

Original Paper

# Enhancing Performance of the National Field Triage Guidelines Using Machine Learning: Development of a Prehospital Triage Model to Predict Severe Trauma

Qi Chen<sup>1\*</sup>, PhD; Yuchen Qin<sup>1\*</sup>, PhD; Zhichao Jin<sup>1\*</sup>, PhD; Xinxin Zhao<sup>2</sup>, PhD; Jia He<sup>1</sup>, PhD; Cheng Wu<sup>1\*</sup>, PhD; Bihan Tang<sup>3</sup>, PhD

<sup>1</sup>Department of Health Statistics, Naval Medical University, Shanghai, China

<sup>2</sup>School of Medicine, Tongji University, Shanghai, China

<sup>3</sup>Department of Health Management, Naval Medical University, Shanghai, China

\*these authors contributed equally

**Corresponding Author:**

Bihan Tang, PhD

Department of Health Management

Naval Medical University

No 800 Xiangyin Road

Shanghai, 200433

China

Phone: 86 02181871425

Fax: 86 02181871425

Email: [mangotangbihan@126.com](mailto:mangotangbihan@126.com)

## Abstract

**Background:** Prehospital trauma triage is essential to get the right patient to the right hospital. However, the national field triage guidelines proposed by the American College of Surgeons have proven to be relatively insensitive when identifying severe traumas.

**Objective:** This study aimed to build a prehospital triage model to predict severe trauma and enhance the performance of the national field triage guidelines.

**Methods:** This was a multisite prediction study, and the data were extracted from the National Trauma Data Bank between 2017 and 2019. All patients with injury, aged 16 years of age or older, and transported by ambulance from the injury scene to any trauma center were potentially eligible. The data were divided into training, internal, and external validation sets of 672,309; 288,134; and 508,703 patients, respectively. As the national field triage guidelines recommended, age, 7 vital signs, and 8 injury patterns at the prehospital stage were included as candidate variables for model development. Outcomes were severe trauma with an Injured Severity Score  $\geq 16$  (primary) and critical resource use within 24 hours of emergency department arrival (secondary). The triage model was developed using an extreme gradient boosting model and Shapley additive explanation analysis. The model's accuracy regarding discrimination, calibration, and clinical benefit was assessed.

**Results:** At a fixed specificity of 0.5, the model showed a sensitivity of 0.799 (95% CI 0.797-0.801), an undertriage rate of 0.080 (95% CI 0.079-0.081), and an overtriage rate of 0.743 (95% CI 0.742-0.743) for predicting severe trauma. The model showed a sensitivity of 0.774 (95% CI 0.772-0.776), an undertriage rate of 0.158 (95% CI 0.157-0.159), and an overtriage rate of 0.609 (95% CI 0.608-0.609) when predicting critical resource use, fixed at 0.5 specificity. The triage model's areas under the curve were 0.755 (95% CI 0.753-0.757) for severe trauma prediction and 0.736 (95% CI 0.734-0.737) for critical resource use prediction. The triage model's performance was better than those of the Glasgow Coma Score, Prehospital Index, revised trauma score, and the 2011 national field triage guidelines RED criteria. The model's performance was consistent in the 2 validation sets.

**Conclusions:** The prehospital triage model is promising for predicting severe trauma and achieving an undertriage rate of  $<10\%$ . Moreover, machine learning enhances the performance of field triage guidelines.

(*J Med Internet Res* 2024;26:e58740) doi: [10.2196/58740](https://doi.org/10.2196/58740)

**KEYWORDS**

severe trauma; field triage; machine learning; prediction model

## Introduction

Trauma is a universal health challenge that places a massive burden on national economies. It causes 4.4 million deaths annually, and an estimated 10% of all years lived with disability [1,2]. The American College of Surgeons Committee on Trauma (ACS-COT) recommends that severe trauma be treated at levels 1 and 2 trauma care facilities [3]. Patients with severe trauma have approximately 25% lower mortality rates when treated at levels 1 or 2 trauma centers than when treated at lower-level or nontrauma centers [4]. Prehospital estimation of injury severity is essential for prehospital triage. It is a critical step for emergency medical service (EMS) providers in making decisions regarding patient destination. Under- and overtriage are incorrect triages. A low-sensitivity triage tool results in many false-negative cases, indicating undertriage and a possible failure in trauma first aid. Conversely, low specificity is associated with a high rate of false-positive cases, indicating overtriage [5].

The national field triage guidelines were initially developed by ACS-COT in 1976 and revised in 2011 and 2021 [6,7]. The national field triage guidelines have been widely implemented in the United States and represent 1 of the few standardized national protocols for EMS. It was developed based on peer-reviewed research, resulting in biased estimates and reduced generalizability [3,8]. A prospective national triage guidelines validation study for identifying high-risk trauma patients found that the guidelines were relatively insensitive in identifying severely injured patients and those requiring early critical resource use [9]. In addition, other triage tools, such as the Glasgow Coma Score (GCS), Prehospital Index (PHI), and revised trauma score (RTS), have not shown ideal predictive performance [5,10-12]. Therefore, it emphasizes the need to improve the prehospital triage tool [13].

Machine learning (ML) development has advanced rapidly in the medical field, notably in trauma medicine, and has demonstrated that the ML model's predictive ability is significantly better than that of the conventional trauma triage tools for mortality outcomes, hospitalization, and critical care admission [14,15]. Therefore, this study aimed to build a

prehospital triage model to predict severe trauma (pTEST) and enhance the performance of the national field triage guideline.

## Methods

### Recruitment

This multisite prediction study was conducted to predict trauma severity during field triage. We developed, validated, and reported our triage model following the TRIPOD (Transparent Reporting of a Multivariable Model for Individual Prognosis or Diagnosis) statement [16], as shown in [Multimedia Appendix 1](#).

### Ethical Considerations

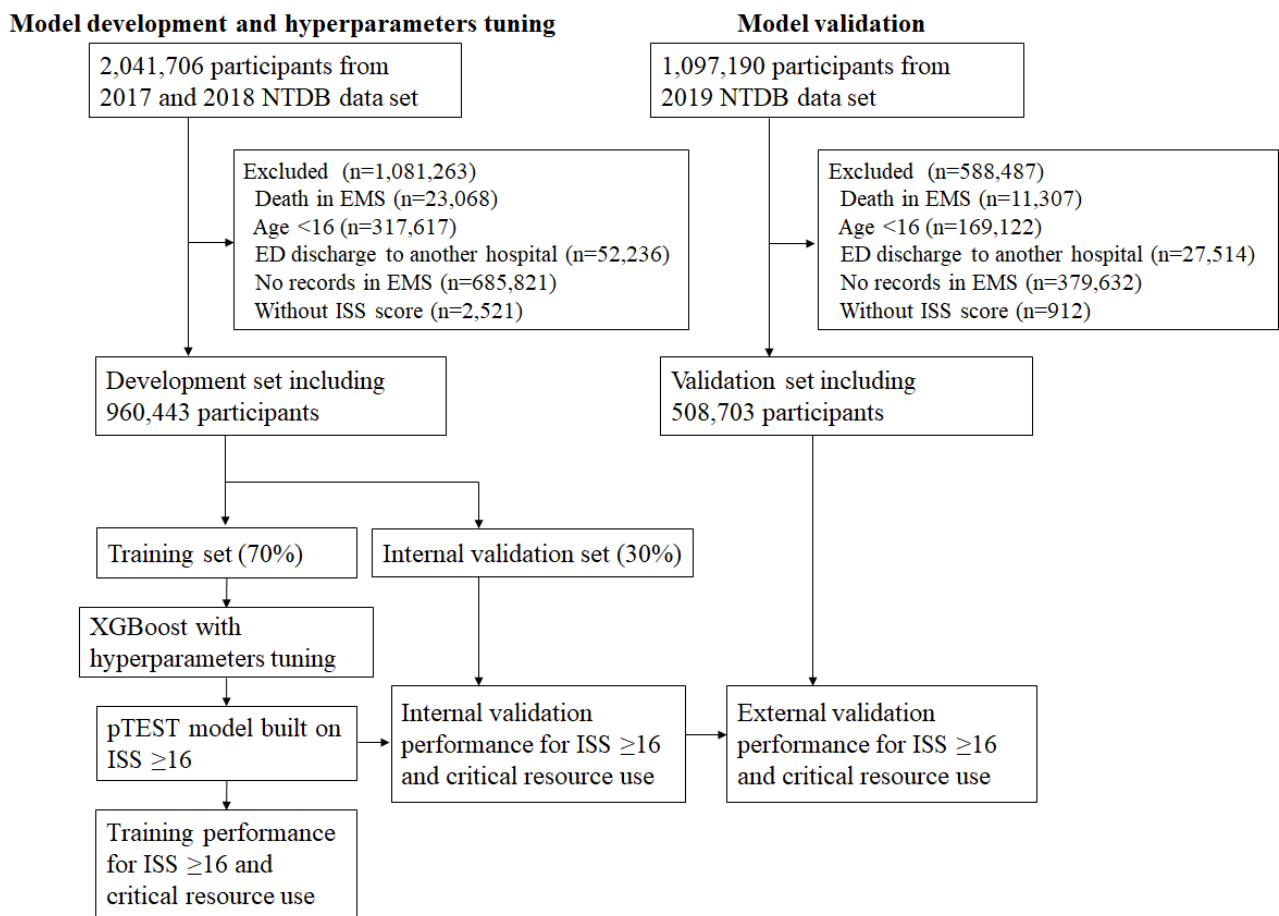
The Naval Medical University Ethics Committee approved the study protocol (reference number NMUEC2022-088). Our study consisted of secondary analyses using the National Trauma Data Bank (NTDB) with primary consent, and the data were anonymized.

