Columbia, Missouri.

Abstract

Purpose: The purpose of this study was to develop a model that accurately predicts mortality among injured children based on components of the initial patient evaluation and that is generalizable to diverse acute care settings. Important predictive variables obtained in an emergency setting are frequently missing in even large national databases, limiting their effectiveness for developing predictions. In this study, a model predicting pediatric trauma mortality was developed using a national database and methods to handle missing data that may avoid biases that can occur restricting analyses to complete cases.

Methods: Records of pediatric patients included in the National Pediatric Trauma Registry (NPTR) between 1996 and 1999 were used as a training set in a logistic regression model to predict hospital mortality using vital signs, Glasgow Coma Scale (GCS) score, and intubation status. Multiple imputation was applied to handle missing data. The model was tested using independent data from the NPTR and National Trauma Data Bank (NTDB).

Results: Complete case analysis identified only GCS-eye and intubation status as predictors of mortality. A model based on complete case analysis had good discrimination (c-index = 0.784) and excellent calibration (Hosmer-Lemeshow c-statistic, 6.8) ($p > 0.05$). Using multiple imputation, three additional predictors of mortality (systolic blood pressure, pulse, and GCS-motor) were identified and improved model performance was observed. The model developed using multiple imputation had excellent discrimination (c-index, 0.947–0.973) in both test datasets. Calibration was better in the NPTR testing set than in the NTDB (Hosmer-Lemeshow c-statistic, 9.2 for NPTR [$p > 0.05$] and 258.2 for NTDB [$p < 0.05$]). At a probability cutoff that minimized misclassification in the training set, the false-negative and false-negative rates of the model were

better than those obtained with either the Revised Trauma Score (RTS) or Pediatric Trauma Score using data from the NPTR testing set. Although the false-positive rates were lower with the RTS using data from the NTDB, the false-negative rates of the proposed model and the RTS were similar in this test dataset.

Conclusions: Using multiple imputation to handle missing data, a model predicting pediatric trauma mortality was developed that compared favorably with existing trauma scores. Application of these methods may produce predictive trauma models that are more statistically reliable and applicable in clinical practice.

Keywords

Child; Hospital mortality; Injury; Models; Statistical; Wounds

Trauma is an important threat to the health of children with injuries, resulting in more deaths in children than all other causes combined.¹ Because injured children have better outcomes at centers with specialized trauma care, appropriate transport to these centers may reduce morbidity and mortality.^{2,3} Early identification of children with severe injuries facilitates transport to designated trauma centers where specialized care can be given. Appropriate triage can also predict the level of initial hospital manpower and other resources needed and determine the need for transfer to a specialized trauma center after arrival to a hospital with limited trauma expertise. Because only 5% of injured children need the resources of a trauma center, a rapid method to classify the severity of injury and to get the patients to the appropriate facility is needed.⁴ Nondesignated trauma centers can treat children with less severe injuries, whereas trauma centers are needed to treat more severely injured children. Undertriage puts children at risk of being treated at a hospital not equipped to deliver specialized trauma care and rehabilitation. Overtriage to trauma centers dilutes the resources needed for caring for severely injured children.

There remains a need for an effective pediatric-specific trauma triage instrument. Features of the optimal triage tool include accuracy, reproducibility, flexibility, and ease of application at the location of injury. A triage tool that is accurate will avoid misclassification and have acceptable overtriage and undertriage rates. A reproducible tool will be accurate when applied in different environments (e.g., in different regions of the country or in a rural versus urban setting) or by different raters. A tool that can be used in only one institution will have limited use and not gain widespread acceptance. A tool that is flexible will be able to easily evolve as the care for injured children improves and the categories of triage change. Because a triage tool needs to be used in the prehospital setting, it should be based on a simple set of variables. On-scene caregivers, who are treating trauma patients and need to rapidly transport critically ill children, and physicians in the emergency room cannot be expected to use a complex triage system.

Several trauma scoring systems have been used to triage injured children. The Revised Trauma Score (RTS) is a physiologic score based on systolic blood pressure (SBP), respiratory rate (RR), and Glasgow Coma Scale (GCS) scores.⁵ Although derived using adult trauma data, the RTS has been validated as a potentially useful triage tool in children.⁶ The Pediatric Trauma Score (PTS) was developed as a pediatric-specific alternative to the

RTS and combines physiologic and anatomic variables including weight, airway status, SBP, central nervous system status, presence of an open wound, and presence of fractures.^{7,8} Although PTS has been shown to correlate with injury severity in children, this score includes subjective variables that may not be easy to obtain in an acute care setting. A significant advantage of the PTS over the RTS has not been shown.^{8,9} The Age-Specific Pediatric Trauma Score (ASPTS) has been recently created from a state trauma database using logistic regression methods and incorporates the use of age-adjusted vital signs and GCS scores to predict injury severity and probability of mortality.¹⁰ Because the specificity of the ASPTS is better than the RTS, the ASPTS is a promising alternative to the RTS and deserves further study as a pediatric trauma triage tool in a separate dataset or in a clinical setting.

The purpose of this study was to develop a model based on components of the initial patient evaluation that accurately predicts hospital mortality among injured children. Although triage decisions can be based on other measures of severity of illness, the model was evaluated as a triage tool using mortality as the principal measure of outcome. Based on triage targets established by a consensus panel of the Florida Trauma Triage Study, the goal of this study was to develop a tool that achieves an overtriage rate of <30% and an undertriage rate of <5%.¹¹ As with previous trauma scores, initial vital signs and GCS scores were considered as potential predictors. In addition to these physiologic parameters, intubation status was also evaluated as a predictor because this feature is easy to identify in an acute care setting and conveys an increased likelihood of severe injury. The model was developed and tested using data obtained from two separate national datasets and compared with currently available triage scores. Because variables obtained in an emergency setting are frequently missing in trauma databases, methods to handle missing data were used to avoid bias in the final model.¹²

METHODS

Data Sources and Subject Selection

This study has been approved by the Institutional Review Board at UMDNJ-Robert Wood Johnson Medical School. The NPTR is a database, started in 1985, that tracks the management and outcome of injured children treated at participating pediatric trauma centers or pediatric hospitals. Data (n = 35,385) from the registry obtained between 1996 and 1999 were used as the training set and additional data obtained between 1999 and 2001 (n = 15,818) were used as the first testing set. Additional testing of the model was performed using data obtained from the NTDB, a database started in 1989 that contains over 730,000 cases from 268 adult and pediatric trauma centers. Because some trauma centers have may have reported data to both the NPTR and NTDB, only records in the NTDB from 2002 and 2003 (n = 16,868) were used to prevent overlap with data from the NPTR. Records were excluded from analysis in the training and testing sets if age was >17 years or was not recorded, vital signs were outside of normal physiologic range (SBP > 200, p > 250, RR > 80), or hospital mortality was not reported.

Development of Prognostic Models

Input variables selected for modeling included the following clinical features recorded in the emergency room: SBP, pulse, RR, components of the GCS score (GCS; GCS-eye, GCS-motor, and GCS-verbal), mechanism of injury (penetrating or blunt), and intubation status (not intubated or intubated). The outcome variable studied was hospital mortality. Pulse is not available in the NTDB and was handled as missing and was imputed as described below.

An analysis was performed to assess the pattern of missing data and evaluate how missing data might impact modeling. Missing data were observed for all input variables except mechanism of injury, with the proportion of missing data varying between 1% and 29% of cases (Table 1). The relationship between the absence of each variable and other covariates was evaluated using univariate logistic regression, and between data absence and mortality using χ^2 analysis.

Modeling only using cases with complete data in all fields (complete case analysis) requires that the cases represent a random sample of the entire dataset. This type of missing data pattern is called *missing completely at random* (MCAR). When data are MCAR, no relationship can be expected between data absence and outcome. A more general missing data pattern, *missing at random* (MAR), occurs when the probability that a data point is missing may depend on the values of other variables that were measured but not on the true value of the data point that is missing.¹³ Most methods for imputing missing data, such as multiple imputation, require that the missing data pattern is MAR. Direct proof that a missing data pattern is MAR requires, however, that the missing data are known. Although a MAR data pattern must be assumed and cannot be proven, a relationship between absence of a given variable and the values of other covariates provides evidence of a potential MAR pattern. Because the missing data pattern was not MCAR (see Results, below), we assumed that the missing data pattern was MAR and applied multiple imputation to impute missing values.

An introduction to multiple imputation can be found in several excellent sources.^{14–16} With the increasing availability of software that can perform multiple imputation, this powerful method now has been applied to a range of clinical problems.^{16–19} Multiple imputation involves three phases. First, missing data are filled in n times to generate n complete datasets. Missing data are replaced in each of the n datasets with possible values that represent the uncertainty of the correct missing value rather than with a single value such as a median, mean, or mode. Each of these n imputed datasets are “complete” with measured and imputed values. The n imputed datasets are then analyzed using conventional methods ordinarily used for complete case analysis. Finally, the results of the statistical analyses on each imputed dataset are combined to give a final result.

Multiple imputation was used to construct 10 imputed datasets from the training data. Values were imputed using a regression that included age, SBP, pulse, RR, GCS-eye, GCS-motor, GCS-verbal, GCS-total, mechanism of injury (penetrating or blunt), and intubation status. Because the normal ranges of SBP, pulse, and RR differ by age, these variables were standardized to improve comparison among all subjects after imputation. The mean and

standard deviation of SBP at each age were obtained from the training set and used to standardize each recorded value of SBP in the imputed datasets using the equation:

$$\frac{\text{Recorded SBP} - \text{mean SBP for patient age}}{\text{STDEV of SBP for patient age}}$$

Pulse and RR were similarly standardized. Standardized vital signs, components of the GCS, and intubation status were then used as predictors in logistic regression analyses performed on each imputed dataset. The results of the analyses performed on each of the 10 imputed datasets were then combined to give the final result.

Model Validation

Because each testing dataset contained missing data, multiple imputation was performed to “complete” each testing dataset. Data from the training set and each testing set were merged. Multiple imputation ($n = 10$) was then performed on the merged datasets. A separate imputation was performed for each testing dataset. After imputation, vital signs in the testing datasets were standardized using mean and standard deviation values obtained from the training dataset. The logistic regression model developed from the training dataset was used to calculate an estimated probability of hospital mortality for each case in the imputed testing datasets. The mean of the probabilities for each case was calculated and compared with the observed mortality to assess the predictive capacity of the model.

Discrimination of the model was evaluated by calculating the area under the ROC curve or c-index. Calibration (goodness of fit) of the model was assessed using the Hosmer-Lemeshow (H-L) statistics. Cases were ordered based on the probability of mortality estimated by the model and grouped into deciles (c-statistic) or decines (h-statistic). The expected and observed number of outcomes in each partition were compared. An H-L statistic < 15.5 (8 degrees of freedom; $p 0.05$) shows that there is no significant difference between observed and predicted values and excellent goodness of fit, while an H-L statistic > 15.5 shows a significant difference between observed and predicted values and poor goodness of fit. Calibration was also graphically displayed by plotting predicted and observed mortality across all risk ranges. The slope of the calibration curve was calculated using linear regression. The R^2 value represents the proportion of variation of the dependent variable (observed mortality rate) that is predicted from the independent variable (predicted mortality rate). An R^2 value of 1.0 indicates that all points lie on a straight line and that the predicted mortality rate is able to predict the actual mortality rate with 100% certainty. A perfectly calibrated model will have a slope of 1 and a y-intercept at 0.