### Source of Data and Patients

Data from the NTDB, the largest aggregation of trauma registry data in the United States assembled by the ACS, were used in this study [17]. In 2017, the ACS Trauma Quality Program transitioned to a new technical vendor and redesigned the NTDB infrastructure. The 2017 and 2018 NTDB datasets comprising 2,041,706 patients were used for pTEST model development, hyperparameter tuning, and internal validation. The 2019 NTDB dataset comprising 1,097,190 patients was used for the external validation of the pTEST model.

Notably, all patients with injury, aged 16 years of age or older, and transported by ground or aerial ambulance from the injury scene to any trauma center were potentially eligible. Due to a lack of crucial information on outcomes and predictors, we excluded patients who died in the EMS, patients discharged from the emergency department (ED) to another hospital, and those without any EMS records or an Injury Severity Score (ISS). In total, 960,443 and 508,703 participants were included in the development and validation sets, respectively. [Figure 1](#) shows the patient selection process.

**Figure 1.** The flowchart of model development and validation. ED: emergency department; EMS: emergency medical service; ISS: Injury Severity Score; NTDB: National Trauma Data Bank; pTEST: prehospital triage model to predict severe trauma.



**Outcome and Predictors**

Defining “severe trauma” is a challenge in prehospital triage model development. It varies widely across studies. The reference standard (primary outcome) of “severe injury” was considered as an ISS  $\geq 16$  as the benchmark to evaluate triage accuracy recommended by the ACS [7,18,19]. The ISS calculated by anatomical criteria is assumed to be consistent with the patient status on the scene and is associated with high mortality [4,19]. However, it does not reflect resource use directly; therefore, we included a composite resource-based early critical resource use measure as the secondary outcome. According to similar studies [9,18], early critical resource use included intubation in the EMS or ED, discharge to the intensive care unit from the ED, surgery for hemorrhage control, interventional radiology procedures, cerebral monitoring, and in-hospital death, all within 24 hours. A detailed definition of severe trauma is provided in [Multimedia Appendix 2](#).

According to a recent field triage protocol review in 2017 [10], the significant predictors of a severely injured patient were age, vital signs, injury patterns, and injury mechanism. In addition, in the US National Guidelines for the Field Triage of Injured Patients in 2011 and 2021 [6,7], age, vital signs, and injury pattern measurements were the field triage’s top priorities. In the 2011 national field triage guidelines, severely injured patients who should be transported preferentially to the highest-level trauma center were identified using the RED

criteria, which included 3 vital signs (GCS, systolic blood pressure, and respiratory rate) and 8 injury patterns. During field triage, time is essential, and the number and complexity of hand-collected variables must be limited. Therefore, as the US National Guideline for the Field Triage recommended and recorded in the NTDB, we incorporated 16 candidate variables in EMS for model development, including age at the time of injury (AGEYEARS), GCS eye (EMSGCSEYE), GCS motor (EMSGCSMOTOR), GCS verbal (EMSGCSVERBAL), systolic blood pressure (EMSSBP), oxygen saturation (EMSPULSEOXIMETRY), respiratory rate (EMSRESPIRATORYRATE), pulse rate (EMSPULSERATE), penetrating injuries (TCCPEN), chest wall instability (TCCCHEST), long-bone fractures (TCCLONGBONE), crushed extremity (TCCCRUSHED), amputation (TCCAMPUTATION), pelvic fracture (TCCPELVIC), skull fracture (TCCSKULLFRACTURE), and paralysis (TCCPARALYSIS). The detailed definitions of candidate variables are listed in [Multimedia Appendix 3](#).

**Model Development**

The pTEST model was developed using the extreme gradient boosting model (XGBoost) and Shapley additive explanation analysis (SHAP). XGBoost is a novel boosting tree-based ensemble algorithm through which new models are created to predict residuals or errors of prior models and then combined to make a final prediction [20]. Recently, XGBoost has been widely used in ML due to its outstanding prediction

performance, ability to use continuous and categorical inputs, lack of data preprocessing, imbalanced data handling capacity, high internal optimization, and relatively modest computational costs [21].

Patients in the development set from the 2017 and 2018 NTDB datasets were randomly grouped into training (70%) and internal validation (30%) sets for model development. Using a grid search, a 10-fold cross-validation process was used on the training set for hyperparameter tuning. The goal of hyperparameter tuning is to find the values that lead to the best-predicted performance. The optimal values of hyperparameters were learning rate=0.04071151, minimum loss reduction required to make a further partition=20.36485, maximum tree depth=14, minimum sum of instance weights needed in a leaf node=39, maximum number of boosting iterations=1051, and subsample ratio of the training instance=0.7763707, with the other hyperparameters set to default. In addition to the training set, the model's reproducibility, transportability, and generalizability were evaluated using internal and external validation sets.

Missing values are an essential concern in trauma triage because there may not always be time to measure critical variables. The absent proportions of the training, internal validation, and external validation sets are shown in [Multimedia Appendix 4](#), with all variables, except pulse oximetry, missing below 6%. XGBoost supports branch directions for predictors with missing values, creating an advantage in real-world situations where XGBoost can still achieve individual prediction without complete prehospital data.

To gain insight into the risk prediction model, we investigated different predictors' contributions based on Shapley values using SHAP, a game theory concept introduced in the 1950s [22,23]. A predictor's SHAP value can be positive or negative, suggesting an increased or decreased probability of severe trauma, respectively. In our study, the SHAP values were visualized at global (dataset level) and local (patient-specific) levels to investigate the predictors' impact. XGBoost and SHAP were implemented using the R (R Core Team) packages *tidy models* and *shapviz*.

### Statistical Analysis Methods

For the sample size calculation, the prevalence of events was set at 17.8%, and the number of predictors was 16 based on the development set. The area under the curve (AUC) of the optimal prehospital triage model in a previous study was 0.68 [11], and our pTEST model was expected to achieve an AUC of 0.7. At least 1871 patients were required for model development or validation using the R package *pmsampsize* [24].

Continuous data are presented as mean and SD, and categorical data are presented as frequencies and percentages (%). The 2-tailed *t* test was used to evaluate the differences in continuous data, which followed a normal distribution and variance homogeneity; otherwise, the Wilcoxon rank-sum test was used.

The differences in categorical data were evaluated using the Pearson chi-square test. The area under the receiver operating characteristic curve was calculated to assess model discrimination. The AUCs between models were compared using the DeLong test. The best thresholds of the models were determined by maximizing the Youden index, and performance metrics, including sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value (NPV), were calculated. Performance metrics 95% CIs were calculated using 500 bootstrap replicates. In addition, the pTEST model is intended to identify severe trauma that requires high sensitivity and NPV to rule it out. Sensitivity and specificity are inversely proportional and a tradeoff needs to be made between sensitivity and specificity. Therefore, the sensitivities and NPV of the different models were compared using a 0.5 specificity as in the previous studies [25,26].

Our study defined the over- and undertriage rates. Overtriage rate (1-PPV) is equal to the number of patients with negative outcomes (ISS <16 or no critical resource use) predicted as positive outcomes divided by the total number of predicted positive outcomes. The undertriage rate (1-NPV) is equal to the number of patients with positive outcomes predicted as negative outcomes divided by the total number of predicted negative outcomes.

In addition, a calibration plot and Brier score were generated to assess how closely the predicted probability approximated the actual probability. The clinical benefit of the models was evaluated using a decision curve analysis method. The discrimination, calibration, and clinical benefit of the pTEST were compared with the GCS, PHI, RTS, and RED criteria of the 2011 National Field Triage Guidelines. Statistical significance was set at  $P<.05$ . All statistical analyses were performed using the R software (version 4.3.1).