The model was also compared with RTS and PTS. A cutoff probability was determined in the training dataset that resulted in the least misclassification of cases (false-positive and false-negative rates assumed to be equivalent). The RTS was calculated in each testing dataset using imputed vital sign and GCS values. A value of $\text{RTS} < 12$ was used to designate a high potential for mortality.⁵ The PTS was calculated for most subjects in the NPTR testing set (93% of records) but was not available in the NTDB. A score of < 9 was considered to indicate a high potential for mortality.⁷ On the basis of these cutoff points, sensitivity, specificity, and predictive values were calculated for the individual scores.

Overtriage (false-positive) rate was defined as (1-specificity), whereas undertriage (false-negative) rate was defined as (1-sensitivity).

Statistical Software

SPSS 12.0 (SPSS Inc., Chicago, IL) was used to perform univariate logistic regression and to analyze continuous variables using the unpaired Student's *t* test, ordinal variables using the Mann-Whitney *U* test, and groups using χ^2 analyses. Multiple imputation and development of the final logistic regression model was performed using SAS 8.2 (PROC MI and PROC MIANALYZE, SAS Institute, Cary, NC).¹⁶

RESULTS

Overview of Data

The mortality in the NPTR training set was 2.9%, in the NPTR testing set was 2.6%, and in the NTDB dataset was 3.1%. Variables used for modeling in the three datasets are shown in Table 1. In the three datasets, standardized SBP and GCS were lower whereas standardized pulse was higher among subjects who died. Subjects who died were more commonly intubated than those who survived. Patients in the NPTR who died were younger than those who survived, whereas those in the NTDB who died were older than those who survived. Although no differences in standardized respiratory rate were observed in the NPTR, the standardized respiratory rate was lower among patients in the NTDB who died. The frequency of penetrating injuries among survivors and non-survivors was similar in the NPTR, but more penetrating injuries were observed among nonsurvivors in the NTDB (Table 1).

Missing variables were frequently observed among those who died. Although there were 1,008 deaths in the NPTR training set, only 17 deaths (0.07%) were observed among 23,689 subjects with complete data for the 7 variables used to model. A logistic regression model based on these data had good discrimination (ability to discriminate between patients who live and those who die; c-index = 0.784) and excellent calibration (accuracy of predicting the mortality rate; H-L c-statistic = 6.4). Lower GCS-eye and intubation status were the only significant predictors of mortality in the final model (Table 2). When evaluated in each testing dataset, discrimination of the model was excellent (c-index = 0.965 in NPTR testing set and 0.940 in NTDB), whereas calibration was not as good (H-L c-statistic = 242.6 in NPTR testing set and 398.1 in NTDB).

Development of an Improved Model

To develop a model with improved performance, the feasibility and effectiveness of multiple imputation was evaluated. As shown in Table 1, missing values were observed for all variables used for modeling except mechanism of injury and age (records were not included for analysis if age was not available). Values were most commonly missing for components of the GCS. Data patterns in which some or all components of GCS were missing were most frequent. A similar missing data pattern was observed in the NPTR and the NTDB (Table 3). In the NPTR training set, missing values of each variable were less common among survivors than among nonsurvivors ($p < 0.001$). Univariate logistic models show that the

absence of variables used to model mortality was potentially associated with other variables (Table 4). GCS-eye, GCS-motor, and intubation status were associated with the absence of each modeling variable, whereas pulse was associated with the absence of all but one variable.

Multiple imputation was performed to complete missing data fields in the NPTR training set. Ten imputations were performed to obtain values for 31,495 previously missing values among the seven variables used for modeling. Data from the original incomplete dataset were similar to those obtained by imputation (Table 5). The addition of age as a covariate was not associated with any improvement in either discrimination or calibration of the model. Compared with the model developed using complete case analysis, the model developed using multiple imputation included three additional significant variables (standardized SBP, standardized pulse, and GCS-motor; Table 6). The model obtained using multiple imputation had improved discrimination (c-index = 0.958 versus 0.784) but decreased calibration (H-L c-statistic, 25.7 versus 6.8). In the final model, each unit decrease in GCS-eye resulted in a 37% increase in risk of death, while each unit decrease in GCS-motor resulted in a 59% increase in risk of death. The risk of mortality was 19 times higher when the subject was intubated.

Assessment of Model Performance

The logistic regression model was tested using independent data from the NPTR and data from the NTDB. For each subject, the probability of mortality was calculated from each logistic regression equation using observed or imputed values of vitals signs, GCS components, and intubation status. The model obtained using multiple imputation had better performance than the model based on complete data when evaluated using NPTR testing set data, but had similar performance using NTDB data (Table 7). Subgroups with a blunt or penetrating mechanism of injury were separately analyzed. Excellent discrimination was observed in the test data (c-index, 0.947–0.973) regardless of mechanism of injury. Model calibration, however, was more variable, being generally better for the NPTR data than the NTDB dataset (Table 8). The calibration curves showed that the model over-predicted rather than under-predicted mortality in the NTDB dataset (Fig. 1). Using probability estimates in the training set, a cut-off probability of 0.009% resulted in the least misclassification (false-positive rate equal to false-negative rate). This cut-off value led to a similar misclassification in the NPTR testing set but higher misclassification in the NTDB (Table 9). The proposed model had a higher positive predictive value and accuracy than either the RTS or PTS using NPTR testing set data ($p < 0.001$), but performed worse than the RTS using NTDB data ($p < 0.001$; Table 9).

DISCUSSION

In the current study, we have developed and tested a model predicting pediatric trauma mortality using the NPTR and NTDB. While the size and case-mix differences of these databases are advantages for developing and testing a prognostic trauma model, both databases have a significant amount of missing data. This problem may be more apparent among variables obtained in a prehospital setting or in the emergency room because absence

variables in these settings may be related to the severity of injury.¹² Although smaller or regional databases may have less missing data, we felt that the advantages of these larger datasets justified their use and the application of methods for handling missing data.