## Results

### Patient Characteristics

**Table 1** shows the patients' baseline characteristics in the training, internal validation, and external validation sets. In these 3 sets, severe trauma proportions were 17.80% (119,690/672,309), 17.80% (51,296/288,134), and 17.08% (86,902/508,703), respectively, and critical resource use proportions were 29.36% (177,570/604,806), 29.56% (76,604/259,148), and 28.17% (129,551/459,843), respectively. Notably, most variables showed statistically significant differences among the 3 sets for large sample sizes, but the differences were minimal. The demographic characteristics, vital signs, and injury patterns of nonsevere and severe trauma are listed in [Multimedia Appendices 5-7](#), and the attributes of noncritical and critical resource users are listed in [Multimedia Appendices 8-10](#). Severe trauma and critical resource users are usually male, air-transported, taken to higher-level trauma centers, and have extreme trauma patterns.

**Table 1.** Baseline characteristics of the patients from the training set, internal validation set, and external validation set.

Characteristics	Training set (n=672,309)	Internal validation set (n=288,134)	External validation set (n=508,703)	P value
<b>Sex (male), n (%)</b>	403,434 (60.01)	172,821 (59.98)	300,711 (59.12)	<.001
<b>Transport mode, n (%)</b>				<.001
Ground	622,489 (92.59)	266,622 (92.53)	474,645 (93.30)	
Helicopter	48,464 (7.21)	20,911 (7.26)	33,346 (6.56)	
Fixed-wing	1356 (0.20)	601 (0.21)	712 (0.14)	
<b>Trauma center level, n (%)</b>				<.001
Level 1	275,723 (55.75)	118,199 (55.72)	206,563 (54.39)	
Level 2	179,419 (36.28)	77,005 (36.30)	142,433 (37.50)	
Level 3	39,397 (7.97)	16,944 (7.99)	30,819 (8.11)	
<b>TCCPEN<sup>a</sup> (yes), n (%)</b>	26,987 (4.01)	11,319 (3.93)	17,658 (3.47)	<.001
<b>TCCCHEST<sup>b</sup> (yes), n (%)</b>	4295 (0.64)	1825 (0.63)	3216 (0.63)	.89
<b>TCCLONGBONE<sup>c</sup> (yes), n (%)</b>	4598 (0.68)	1867 (0.65)	3213 (0.63)	.002
<b>TCCCRUSHED<sup>d</sup> (yes), n (%)</b>	3188 (0.47)	1374 (0.48)	2712 (0.53)	<.001
<b>TCCAMPUTATION<sup>e</sup> (yes), n (%)</b>	832 (0.12)	372 (0.13)	613 (0.12)	.58
<b>TCCPELVIC<sup>f</sup> (yes), n (%)</b>	7495 (1.11)	3346 (1.16)	5852 (1.15)	.07
<b>TCCSKULLFRACTURE<sup>g</sup> (yes), n (%)</b>	5127 (0.76)	2237 (0.78)	4087 (0.80)	.04
<b>TCCPARALYSIS<sup>h</sup> (yes), n (%)</b>	4379 (0.65)	1887 (0.65)	3077 (0.60)	.003
<b>ISS<sup>i</sup> score (≥16), n (%)</b>	119,690 (17.80)	51,296 (17.80)	86,902 (17.08)	<.001
<b>Surgery for hemorrhage control (yes), n (%)</b>	14,714 (2.45)	6441 (2.50)	10,886 (2.37)	.002
<b>Cerebral monitor (yes), n (%)</b>	8527 (1.42)	3563 (1.38)	5819 (1.27)	<.001
<b>Interventional radiology procedures (yes), n (%)</b>	5154 (0.86)	2182 (0.85)	3848 (0.84)	.59
<b>Discharge to the ICU<sup>j</sup> from ED<sup>k</sup> (yes), n (%)</b>	142,422 (21.46)	61,384 (21.58)	105,670 (21.06)	<.001
<b>In-hospital death within 24 hours (yes), n (%)</b>	9303 (1.38)	3987 (1.38)	6371 (1.25)	<.001
<b>Intubation in the EMS<sup>l</sup> or ED (yes), n (%)</b>	72,755 (10.82)	31,080 (10.79)	49,779 (9.79)	<.001
<b>Critical resource use (yes), n (%)</b>	177,570 (29.36)	76,604 (29.56)	129,551 (28.17)	<.001
<b>RED criteria (yes), n (%)</b>	87,577 (13.03)	37,448 (13.00)	62,194 (12.23)	<.001
<b>Age (years), mean (SD)</b>	53.12 (21.86)	53.22 (21.84)	54.40 (21.77)	<.001
<b>EMSSBP<sup>m</sup> (mm Hg), mean (SD)</b>	139.89 (28.37)	139.92 (28.42)	140.74 (28.62)	<.001
<b>EMSPULSERATE<sup>n</sup> (n/minute), mean (SD)</b>	90.58 (20.31)	90.58 (20.29)	90.34 (20.41)	<.001
<b>EMSRESPIRATORYRATE<sup>o</sup> (n/minute), mean (SD)</b>	18.42 (4.73)	18.44 (4.79)	18.46 (4.82)	.08
<b>EMSPULSEOXIMETRY<sup>p</sup>, mean (SD)</b>	96.26 (5.50)	96.26 (5.48)	96.19 (5.44)	<.001
<b>EMSGCSEYE<sup>q</sup>, mean (SD)</b>	3.81 (0.65)	3.81 (0.65)	3.82 (0.63)	<.001
<b>EMSGCSVERBAL<sup>r</sup>, mean (SD)</b>	4.60 (0.95)	4.60 (0.95)	4.61 (0.93)	.02
<b>EMSGCSMOTOR<sup>s</sup>, mean (SD)</b>	5.73 (0.97)	5.73 (0.97)	5.75 (0.94)	<.001
<b>EMSTOTALGCS<sup>t</sup>, mean (SD)</b>	14.12 (2.44)	14.12 (2.43)	14.17 (2.36)	<.001
<b>Total time spent in ED (minutes), mean (SD)</b>	189.94 (150.35)	207.75 (400.72)	207.68 (1355.03)	<.001



Characteristics	Training set (n=672,309)	Internal validation set (n=288,134)	External validation set (n=508,703)	P value
<b>Length of stay (days), mean (SD)</b>	6.20 (9.67)	6.21 (8.51)	6.21 (8.38)	.001
<b>ISS score, mean (SD)</b>	9.69 (8.37)	9.69 (8.32)	9.55 (8.20)	<.001
<b>PHI<sup>u</sup> score, mean (SD)</b>	1.32 (2.23)	1.32 (2.23)	1.31 (2.22)	.26
<b>RTS<sup>v</sup> score, mean (SD)</b>	11.73 (0.88)	11.73 (0.89)	11.74 (0.86)	.045

<sup>a</sup>TCCPEN: penetrating injuries.

<sup>b</sup>TCCCHEST: chest wall instability.

<sup>c</sup>TCCCLONGBONE: long-bone fractures.

<sup>d</sup>TCCCRUSHED: crushed extremity.

<sup>e</sup>TCCAMPUTATION: amputation.

<sup>f</sup>TCCPELVIC: pelvic fracture.

<sup>g</sup>TCCSKULLFRACTURE: skull fracture.

<sup>h</sup>TCCPARALYSIS: paralysis.

<sup>i</sup>ISS: Injury Severity Score.

<sup>j</sup>ICU: intensive care unit.

<sup>k</sup>ED: emergency department.

<sup>l</sup>EMS: emergency medical service.

<sup>m</sup>EMSSBP: systolic blood pressure.

<sup>n</sup>EMSPULSERATE: pulse rate.

<sup>o</sup>EMSRESPIRATORYRATE: respiratory rate.

<sup>p</sup>EMSPULSEOXIMETRY: oxygen saturation.

<sup>q</sup>EMSGCSEYE: Glasgow Coma Score eye.

<sup>r</sup>EMSGCSVERBAL: Glasgow Coma Score verbal.

<sup>s</sup>EMSGCSMOTOR: Glasgow Coma Score motor.

<sup>t</sup>EMSTOTALGCS: Glasgow Coma Score total.

<sup>u</sup>PHI: Prehospital Index.

<sup>v</sup>RTS: revised trauma score.

## Model Performance

For predicting severe trauma, we compared the performance metrics of the other models at the same specificity fixed at a moderate number of 0.5. The pTEST model showed a higher sensitivity of 0.799 (95% CI 0.797-0.801), a lower undertriage rate of 0.080 (95% CI 0.079-0.081), and a lower overtriage rate of 0.743 (95% CI 0.742-0.743) in the training set (Table 2). In addition, for critical resource use prediction fixed at a specificity of 0.5, the pTEST model showed a higher sensitivity of 0.774

(95% CI 0.772-0.776), lower undertriage rate of 0.158 (95% CI 0.157-0.159), and lower overtriage rate of 0.609 (95% CI 0.608-0.609) than the other models in the training set (Multimedia Appendix 11). We validated the pTEST model performance using 2 validation sets and obtained consistent results (Table 2 and Multimedia Appendix 11). The model performance metrics for predicting severe trauma and critical resource use at the best thresholds with the maximum Youden index are listed in Multimedia Appendices 12 and 13, demonstrating a higher pTEST Youden index than other models.

**Table 2.** Model performance metrics for predicting severe trauma fixed at a specificity of 0.5.

Prediction tool	Specificity	Sensitivity (95% CI)	Accuracy, AUC (95% CI)	Undertriage rate (1-NPV), AUC (95% CI)	Overtriage rate (1-PPV), AUC (95% CI)	Youden index
<b>Training set</b>						
pTEST <sup>a</sup>	0.500	0.799 (0.797-0.801)	0.553 (0.553-0.554)	0.080 (0.079-0.081)	0.743 (0.742-0.743)	1.299
GCS <sup>b</sup>	0.500	0.682 (0.680-0.684)	0.532 (0.532-0.532)	0.119 (0.119-0.120)	0.774 (0.774-0.775)	1.182
PHI <sup>c</sup>	0.500	0.711 (0.709-0.713)	0.536 (0.536-0.537)	0.107 (0.107-0.108)	0.772 (0.771-0.772)	1.211
RTS <sup>d</sup>	0.500	0.634 (0.632-0.636)	0.523 (0.523-0.523)	0.132 (0.132-0.133)	0.791 (0.791-0.792)	1.134
RED criteria	0.500	0.620 (0.619-0.621)	0.521 (0.521-0.522)	0.141 (0.141-0.142)	0.788 (0.788-0.789)	1.120
<b>Internal validation set</b>						
pTEST	0.500	0.794 (0.791-0.798)	0.552 (0.552-0.553)	0.082 (0.081-0.083)	0.744 (0.743-0.745)	1.294
GCS	0.500	0.682 (0.680-0.685)	0.532 (0.532-0.533)	0.120 (0.119-0.120)	0.774 (0.773-0.775)	1.182
PHI	0.500	0.710 (0.707-0.714)	0.536 (0.536-0.537)	0.108 (0.106-0.109)	0.772 (0.771-0.773)	1.210
RTS	0.500	0.633 (0.631-0.636)	0.523 (0.523-0.523)	0.132 (0.132-0.133)	0.791 (0.791-0.792)	1.133
RED criteria	0.500	0.620 (0.619-0.622)	0.521 (0.521-0.522)	0.141 (0.141-0.142)	0.788 (0.788-0.789)	1.120
<b>External validation set</b>						
pTEST	0.500	0.794 (0.792-0.797)	0.550 (0.550-0.551)	0.078 (0.077-0.079)	0.753 (0.753-0.754)	1.294
GCS	0.500	0.681 (0.679-0.683)	0.531 (0.530-0.531)	0.115 (0.115-0.116)	0.783 (0.782-0.783)	1.181
PHI	0.500	0.709 (0.706-0.711)	0.535 (0.534-0.535)	0.103 (0.103-0.104)	0.781 (0.780-0.782)	1.209
RTS	0.500	0.633 (0.631-0.635)	0.522 (0.522-0.522)	0.127 (0.126-0.128)	0.799 (0.799-0.800)	1.133
RED criteria	0.500	0.616 (0.615-0.618)	0.520 (0.520-0.520)	0.137 (0.136-0.137)	0.798 (0.797-0.798)	1.116

<sup>a</sup>pTEST: prehospital triage model to predict severe trauma.

<sup>b</sup>GCS: Glasgow Coma Score.

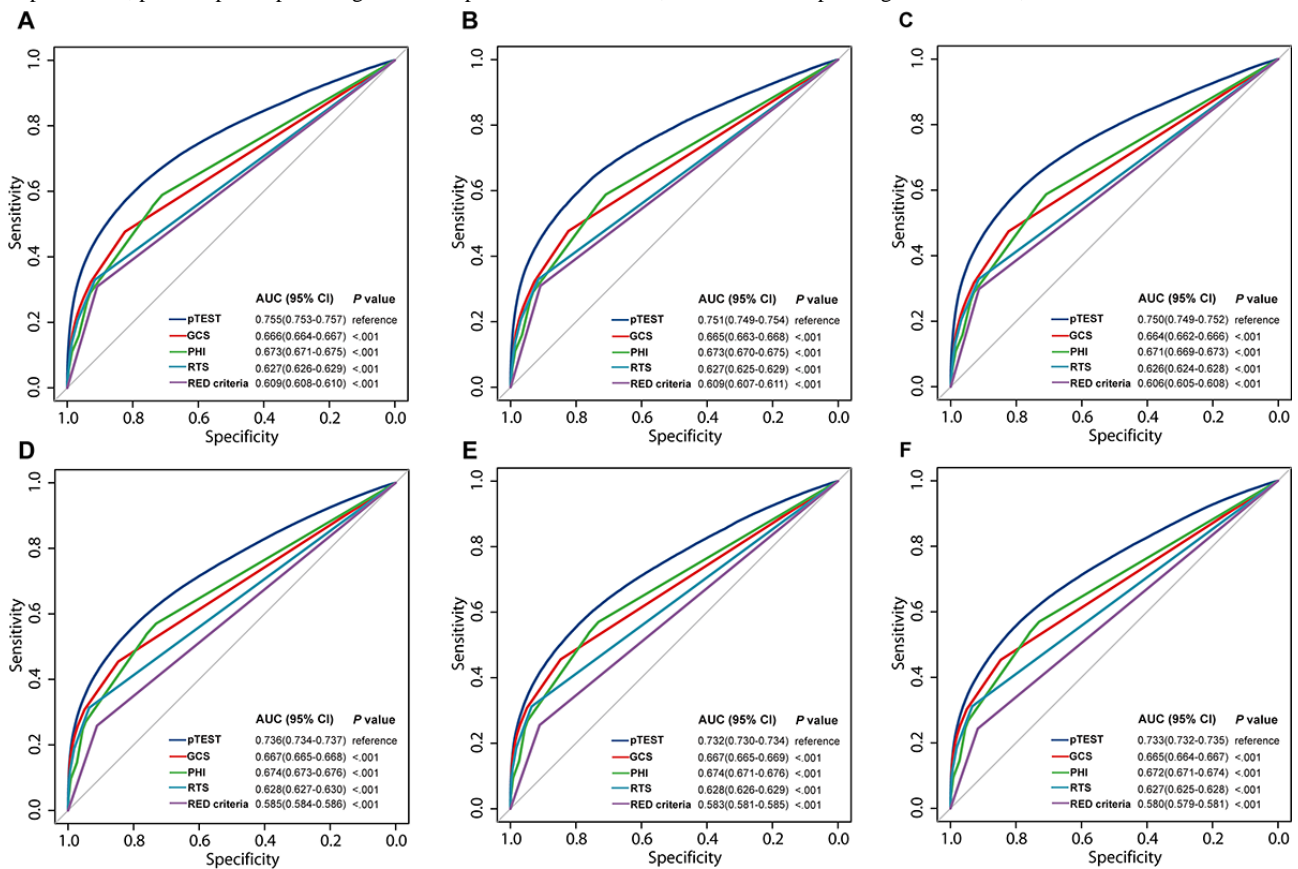
<sup>c</sup>PHI: Prehospital Index.

<sup>d</sup>RTS: revised trauma score.

In [Figure 2](#), pTEST AUCs for severe trauma prediction were 0.755 (95% CI 0.753-0.757), 0.751 (95% CI 0.749-0.754), and 0.750 (95% CI 0.749-0.752) in training, internal validation, and external validation sets, respectively, and the AUCs for predicting critical resource use were 0.736 (95% CI 0.734-0.737), 0.732 (95% CI 0.730-0.734), and 0.733 (95% CI 0.732-0.735), respectively, demonstrating better discrimination ability than GCS, PHI, RTS, and RED criteria. [Multimedia](#)

[Appendix 14](#) depicts the pTEST model predicted outcome probability as a waterfall plot. The calibration curves in [Multimedia Appendix 15](#) show that the severe trauma predicted probability and pTEST critical resource use agreed with the proportion observed using the smallest Brier score. In [Multimedia Appendix 16](#), pTEST provides a consistently higher net benefit across a broad range of risk thresholds (10%-100%) than the 2 default strategies and other models.

**Figure 2.** The ROC curves of 5 models. (A) Predicting severe trauma in the training set. (B) Predicting severe trauma in the internal validation set. (C) Predicting severe trauma in the external validation set. (D) Predicting critical resource use in training set. (E) Predicting critical resource use in the internal validation set. (F) Predicting critical resource use in the external validation set. AUC: area under the curve; GCS: Glasgow Coma Score; PHI: Prehospital Index; pTEST: prehospital triage model to predict severe trauma; ROC: receiver operating characteristic; RTS: revised trauma score.