In previous trauma studies, two methods have been used to handle missing data. The most common method has been to restrict the analysis to subjects for whom values of all modeling variables are complete.¹⁰ When the number of records with missing variables represents a small fraction of the total number of records, complete case analysis is an acceptable method for handling missing data, since exclusion of records with incomplete data may have only minimal impact on the relative contribution of predictors. When complete case analysis is applied to datasets with a significant amount of missing data, omission of incomplete data may lead to biased results when the remaining cases are not representative of the larger population that the data are in-tended to represent. Biases that may result include both inclusion of nonpredictive covariates or omission of predictive covariates. Because of these limitations, it is appropriate to assess the likelihood that data are MCAR before proceeding with complete case analysis. Other methods for handling missing data have included ad hoc methods such as inserting the mean, median, or mode values in missing data fields.²⁰ These methods are convenient but generally produce results with variances that are biased toward zero. Because of these biases, standard measures for uncertainty of potential predictors such as standard error and *p* values can be inaccurate because these will not convey the true uncertainty of the missing data. Inaccurate estimates of uncertainty may result in a biased assessment of the relative contribution of individual predictors.¹³ Multiple imputation is a method for handling missing data that avoids the biases associated with complete case analysis or single imputation. With this method, statistical inferences can be made based on information contained in available complete data fields while reflecting the uncertainty related to data absence. Although this method has been increasingly used in a wide range of clinical studies, this method has, to our knowledge, been applied in only one study using trauma data.¹²

We first developed a predictive model using only records with complete data. After exclusion of records with at least one missing variable, only 69% of records remained from the original dataset. An analysis of the missing data pattern showed that components of the GCS were usually missing together. For this reason, it is unlikely that a step-wise backward elimination method would lead to an effective increase in the sample size because of the previously observed importance of GCS as a predictor of trauma mortality. A backward elimination method was also avoided because of the tendency for this methodology to lead to data overfitting. There was strong evidence that an analysis restricted to complete cases would lead to a less precise model. Since the mortality in the complete data substantially differed from that observed in the entire training dataset (0.07% versus 3%), data in the training set was unlikely to be MCAR. The observation that univariate analysis showed a strong correlation between the absence of potential predictors and mortality supported, but did not prove, that the data were not MCAR.

When the analysis was performed on the 23,695 cases with complete data, only GCS-eye and intubation status were identified as significant predictors of mortality. Using multiple imputation, we were able to increase the sample size by 45% (10,647 subjects) and increase

the number of subjects who died from 17 to 1,008. With the added statistical power, three additional variables were observed to be significant predictors of mortality (SBP, pulse, and GCS-motor), and the resulting model had improved higher discrimination. The inclusion of additional predictors suggested that information contained in the incomplete records was relevant to mortality prediction and was being incorporated into the final model. The application of multiple imputation provided insight into the relative contribution of individual predictors that was not achievable using complete case analysis alone.

Similar to previous triage tools, initial vital signs and components of the GCS were observed to be significant predictors of hospital mortality.⁵⁻¹⁰ In contrast to other studies that have found RR to be an important predictor of mortality, either alone or in combination with other variables, we did not observe RR to be an important predictor of hospital mortality. This difference may be attributable to our representation of RR as a monotonic variable rather than scoring RR for the predictive value of both tachypnea and diminished RR, as is done in the RTS and ASPTS. Similarly, components of the initial GCS were observed to be significant predictors of hospital mortality, most likely reflecting the importance of head injury as a cause of mortality among injured children. The relative contribution of GCS-motor to the final model is consistent with the previous finding that GCS-motor contains most of the predictive power of the GCS in relationship to mortality.²¹ It is interesting to note that GCS-motor was not identified as a significant covariate in the model based only on complete cases. Intubation status proved to be the covariate with the largest predictive capacity. Although anatomic variables that require interpretation have been described as a limitation of other trauma scores, intubation status is straightforward enough to determine even in an acute care setting and easily conveys severity of injury.

To evaluate performance of the proposed model, we first applied the methods described by Justice et al.²² These authors describe a scheme for evaluating models using the related concepts of *accuracy* and *generalizability*. Accuracy is assessed using conventional methods for evaluating discrimination and calibration. Generalizability is assessed by evaluating whether the accuracy of the model is both *reproducible* and *transportable*. Reproducibility is defined as the maintenance of accuracy in data obtained from the same source as the training set. Transportability is defined as the maintenance of accuracy in data obtained from a different but related population or collected using methods different from those used to obtain the training data.

Based on performance using the NPTR training set, the proposed model was found to be accurate since it satisfied conventional standards of discrimination and calibration. The model was also observed to be reproducible because it had adequate discrimination and calibration using unique data from the NPTR, the same dataset used for training. The transportability of the model was more limited. Using the NTDB testing data, discrimination of the model was excellent but calibration was not as good. An analysis of the calibration curves shows that application of the model in the NTDB resulted in overtriage because predicted mortality was generally higher than actual mortality. While the reproducibility of predictive models has been frequently performed, evaluation of transport-ability has frequently been omitted in previous studies.^{10,20} The current findings emphasize that

assessment of reproducibility alone may lead to an overoptimistic assessment of the predictive capacity of a model.

Several explanations may account for the more limited transportability of the model to the NTDB database. The NPTR and NTDB datasets were obtained from different types of trauma centers. Centers reporting to the NPTR were generally pediatric trauma centers, whereas hospitals reporting to the NTDB included both adult and pediatric trauma centers. These differences are important since the outcome of children treated at pediatric trauma centers may be better than those treated at adult trauma centers.^{2,3} In addition, the profile of children in each dataset differed; children in the NPTR were younger and less frequently sustained penetrating injuries than those in NTDB. Although we deliberately used a minimal set of predictors in the model, additional variables may need to be added to the proposed model to improve calibration across datasets obtained from different populations. Potential variables to include in future refinements of the model include the mechanism or intent of injury, transport time, and prehospital treatments.

A second measure of model performance relates to its clinical applicability. Although a predictive model may be statistically valid, it may not have clinical value.²³ One way that we assessed the clinical value of our model was to evaluate it against predetermined standards of overtriage and undertriage. The model achieved our prestudy goal of achieving an undertriage rate of <5% using both testing datasets, comparing favorably with the RTS and PTS. In contrast, the model achieved the prestudy goal of an overtriage rate of <30% only using data from the NPTR. An advantage of the model over the RTS and PTS was observed using NPTR data but not using NTDB data. Overtriage and undertriage are both critical factors to measure when evaluating a triage tool but do not have equivalent importance. Misclassifying patients who have high risk for mortality (undertriage) may have an impact on outcome, whereas misclassifying those who have low risk for mortality (overtriage) may lead to increased resource utilization. Emphasizing undertriage as a measure of clinical value, the proposed model performed equivalently to both the RTS and PTS.

There are limitations of the proposed model that will serve as the basis for future modification and evaluation. One limitation of the proposed model is that it predicts mortality alone. Although mortality is a simple criterion upon which to base triage decisions, it is a complex variable. The implications for resource utilization and trauma center readiness can be different when a death occurs in the emergency room versus when it occurs at the end of an extended hospitalization. For this reason, predicting the time of trauma death may be as important as predicting death.

In contrast to previous trauma triage tools, the proposed model does not rely on a simple score, but depends on a series of calculations that includes standardization of vital signs for age. Although decision tools that use a score to aid clinical decision making have been successful in other settings, the poor usability of even simple triage scores in a prehospital and acute care settings has been recognized.²⁴ The age-dependency of vital signs necessarily requires a complex method for handling these variables in pediatric trauma patients. For example, the ASPTS uses scoring based on age-related vital sign ranges.¹⁰ As medical care moves toward electronic acquisition and storage of data, the use of predictive models that

require more than simple calculations will become more practical. The problem of calculation and immediate interpretation by the user will be replaced by the challenge of providing devices that have adequate user interface and feedback.

The proposed model has additional limitations that are also found in other triage tools. The GCS is an important component of the model that can be difficult to assess in an acute care setting. Because of the importance of head trauma as a cause of pediatric trauma mortality, this variable is common to other trauma scores that have been evaluated in children, including the RTS, PTS, and ASPTS.⁵⁻¹⁰ Components of the GCS, particularly GCS-verbal, can be difficult to assess in critically ill or intubated patients.²⁵ Age-related adjustments are also needed when the GCS is applied to younger children, making application of GCS to these patients more difficult. The frequent absence of components of the GCS reflects the difficulties related to assessing this variable. Previous studies have handled this problem by either omitting records with incomplete GCS or using the lowest possible score when components of the GCS (particularly GCS-verbal in intubated patients) cannot be accurately accessed.²⁶ Application of regression methods such as multiple imputation to predict missing GCS components is a preferred solution since this will prevent loss of valuable data and avoid biases that may arise from ad hoc methods of data completion.

The proposed model is also intolerant of missing data. Similar to other trauma triage tools, every variable is required before a prediction can be obtained. In prehospital or other acute care settings, even simple variables such as vital signs may not be easy to immediately obtain because of the need for rapid stabilization and transport. Estimation of missing values using methods such as multiple imputation or Bayesian methods may become practical even in a prehospital setting as computer technologies are incorporated into these locations. We envision that a computer-based prediction model similar to that proposed here will be incorporated directly into the electronic medical record of each injured patient. When needed predictors are missing or have not yet been obtained, available data from an individual patient can be added to a larger existing dataset and multiple imputation performed to “complete” missing data fields for that patient. The now “complete” patient data can then be used to estimate the probability of mortality using a logistic regression equation similar to that described.

Early and accurate triage will likely reduce the morbidity and mortality of injured children.^{2,3} Although it may have higher accuracy than existing triage tools, the proposed model will require further validation that necessarily includes prospective testing. The current study highlights the challenges of developing a triage tool and proposes a preliminary model for usage in clinical settings. The goal of our work in this area is to develop a triage tool that meets acceptable standards of undertriage and overtriage and can easily be implemented in real world settings.

ACKNOWLEDGMENTS

We thank John E. Kolassa, PhD, Department of Statistics, Rutgers University, Piscataway, New Jersey, for his advice on the statistical analyses used in this paper. We also thank John R. Clarke, MD, for reviewing the manuscript.

Presented as an oral paper to the Society for Medical Decision Making, October 18–22, 2003, Chicago, Illinois.

Supported by grant R03/HD042561 from the National Institute of Child Health and Human Development, National Institutes of Health.