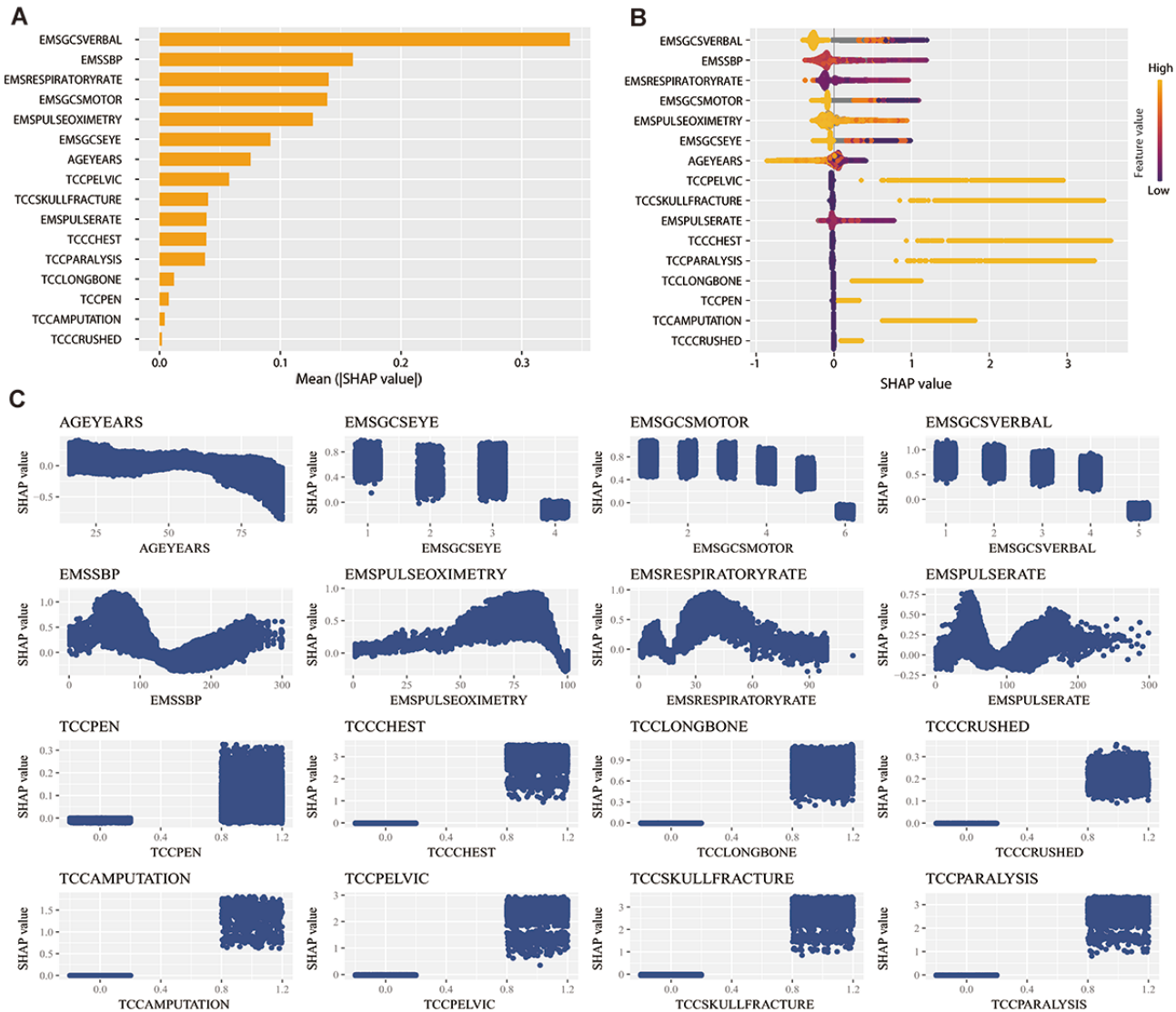
### Model Interpretation

As shown in the SHAP summary plots (Figure 3A), the contributions of the variables to the pTEST model for severe trauma prediction were evaluated using the average absolute SHAP values; the top 5 important variables were EMSGCSVERBAL, EMSSBP, EMSRESPIRATORYRATE, EMSGCSMOTOR, and EMPULSEOXIMETRY. Figure 3B lists the impact of the different variables illustrated by the SHAP values for severe trauma prediction. Figure 3C shows each variable’s SHAP values versus measured values. Figure 3B, C

shows that the higher the EMSGCSVERBAL score, the lower the probability of severe trauma (“negative” impact). Similarly, AGEYEARS, EMSGCSEYE, EMSGCSMOTOR, and EMPULSEOXIMETRY negatively contributed to the predicted probability. In contrast, 8 injury patterns contributed positively to the predicted probability. The SHAP summary and dependence plots of the pTEST model for critical resource use prediction are shown in Multimedia Appendix 17. The personalized feature attributes for 2 representative patients with and without severe trauma in the training set are provided in Multimedia Appendix 18.



**Figure 3.** The global model explanation for predicting severe trauma by the SHAP method in the training set. (A) SHAP summary bar plot of the average SHAP value for each variable. (B) SHAP summary dot plot. In each variable, a dot is made for each single patient, representing the SHAP value of this variable. The colors of the dots demonstrate the actual values of the features, and the dots are stacked vertically to show density. (C) SHAP dependence plot. Each dependence plot shows the association between the actual value and the SHAP value of the variable, and each dot represents a single patient. AGEYEARS: age at the time of injury; EMSGCSEYE: Glasgow Coma Score eye; EMSGCSMOTOR: Glasgow Coma Score motor; EMSGCSVERBAL: Glasgow Coma Score verbal; EMSSBP: systolic blood pressure; EMSPULSEOXIMETRY: oxygen saturation; EMSRESPIRATORYRATE: respiratory rate; EMSPULSERATE: pulse rate; SHAP: Shapley additive explanation analysis; TCCPEN: penetrating injuries; TCCCHEST: chest wall instability; TCCLONGBONE: long-bone fractures; TCCCRUSHED: crushed extremity; TCCAMPUTATION: amputation; TCCPELVIC: pelvic fracture; TCCSKULLFRACTURE: skull fracture; TCCPARALYSIS: paralysis.



### Subgroup Analysis

In [Multimedia Appendix 19](#), subgroup analyses were performed according to age, sex, transport mode, trauma type, and prehospital time. The pTEST model AUCs for severe trauma prediction in patients 60 years of age or older were relatively low at 0.717 (95% CI 0.714-0.720), 0.712 (95% CI 0.708-0.717), and 0.710 (95% CI 0.707-0.714) in the 3 sets, respectively. The AUCs in patients with penetrating injuries were relatively high at 0.815 (95% CI 0.810-0.820), 0.810 (95% CI 0.802-0.817), and 0.810 (95% CI 0.805-0.816) in the 3 sets, respectively. The high proportion of severe trauma in helicopter-transported patients (approximately 41%) led to a high undertriage rate (>0.2) and a low overtriage rate (<0.5). [Multimedia Appendix 20](#) illustrates the pTEST model performance in critical resource use prediction in different subgroups.

### Discussion

#### Principal Findings

In this multisite, large-sample study, we present a prehospital trauma triage tool, pTEST, for severe trauma prediction in EMS. To our knowledge, this is the first study combining ML with national triage guidelines. The pTEST performed optimally in internal and external validations. In addition, its diagnostic accuracy was evaluated using anatomical and resource-based outcomes, and the resource-based outcome is a better alternative to determine the need for specialized trauma care [27]. Furthermore, the pTEST was developed based on national triage guidelines, a globally adopted standard in many organizations. Therefore, the pTEST model can be conveniently applied in EMS practice and may have global relevance.