REFERENCES

1. Gotschall CS. Epidemiology of childhood injury. In: Eichelburger MR, ed. *Pediatric Trauma: Prevention, Acute Care* St Louis: Mosby Year Book;1993:16–19.
2. Potoka DA, Schall LC, Gardner MJ, Stafford PW, Peitzman AB, Ford HR. Impact of pediatric trauma centers on mortality in a statewide system. *J Trauma Injury Infect Crit Care* 2000;49:237–245.
3. Potoka DA, Schall LC, Ford HR. Improved functional outcome for severely injured children treated at pediatric trauma centers. *J Trauma Injury Infect Crit Care* 2001;51:824–832.
4. Champion HR. Field triage of trauma patients. *Ann Emerg Med* 1982;11:160–161. [PubMed: 7065493]
5. Champion HR, Sacco WJ, Copes WS, Gann DS, Gennarelli TA, Flanagan ME. A revision of the trauma score. *J Trauma Injury Infect Crit Care* 1989;29:623–629.
6. Eichelberger MR, Gotschall CS, Sacco WJ, et al. A comparison of the trauma score, the revised trauma score, and the pediatric trauma score. *Ann Emerg Med* 1989;18:1053–1058. [PubMed: 2802280]
7. Tepas JJ, 3rd, Mollitt DL, Talbert JL, et al. The pediatric trauma score as a predictor of injury severity in the injured child. *J Pediatr Surg* 1987;22:14–18. [PubMed: 3102714]
8. Tepas JJ, 3rd, Ramenofsky ML, Mollitt DL, et al. The pediatric trauma score as a predictor of injury severity: an objective assessment. *J Trauma Injury Infect Crit Care* 1988;28:425–429.
9. Kaufmann CR, Maier RV, Rivara FP, et al. Evaluation of the pediatric trauma score. *JAMA* 1990;263:69–72. [PubMed: 2293691]
10. Potoka DA, Schall LC, Ford HR. Development of a novel age-specific pediatric trauma score. *J Pediatr Surg* 2001;36:106–112. [PubMed: 11150447]
11. Phillips S, Rond PC, 3rd, Kelly SM, et al. The need for pediatric-specific triage criteria: results from the Florida Trauma Triage Study. *Pediatr Emerg Care* 1996;12:394–399. [PubMed: 8989783]
12. Joseph L, Belisle P, Tamim H, et al. Selection bias found in interpreting analyses with missing data for the prehospital index for trauma. *J Clin Epidemiol* 2004;57:147–53. [PubMed: 15125624]
13. Schafer JL. Multiple imputation: a primer. *Stat Methods Med Res* 1999;8:3–15. [PubMed: 10347857]
14. Schafer JL. *Analysis of incomplete multivariate data* London: Chapman and Hall; 1997.
15. Little R, Rubin D. *Statistical analysis with missing data* 2nd ed. Hoboken, NJ: Wiley-Interscience; 2002.
16. SAS OnlineDoc, Version 8, SAS Institute Inc, Cary, NC: SAS Institute Inc., 2000.
17. Stadler WM, Huo D, George C, et al. Prognostic factors for survival with gemcitabine plus 5-fluorouracil based regimens for metastatic renal cancer. *J Urol* 2003;170:1141–1145. [PubMed: 14501711]
18. Kagan RS, Joseph L, Dufresne C, et al. Prevalence of peanut allergy in primary-school children in Montreal. *Can J Allergy Clin Immunol* 2003;112:1223–1228. [PubMed: 14657887]
19. Barzi F, Woodward M. Imputations of missing values in practice: results from imputations of serum cholesterol in 28 cohort studies. *Am J Epidemiol* 2004;160:34–45. [PubMed: 15229115]
20. DiRusso SM, Chahine AA, Sullivan T, et al. Development of a model for prediction of survival in pediatric trauma patients: comparison of artificial neural networks and logistic regression. *J Pediatr Surg* 2002;37:1098–1104. [PubMed: 12077780]
21. Healey C, Osler TM, Rogers FB, et al. Improving the Glasgow Coma Scale score: motor score alone is a better predictor. *J Trauma Injury Infect Crit Care* 2003;54:671–678.
22. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med* 1999;130:515–524. [PubMed: 10075620]
23. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med* 2000;19:453–473. [PubMed: 10694730]

24. Tepas JJ. Discussion of Potoka DA, Schall LC, Ford HR. Development of a novel age-specific pediatric trauma score. *J Pediatr Surg* 2001;36:106–112. [PubMed: 11150447]
25. Gabbe BJ, Cameron PA, Finch CF. Is the revised trauma score still useful? *ANZ J Surg* 2003;73:944–948 [PubMed: 14616576]
26. Meredith W, Rutledge R, Fakhry SM, et al. The conundrum of the Glasgow Coma Scale in intubated patients: a linear regression prediction of the Glasgow verbal score from the Glasgow eye and motor scores. *J Trauma* 1998;44:839–44. [PubMed: 9603086]

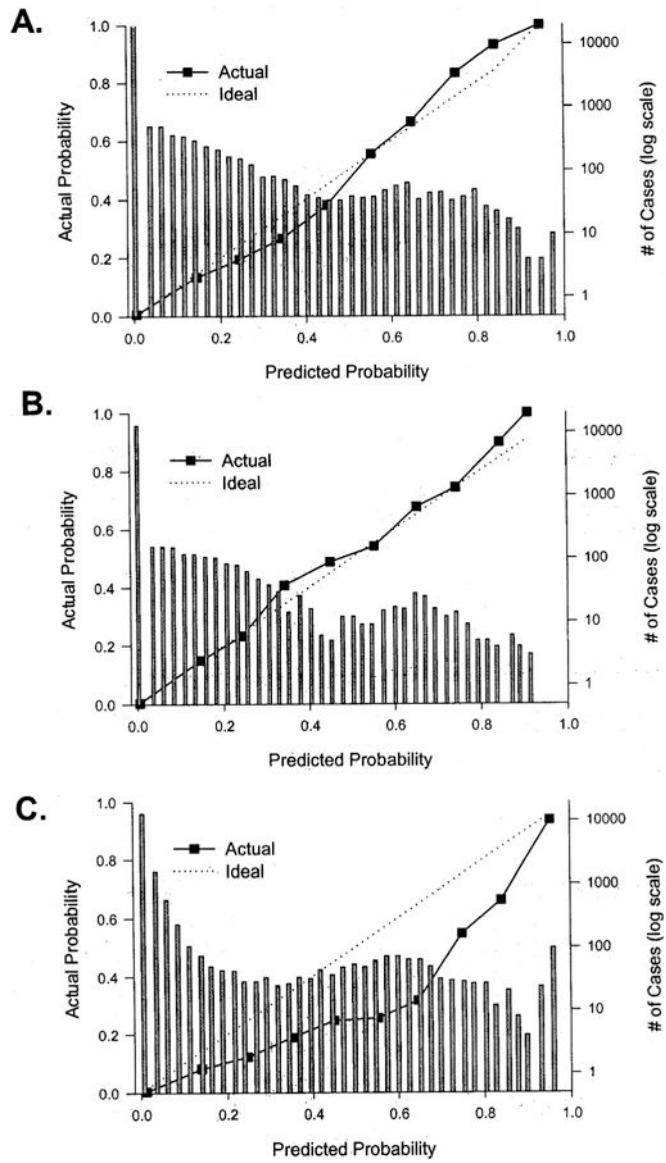


Fig. 1. Calibration curve showing actual mortality rates across deciles of predicted mortality risk for test cases. Data obtained from NPTR training set (A), NPTR testing set (B) and NTDB (C). The line with square bullets represents the actual prediction in each decile; the dashed line represents perfect prediction. The vertical bars indicate the number of cases in each risk group.

Table 1

Univariate Analysis of Variables in Training and Testing Sets

Variable	NPTR Training Set			NPTR Testing Set			NTDB Testing Set		
	All Patients (n = 34,342)	Alive (n = 33,334)	Dead (n = 1,008)	All Patients (n = 14,200)	Alive (n = 13,824)	Dead (n = 376)	All Patients (n = 16,868)	Alive (n = 16,339)	Dead (n = 529)
Age, years (mean ± SD)	8 ± 5 (0)	8 ± 5 (0)	7 ± 5 (0) [†]	7.8 ± 5.0 (0)	7.9 ± 5.0 (0)	7.1 ± 5.5 (0) [*]	10.6 ± 5.6 (0)	10.5 ± 5.6 (0)	11.3 ± 6.3 (0) [*]
Standardized SBP (mean ± SD)	0.0 ± 1.0 (6)	0.0 ± 1.0 (5)	-0.8 ± 1.8 (23) [†]	0.0 ± 1.0 (7)	0.0 ± 1.0 (6)	-0.5 ± 1.6 (24) [†]	0.0 ± 1.3(16)	0.1 ± 1.1(16)	-2.2 ± 3.1 (11) [†]
Standardized pulse (mean ± SD)	0.0 ± 1.0 (2)	0.0 ± 1.0 (1)	0.2 ± 1.7 (21) [†]	0.0 ± 1.0 (2)	0.0 ± 1.0 (1)	0.4 ± 1.6 (20) [†]	-(100)	-(100)	-(100)
Standardized RR (mean ± SD)	0.0 ± 1.0 (12)	0.0 ± 1.0 (9)	0.1 ± 1.8 (92)	0.0 ± 1.1 (12)	0.0 ± 1.1(9)	0.39 ± 1.6 (93)	-0.3 ± 1.2 (14)	-0.2 ± 1.1 (13)	-2.4 ± 2.1 (29) [†]
GCS									
Eye (median)	4 (22)	4 (20)	2 (28) [†]	4 (25)	4 (25)	3 (16) [†]	4 (22)	4 (23)	1 (13) [†]
Verbal (median)	5 (29)	5 (25)	3 (97) [†]	5 (33)	5 (32)	4 (97) [†]	5 (24)	5 (24)	1 (14) [†]
Motor (median)	6 (26)	6 (23)	5 (49) [†]	6 (31)	6 (30)	5 (54) [†]	6 (23)	6 (23)	1 (13) [†]
Total (median)	15 (24)	15 (21)	10 (97) [†]	15 (23)	15 (21)	11 (97) [†]	15 (22)	15 (22)	3 (11) [†]
Indubated (%)	12 (1)	9 (1)	97 (6) [†]	11 (1)	8 (1)	94 (3) [†]	8 (0)	6 (0)	61 (0) [†]
Penetrating injury (%)	7 (0)	7 (0)	8 (0)	6 (0)	6 (0)	6 (0)	7 (6)	7 (6)	17 (6) [†]

% missing indicated in parentheses.

* $p < 0.01$ compared to surviving patients.

[†] $p < 0.001$.

NPTR, National Pediatric Trauma Registry; NTDB, National Trauma Data Bank; SD, standard deviation, SBP, systolic blood pressure; ROC, receiver operating characteristic; —, pulse not available in the NTDB dataset.

Table 2

Logistic Regression Analysis of Risk of Death in Pediatric Trauma Patients Based on Complete Data (n 23,695) in NPTR Training Set

Variable	Coefficient	SE	<i>p</i> value	Odds Ratio	95% CI Lower Limit	95% CI Upper Limit
Intercept	-1.95	1.21	0.107			
SBP—standardized	-0.01	0.23	0.942	0.98	0.61	1.57
Pulse—standardized	0.44	0.23	0.058	1.55	0.98	2.45
RR—standardized	-0.18	0.23	0.436	0.83	0.52	1.32
GCS-eye	-0.77	0.36	0.036	0.46	0.22	0.95
GCS-motor	-0.37	0.31	0.228	0.68	0.37	1.26
GCS-verbal	-0.13	0.34	0.693	0.87	0.44	1.73
Intubated	2.63	0.67	0.0001	13.96	13.68	52.87

SE, standard error; CI, confidence interval; SBP, systolic blood pressure; RR, respiratory rate; GCS, Glasgow Coma Scale.

Table 3

Missing Data Patterns in the NPTR (Training and Testing Sets) and NTDB *

	SBP	Pulse	RR	GCS-eye	GCS-motor	GCS-verbal	GCS-total	Intubation Status	Cases With Pattern (%)
NPTR									
	+	+	+	+	+	+	+	+	67
	+	+	+	-	-	-	-	+	10
	+	+	+	-	-	-	+	+	5
	+	+	-	-	-	-	-	+	3
	+	+	-	+	-	-	-	+	3
	-	+	+	+	+	+	+	+	2
	+	+	-	+	+	-	-	+	2
	-	+	+	-	-	-	-	+	2
NTDB									
	+	-	+	+	+	+	+	+	68
	-	-	-	-	-	-	-	+	10
	+	-	+	-	-	-	-	+	9
	-	-	+	+	+	+	+	+	3
	+	-	-	+	+	+	+	+	3
	-	-	+	-	-	-	-	+	2

+, data is present

-, data is absent

* missing data patterns are shown that occur in at least 2% of subjects.

NPTR, National Pediatric Trauma Registry; NTDB, National Trauma Data Bank; SBP, systolic blood pressure; RR, respiratory rate; GCS, Glasgow Coma Scale.