## Comparison With Previous Studies

Consistent with previous studies [8], we also demonstrated the poor performance of the RED criteria from the national triage guidelines. Globally, all triage guidelines are based on a criteria checklist, including vital signs, injury type, and mechanism of injury [7,28,29]. These guidelines are simplistic—patients meeting any 1 of the criteria should be transported to the highest-level trauma center. In reality, there was an interaction among variables and nonlinear effects of continuous variables (Figure 3C). XGBoost, as a nonlinear ensemble method, can train a more accurate classifier from several weak classifiers and has other benefits, such as dealing with missing values and interaction, avoiding overfitting, and accelerating the training speed by parallel calculation [30]. Our study incorporated XGBoost into national field triage guidelines and developed the pTEST model to enhance performance. The pTEST model included age, 7 vital signs, and 8 injury patterns. We did not perform further variable selection because these variables are the most important in the national field triage guidelines, and the overall number is moderate for field triage. The pTEST model did not achieve an undertriage rate of <5% or an overtriage rate of <35%, as the ACS Committee on Trauma targeted. However, 2 aspects must be noted. First, the definition of the undertriage rate by the ACS is different from that used in our study. For example, the undertriage rate in ACS is equal to the number of patients with ISS  $\geq 16$  transported to a low-level trauma center or nontrauma center divided by the total number of patients transported to a low-level trauma center or nontrauma center [19]. In contrast, the undertriage rate in our study is equal to the number of patients with ISS  $\geq 16$  predicted as ISS <16 divided by the total number of patients predicted ISS <16. In the national field triage guideline, the patients predicted as ISS <16 can be transported to a low-level or nontrauma center. If the patients with predicted ISS <16 and  $\geq 16$  are transferred to low and high-level trauma centers, respectively, then the undertriage rate in our study is the same as that in ACS, but this is unlikely in practice. Second, the under- and overtriage rates are affected by severe trauma proportion. Since most (>90%) of our study population came from level 1 and 2 trauma centers, the severe trauma proportion was approximately 18%, resulting in undertriage rate overestimation and overtriage rate underestimation. Based on previous studies with good sample representativeness [9], we assumed that the proportion of severe trauma was 3%. Keeping the sensitivity (0.794) and specificity (0.5) of the pTEST model and the sample size ( $n=508,703$ ) in the external validation set unchanged, the under- and overtriage rates were 1.26% and 95.3%, respectively, in the external validation set (Multimedia Appendix 20), meeting the ACS undertriage rate target.

The pTEST model and the national field triage guidelines in subgroup analyses were particularly insensitive among older adults [9]. Possible explanations include different physiological responses to injury [31], medication use that potentially worsens injury [32], high prevalence of frailty, and comorbidities [33]. Previous studies have explored elderly-specific triage criteria [15,34,35]. The pTEST model performed better for penetrating traumas. Notably, several studies have found that penetration is a strong severe trauma predictor, and severely penetrated

injured patients are more easily recognized [10]. The undertriage rate was high in patients transported using helicopters. Patients with a high proportion of severe trauma, such as those experiencing large-scale casualties, should be transferred to high-level trauma centers without field triage to reduce the undertriage rate.

Previous studies have reported controversial results regarding the use of ML in medical prediction issues [36]. Overall, in studies with a limited number of predictors, ML does not demonstrate advantages over traditional models, such as logistic regression [37], whereas, for studies with many predictors, advanced ML may have an advantage [38]. A recent review, including 14 studies, demonstrated that the predictive ability of ML-based models was significantly better than that of conventional trauma triage tools for outcomes of mortality, hospitalization, and critical care admission, and XGBoost was the most commonly used ML algorithm [14]. In this study, the relatively large number of predictors and sufficient amount of data tended to favor ML applications. We built the pTEST model using XGBoost but did not evaluate other ML methods. We believe an excellent model can be created using a large sample size, an advanced ML method, and robust hyperparameter tuning. In addition, we minimized the risk of chance findings and overfitting by avoiding exploring other modeling strategies.

## Limitations

This study had some limitations. First, most of our study population were from level 1 and 2 trauma centers, and the proportion of patients with severe trauma (approximately 18%) was significantly higher than that of all prehospital trauma patients (approximately 3%) [9]. However, unlike PPV and NPV, the sensitivity, specificity, and AUC of the pTEST model were not affected by the proportion of severe trauma, and the high sensitivity, specificity, and AUC objectively reflect the pTEST model's good performance. In addition, some emergency resources may be unavailable in low-level trauma and nontrauma centers [39], and samples from high-level trauma centers make it possible to evaluate the pTEST model with critical resource use as the end point. Second, the pTEST model was not developed into a software application, as in other studies [40], because software development requires adaptation to existing information systems in EMS, which is a complex project. However, in the future, EMS providers can develop software based on available data and programs. Third, the 2017-2019 NTDB data followed the 2011 national field triage guidelines, and the latest guidelines have been revised in 2021. An additional "active bleeding" has been added to high-risk trauma types [7]. The new guidelines will take several years to be implemented, and our model must be further validated and updated as necessary.

## Conclusions

We constructed a prehospital triage model, pTEST, to predict severe trauma and achieved an undertriage rate of <10%. Moreover, our study demonstrated that ML is a promising method for enhancing field triage guidelines performance. In the future, we will validate our pTEST model using populations from different countries and casualty backgrounds. In addition,

software must be developed to increase user convenience of the pTEST model in the EMS.

---

### Acknowledgments

The study was funded by grants (2023YFC3107201 and 2023YFC3107202) from the National Key Research and Development Program of China, grant (2022-JCJQ-QT-012) from Young Elite Scientists Sponsorship Program by China Association for Science and Technology, and grant (2022JC011) from Shanghai Emerging Cross Disciplinary Research Project. The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the study; and decision to submit the study for publication. BT had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

---

### Data Availability

The datasets generated and analyzed during this study are available in ACS Trauma Quality Programs.

---

### Authors' Contributions

QC, CW, and BT discussed and developed the study question for this report. QC, YQ, and ZJ conducted the data extraction and statistical analysis, which was validated by BT. All authors were involved in the interpretation of the data and discussed the results. QC wrote the first draft of this paper. All authors agreed on the final draft of this study. The corresponding author has the right to grant on behalf of all authors.

---

### Conflicts of Interest

None declared.

---

### Multimedia Appendix 1

TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) checklist.  
[\[DOCX File , 32 KB-Multimedia Appendix 1\]](#)

---

### Multimedia Appendix 2

Detailed definition of severe trauma and critical resource use.  
[\[DOCX File , 23 KB-Multimedia Appendix 2\]](#)

---

### Multimedia Appendix 3

The candidate variables for model development.  
[\[DOCX File , 23 KB-Multimedia Appendix 3\]](#)

---

### Multimedia Appendix 4

The missing proportion of the candidate variables for model development. (A) The missing proportion in training set. (B) The missing proportion in internal validation set. (C) The missing proportion in external validation set.  
[\[DOCX File , 649 KB-Multimedia Appendix 4\]](#)

---

### Multimedia Appendix 5

The demographic characteristics, vital signs, and injury patterns between nonsevere traumas and severe traumas in training set.  
[\[DOCX File , 40 KB-Multimedia Appendix 5\]](#)

---

### Multimedia Appendix 6

The demographic characteristics, vital signs, and injury patterns between nonsevere traumas and severe traumas in internal validation set.  
[\[DOCX File , 41 KB-Multimedia Appendix 6\]](#)

---

### Multimedia Appendix 7

The demographic characteristics, vital signs, and injury patterns between nonsevere traumas and severe traumas in external validation set.  
[\[DOCX File , 40 KB-Multimedia Appendix 7\]](#)

---

### Multimedia Appendix 8

The demographic characteristics, vital signs, and injury patterns between no critical resource users and critical resource users in training set.

[\[DOCX File , 37 KB-Multimedia Appendix 8\]](#)

---

### Multimedia Appendix 9

The demographic characteristics, vital signs, and injury patterns between no critical resource users and critical resource users in internal validation set.

[\[DOCX File , 37 KB-Multimedia Appendix 9\]](#)

---

### Multimedia Appendix 10

The demographic characteristics, vital signs, and injury patterns between no critical resource users and critical resource users in external validation set.

[\[DOCX File , 37 KB-Multimedia Appendix 10\]](#)

---

### Multimedia Appendix 11

Model performance metrics for predicting critical resource use fixed at a specificity of 0.5.

[\[DOCX File , 25 KB-Multimedia Appendix 11\]](#)

---

### Multimedia Appendix 12

Model performance metrics for predicting severe trauma at the best thresholds with maximum Youden index.

[\[DOCX File , 25 KB-Multimedia Appendix 12\]](#)

---

### Multimedia Appendix 13

Model performance metrics for predicting critical resource use at the best thresholds with maximum Youden index.

[\[DOCX File , 26 KB-Multimedia Appendix 13\]](#)

---

### Multimedia Appendix 14

Waterfall plot for the predicted probability of pTEST model related to the outcome. (A) Predicted probability of severe trauma in training set. (B) Predicted probability of severe trauma in internal validation set. (C) Predicted probability of severe trauma in external validation set. (D) Predicted probability of critical resource use in training set. (E) Predicted probability of critical resource use in internal validation set. (F) Predicted probability of critical resource use in external validation set. pTEST: prehospital triage model to predict severe trauma.

[\[DOCX File , 883 KB-Multimedia Appendix 14\]](#)

---

### Multimedia Appendix 15

Calibration curves of 5 models. (A) Predicting severe trauma in training set. (B) Predicting severe trauma in internal validation set. (C) Predicting severe trauma in external validation set. (D) Predicting critical resource use in training set. (E) Predicting critical resource use in internal validation set. (F) Predicting critical resource use in external validation set.