Table 4

Associations Between Missingness Data and Variables in NPTR Training Dataset and Mortality

Absence of:	Variables Used for Imputation								Penetrating Injury	Intubation Status	Outcome Mortality
	Age	SBP	P	RR	GCS-eye	GCS-motor	GCS-verbal	GCS-total			
SBP	X	NA	X	X	X	X	X	—	X	X	X
Pulse	—	X	NA	—	X	X	—	X	—	X	X
RR	X	—	X	NA	X	X	X	X	X	X	X
GCS-eye	X	X	X	X	NA	X	X	X	X	X	X
GCS-motor	X	—	X	X	X	NA	X	X	—	X	X
GCS-verbal	X	X	X	X	X	X	NA	X	—	X	X
Intubation Status	—	—	—	—	X	X	—	—	X	NA	X

X, associated using univariate logistic regression ($p < 0.05$); —, no association using univariate logistic regression; NA, not applicable; SBP, systolic blood pressure; RR, respiratory rate; GCS, Glasgow Coma Scale.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 5

Comparison of Values from the Original NPTR Training Dataset with Values From the 10 Imputed Datasets

Variable	Original Data	Imputed Datasets
SBP—standardized (mean [range])	0.0(−7.5–4.9)	0.0(−7.4–5.0)
Pulse—standardized (mean [range])	0.0(−5.1–5.6)	0.0(−5.1–6.6)
RR—standardized (mean [range])	0.0(−4.7–11.5)	0.0(−4.4–11.9)
GCS-eye (median [range])	4(1–4)	4(1–4)
GCS-motor (median [range])	5(1–5)	5(1–5)
GCS-verbal (median [range])	6(1–6)	6(1–6)
Intubated (%)	7	11

SBP, systolic blood pressure; RR, respiratory rate; GCS, Glasgow Coma Scale.

Table 6

Logistic Regression Analysis of Risk of Death in Pediatric Trauma Patients Based on Imputed NPTR Training Data (n = 34,342)

Variable	Coefficient	SE	P value	Odds Ratio	95% CI Lower Limit	95% CI Upper Limit
Intercept	-1.471	0.331				
SBP—standardized	-0.387	0.034	<0.0001	0.679	0.635	0.726
Pulse—standardized	-0.131	0.037	0.0006	0.877	0.815	0.944
RR—standardized	-0.007	0.038	0.852	0.993	0.922	1.071
GCS-eye	-0.375	0.055	<0.0001	0.687	0.617	0.765
GCS-motor	-0.589	0.038	<0.0001	0.555	0.515	0.598
GCS-verbal	0.020	0.070	0.776	1.020	0.888	1.172
Intubated	2.953	0.170	<0.0001	19.163	13.736	26.775

CI, confidence interval; SE, standard error; SBP, systolic blood pressure; RR, respiratory rate; GCS, Glasgow Coma Scale.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 7

Comparison of Discrimination and Calibration of Models Developed Using Complete Case Analysis and Multiple Imputation

Dataset	Model Based on Complete Case Analysis	Model Based on Multiple Imputation
Discrimination (c-index)		
NPTR testing data	0.965	0.972
NTDB	0.940	0.947
Calibration (H-L c-statistic)		
NPTR testing data	242.6	9.2
NTDB	398.1	258.2
Calibration (R ²)		
NPTR testing data	0.793	0.996
NTDB	0.922	0.877

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 8

Evaluation of Discrimination and Calibration on Testing Datasets

Testing Dataset	Records (n)	Actual Mortality (%)	C-index (95% CI)	H-L c-statistic	H-L h-statistic	R ²
NPTR testing data (all cases)	14,200	2.6	0.972 (0.966, 0.978)	9.2	6.8	0.996
NPTR testing data (blunt mechanism)	13,312	2.7	0.973 (0.967, 0.979)	8.4	11.8	0.987
NPTR testing data (penetrating mechanism)	888	2.6	0.952 (0.916, 0.988)	12.7	19.0	0.451
NTDB (all cases [*])	16,868	3.1	0.947 (0.936, 0.959)	258.2	377.3	0.877
NTDB (blunt mechanism [*])	14,700	2.8	0.953 (0.941, 0.965)	209.2	411.6	0.849
NTDB (penetrating mechanism [*])	1,080	7.8	0.966 (0.943, 0.990)	19.9	18.4	0.696

^{*}Pulse estimated using multiple imputation in all cases. CI, confidence interval; H-L, Hosmer-Lemeshow.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 9

Comparison of the Predictive Validity of the Proposed Model with the Revised Trauma Score (RTS) and Pediatric Trauma Score (PTS)*

Testing Dataset	Sensitivity	Specificity	Undertriage (1-sensitivity)	Overtriage (1-sensitivity)	Positive Predictive Value	Negative Predictive Value	Accuracy
NPTR testing data							
Proposed model	96(94–98)	91(90–91)	4(2–6)	9(9–10)	22(20–24)	99 (99–99)	91(90–91)
RTS	95(93–97)	71(70–72) ^{//}	5(3–7)	29(28–30) ^{//}	8(7–9) ^{//}	99 (99–99)	72(71–72) ^{//}
PTS [‡]	99(98–100) [§]	71(70–71) ^{//}	1(0–2) [§]	29(29–30) ^{//}	7(6–8) ^{//}	99 (99–99)	71(70–72) ^{//}
NTDB [‡]							
Proposed model	97(95–98)	60(59–61)	3(2–5)	40(39–41)	7(7–8)	99 (99–99)	61(60–62)
RTS	94(92–96)	70(69–71) ^{//}	6(4–8)	30(29–31) ^{//}	9(9–10) ^{//}	99 (99–99)	71(70–72) ^{//}

* All indicated values are percentages; 95% confidence intervals indicated in parentheses.

[‡] based on values available in the NPTR database (n = 13,135).

[‡] PTS not available in the NTDB.

[§] $P < 0.01$.

^{//} $P < 0.001$ compare to proposed model