[\[DOCX File , 300 KB-Multimedia Appendix 15\]](#)

---

### Multimedia Appendix 16

DCA curves of 5 models. (A) Predicting severe trauma in training set. (B) Predicting severe trauma in internal validation set. (C) Predicting severe trauma in external validation set. (D) Predicting critical resource use in training set. (E) Predicting critical resource use in internal validation set. (F) Predicting critical resource use in external validation set. DCA: decision curve analysis.

[\[DOCX File , 496 KB-Multimedia Appendix 16\]](#)

---

### Multimedia Appendix 17

Global model explanation for predicting critical resource use by the SHAP method in training set. (A) SHAP summary bar plot of the average SHAP value for each variable. (B) SHAP summary dot plot. (C) SHAP dependence plot. SHAP: Shapley additive explanation analysis.

[\[DOCX File , 523 KB-Multimedia Appendix 17\]](#)

---

### Multimedia Appendix 18

Local model explanation for predicting severe trauma by the SHAP method in training set. (A) Waterfall plot of risks contributed by each variable for a patient at high risk of severe trauma. (B) Waterfall plot for a patient at low risk of severe trauma. SHAP: Shapley additive explanation analysis.

[\[DOCX File , 310 KB-Multimedia Appendix 18\]](#)

### Multimedia Appendix 19

The performance of the pTEST model for predicting severe trauma in different subgroups. pTEST: prehospital triage model to predict severe trauma.

[\[DOCX File , 31 KB-Multimedia Appendix 19\]](#)

### Multimedia Appendix 20

The performance of the pTEST model for predicting critical resource use in different subgroups. The undertriage and overtriage rates of the pTEST in external validation set with the fixed sensitivity, specificity, and sample size. pTEST: prehospital triage model to predict severe trauma.

[\[DOCX File , 30 KB-Multimedia Appendix 20\]](#)

### References

1. GBD 2019 Diseases Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the global burden of disease study 2019. *Lancet*. 2020;396(10258):1204-1222. [\[FREE Full text\]](#) [doi: [10.1016/S0140-6736\(20\)30925-9](https://doi.org/10.1016/S0140-6736(20)30925-9)] [Medline: [33069326](https://pubmed.ncbi.nlm.nih.gov/33069326/)]
2. Injuries and violence. World Health Organization. 2021. URL: <https://www.who.int/news-room/fact-sheets/detail/injuries-and-violence> [accessed 2024-02-22]
3. Morris RS, Karam BS, Murphy PB, Jenkins P, Milia DJ, Hemmila MR, et al. Field-triage, hospital-triage and triage-assessment: a literature review of the current phases of adult trauma triage. *J Trauma Acute Care Surg*. 2021;90(6):e138-e145. [doi: [10.1097/TA.0000000000003125](https://doi.org/10.1097/TA.0000000000003125)] [Medline: [33605709](https://pubmed.ncbi.nlm.nih.gov/33605709/)]
4. MacKenzie EJ, Rivara FP, Jurkovich GJ, Nathens AB, Frey KP, Egleston BL, et al. A national evaluation of the effect of trauma-center care on mortality. *N Engl J Med*. 2006;354(4):366-378. [doi: [10.1056/NEJMs052049](https://doi.org/10.1056/NEJMs052049)] [Medline: [16436768](https://pubmed.ncbi.nlm.nih.gov/16436768/)]
5. Gianola S, Castellini G, Biffi A, Porcu G, Fabbri A, Ruggieri MP, et al. Italian National Institute of Health guideline working group. Accuracy of pre-hospital triage tools for major trauma: a systematic review with meta-analysis and net clinical benefit. *World J Emerg Surg*. 2021;16(1):31. [\[FREE Full text\]](#) [doi: [10.1186/s13017-021-00372-1](https://doi.org/10.1186/s13017-021-00372-1)] [Medline: [34112209](https://pubmed.ncbi.nlm.nih.gov/34112209/)]
6. Sasser SM, Hunt RC, Faul M, Sugerman D, Pearson WS, Dulski T, et al. Guidelines for field triage of injured patients: recommendations of the national expert panel on field triage, 2011. *MMWR Recomm Rep*. 2012;61(RR-1):1-20. [\[FREE Full text\]](#) [Medline: [22237112](https://pubmed.ncbi.nlm.nih.gov/22237112/)]
7. Newgard CD, Fischer PE, Gestring M, Michaels HN, Jurkovich GJ, Lerner EB, et al. National guideline for the field triage of injured patients: recommendations of the national expert panel on field triage, 2021. *J Trauma Acute Care Surg*. 2022;93(2):e49-e60. [\[FREE Full text\]](#) [doi: [10.1097/TA.0000000000003627](https://doi.org/10.1097/TA.0000000000003627)] [Medline: [35475939](https://pubmed.ncbi.nlm.nih.gov/35475939/)]
8. Lupton JR, Davis-O'Reilly C, Jungbauer RM, Newgard CD, Fallat ME, Brown JB, et al. Under-triage and over-triage using the field triage guidelines for injured patients: a systematic review. *Prehosp Emerg Care*. 2023;27(1):38-45. [doi: [10.1080/10903127.2022.2043963](https://doi.org/10.1080/10903127.2022.2043963)] [Medline: [35191799](https://pubmed.ncbi.nlm.nih.gov/35191799/)]
9. Newgard C, Fu R, Zive D, Rea T, Malveau S, Daya M, et al. Prospective validation of the national field triage guidelines for identifying seriously injured persons. *J Am Coll Surg*. 2016;222(2):146-158.e2. [\[FREE Full text\]](#) [doi: [10.1016/j.jamcollsurg.2015.10.016](https://doi.org/10.1016/j.jamcollsurg.2015.10.016)] [Medline: [26712244](https://pubmed.ncbi.nlm.nih.gov/26712244/)]
10. van Rein EAJ, Houwert R, Gunning A, Lichtveld R, Leenen L, van Heijl M. Accuracy of prehospital triage protocols in selecting severely injured patients: a systematic review. *J Trauma Acute Care Surg*. 2017;83(2):328-339. [doi: [10.1097/TA.0000000000001516](https://doi.org/10.1097/TA.0000000000001516)] [Medline: [28452898](https://pubmed.ncbi.nlm.nih.gov/28452898/)]
11. Malik NS, Chernbumroong S, Xu Y, Vassallo J, Lee J, Bowley DM, et al. The BCD triage sieve outperforms all existing major incident triage tools: comparative analysis using the UK national trauma registry population. *EClinicalMedicine*. 2021;36:100888. [\[FREE Full text\]](#) [doi: [10.1016/j.eclinm.2021.100888](https://doi.org/10.1016/j.eclinm.2021.100888)] [Medline: [34308306](https://pubmed.ncbi.nlm.nih.gov/34308306/)]
12. Bhalla MC, Frey J, Rider C, Nord M, Hegerhorst M. Simple triage algorithm and rapid treatment and sort, assess, lifesaving, interventions, treatment, and transportation mass casualty triage methods for sensitivity, specificity, and predictive values. *Am J Emerg Med*. 2015;33(11):1687-1691. [doi: [10.1016/j.ajem.2015.08.021](https://doi.org/10.1016/j.ajem.2015.08.021)] [Medline: [26349777](https://pubmed.ncbi.nlm.nih.gov/26349777/)]
13. Voskens FJ, van Rein EAJ, van der Sluijs R, Houwert RM, Lichtveld RA, Verleisdonk EJ, et al. Accuracy of prehospital triage in selecting severely injured trauma patients. *JAMA Surg*. 2018;153(4):322-327. [\[FREE Full text\]](#) [doi: [10.1001/jamasurg.2017.4472](https://doi.org/10.1001/jamasurg.2017.4472)] [Medline: [29094144](https://pubmed.ncbi.nlm.nih.gov/29094144/)]



14. Adebayo O, Bhuiyan ZA, Ahmed Z. Exploring the effectiveness of artificial intelligence, machine learning and deep learning in trauma triage: a systematic review and meta-analysis. *Digit Health*. 2023;9:20552076231205736. [FREE Full text] [doi: [10.1177/20552076231205736](https://doi.org/10.1177/20552076231205736)] [Medline: [37822960](https://pubmed.ncbi.nlm.nih.gov/37822960/)]
15. Garwe T, Newgard CD, Stewart K, Wan Y, Cody P, Cutler J, et al. Enhancing utility of interfacility triage guidelines using machine learning: development of the geriatric interfacility trauma triage score. *J Trauma Acute Care Surg*. 2023;94(4):546-553. [FREE Full text] [doi: [10.1097/TA.0000000000003846](https://doi.org/10.1097/TA.0000000000003846)] [Medline: [36404409](https://pubmed.ncbi.nlm.nih.gov/36404409/)]
16. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ*. 2015;350:g7594. [FREE Full text] [doi: [10.1136/bmj.g7594](https://doi.org/10.1136/bmj.g7594)] [Medline: [25569120](https://pubmed.ncbi.nlm.nih.gov/25569120/)]
17. Hashmi ZG, Kaji AH, Nathens AB. Practical guide to surgical data sets: national trauma data bank (NTDB). *JAMA Surg*. 2018;153(9):852-853. [doi: [10.1001/jamasurg.2018.0483](https://doi.org/10.1001/jamasurg.2018.0483)] [Medline: [29617536](https://pubmed.ncbi.nlm.nih.gov/29617536/)]
18. van der Sluijs R, Lokerman RD, Waalwijk JF, de Jongh MAC, Edwards MJR, den Hartog D, et al. Accuracy of pre-hospital trauma triage and field triage decision rules in children (P2-T2 study): an observational study. *Lancet Child Adolesc Health*. 2020;4(4):290-298. [doi: [10.1016/S2352-4642\(19\)30431-6](https://doi.org/10.1016/S2352-4642(19)30431-6)] [Medline: [32014121](https://pubmed.ncbi.nlm.nih.gov/32014121/)]
19. Committee on Trauma. Resources for optimal care of the injured patient 2014. American College of Surgeons. 2014. URL: <https://www.facs.org/media/yu0laoqz/resources-for-optimal-care.pdf> [accessed 2024-09-20]
20. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. 2016. Presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Aug 13-17, 2016; San Francisco, CA. [doi: [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785)]
21. Sharma V, Kulkarni V, Jess E, Gilani F, Eurich D, Simpson SH, et al. Development and validation of a machine learning model to estimate risk of adverse outcomes within 30 days of opioid dispensation. *JAMA Netw Open*. 2022;5(12):e2248559. [FREE Full text] [doi: [10.1001/jamanetworkopen.2022.48559](https://doi.org/10.1001/jamanetworkopen.2022.48559)] [Medline: [36574245](https://pubmed.ncbi.nlm.nih.gov/36574245/)]
22. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell*. 2020;2(1):56-67. [FREE Full text] [doi: [10.1038/s42256-019-0138-9](https://doi.org/10.1038/s42256-019-0138-9)] [Medline: [32607472](https://pubmed.ncbi.nlm.nih.gov/32607472/)]
23. Fahmy AS, Csecs I, Arafati A, Assana S, Yankama TT, Al-Otaibi T, et al. An explainable machine learning approach reveals prognostic significance of right ventricular dysfunction in nonischemic cardiomyopathy. *JACC Cardiovasc Imaging*. 2022;15(5):766-779. [FREE Full text] [doi: [10.1016/j.jcmg.2021.11.029](https://doi.org/10.1016/j.jcmg.2021.11.029)] [Medline: [35033500](https://pubmed.ncbi.nlm.nih.gov/35033500/)]
24. Riley RD, Ensor J, Snell KIE, Harrell FE, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ*. 2020;368:m441. [doi: [10.1136/bmj.m441](https://doi.org/10.1136/bmj.m441)] [Medline: [32188600](https://pubmed.ncbi.nlm.nih.gov/32188600/)]
25. He J, Wang B, Tao J, Liu Q, Peng M, Xiong S, et al. Accurate classification of pulmonary nodules by a combined model of clinical, imaging, and cell-free DNA methylation biomarkers: a model development and external validation study. *Lancet Digit Health*. 2023;5(10):e647-e656. [FREE Full text] [doi: [10.1016/S2589-7500\(23\)00125-5](https://doi.org/10.1016/S2589-7500(23)00125-5)] [Medline: [37567793](https://pubmed.ncbi.nlm.nih.gov/37567793/)]
26. Maxim LD, Niebo R, Utell MJ. Screening tests: a review with examples. *Inhal Toxicol*. 2014;26(13):811-828. [FREE Full text] [doi: [10.3109/08958378.2014.955932](https://doi.org/10.3109/08958378.2014.955932)] [Medline: [25264934](https://pubmed.ncbi.nlm.nih.gov/25264934/)]
27. Lerner EB, Cushman JT, Drendel AL, Badawy M, Shah MN, Guse CE, et al. Effect of the 2011 revisions to the field triage guidelines on under- and over-triage rates for pediatric trauma patients. *Prehosp Emerg Care*. 2017;21(4):456-460. [doi: [10.1080/10903127.2017.1300717](https://doi.org/10.1080/10903127.2017.1300717)] [Medline: [28489471](https://pubmed.ncbi.nlm.nih.gov/28489471/)]
28. Hamada S, Gauss T, Duchateau F, Truchot J, Harrois A, Raux M, et al. Evaluation of the performance of French physician-staffed emergency medical service in the triage of major trauma patients. *J Trauma Acute Care Surg*. 2014;76(6):1476-1483. [doi: [10.1097/TA.0000000000000239](https://doi.org/10.1097/TA.0000000000000239)] [Medline: [24854319](https://pubmed.ncbi.nlm.nih.gov/24854319/)]
29. Dinh MM, Bein KJ, Oliver M, Veillard AS, Ivers R. Refining the trauma triage algorithm at an Australian major trauma centre: derivation and internal validation of a triage risk score. *Eur J Trauma Emerg Surg*. 2014;40(1):67-74. [doi: [10.1007/s00068-013-0315-1](https://doi.org/10.1007/s00068-013-0315-1)] [Medline: [26815779](https://pubmed.ncbi.nlm.nih.gov/26815779/)]
30. Dupont T, Kentish-Barnes N, Pochard F, Duchesnay E, Azoulay E. Prediction of post-traumatic stress disorder in family members of ICU patients: a machine learning approach. *Intensive Care Med*. 2024;50(1):114-124. [doi: [10.1007/s00134-023-07288-1](https://doi.org/10.1007/s00134-023-07288-1)] [Medline: [38112774](https://pubmed.ncbi.nlm.nih.gov/38112774/)]
31. Heffernan DS, Thakkar RK, Monaghan SF, Ravindran R, Adams JC, Kozloff MS, et al. Normal presenting vital signs are unreliable in geriatric blunt trauma victims. *J Trauma*. 2010;69(4):813-820. [doi: [10.1097/TA.0b013e3181f41af8](https://doi.org/10.1097/TA.0b013e3181f41af8)] [Medline: [20938267](https://pubmed.ncbi.nlm.nih.gov/20938267/)]
32. Nishijima DK, Shahlaie K, Sarkar K, Rudisill N, Holmes JF. Risk of unfavorable long-term outcome in older adults with traumatic intracranial hemorrhage and anticoagulant or antiplatelet use. *Am J Emerg Med*. 2013;31(8):1244-1247. [FREE Full text] [doi: [10.1016/j.ajem.2013.04.035](https://doi.org/10.1016/j.ajem.2013.04.035)] [Medline: [23759685](https://pubmed.ncbi.nlm.nih.gov/23759685/)]
33. Galimberti S, Graziano F, Maas A, Isernia G, Lecky F, Jain S, et al. Effect of frailty on 6-month outcome after traumatic brain injury: a multicentre cohort study with external validation. *Lancet Neurol*. 2022;21(2):153-162. [doi: [10.1016/S1474-4422\(21\)00374-4](https://doi.org/10.1016/S1474-4422(21)00374-4)] [Medline: [35065038](https://pubmed.ncbi.nlm.nih.gov/35065038/)]
34. Ichwan B, Darbha S, Shah MN, Thompson L, Evans DC, Boulger CT, et al. Geriatric-specific triage criteria are more sensitive than standard adult criteria in identifying need for trauma center care in injured older adults. *Ann Emerg Med*. 2015;65(1):92-100.e3. [doi: [10.1016/j.annemergmed.2014.04.019](https://doi.org/10.1016/j.annemergmed.2014.04.019)] [Medline: [24908590](https://pubmed.ncbi.nlm.nih.gov/24908590/)]

35. Newgard CD, Holmes JF, Haukoos JS, Bulger EM, Staudenmayer K, Wittwer L, et al. Improving early identification of the high-risk elderly trauma patient by emergency medical services. *Injury*. 2016;47(1):19-25. [FREE Full text] [doi: [10.1016/j.injury.2015.09.010](https://doi.org/10.1016/j.injury.2015.09.010)] [Medline: [26477345](https://pubmed.ncbi.nlm.nih.gov/26477345/)]
36. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. 2019;110:12-22. [doi: [10.1016/j.jclinepi.2019.02.004](https://doi.org/10.1016/j.jclinepi.2019.02.004)] [Medline: [30763612](https://pubmed.ncbi.nlm.nih.gov/30763612/)]
37. Gravesteyn BY, Nieboer D, Ercole A, Lingsma HF, Nelson D, van Calster B, et al. Machine learning algorithms performed no better than regression models for prognostication in traumatic brain injury. *J Clin Epidemiol*. 2020;122:95-107. [FREE Full text] [doi: [10.1016/j.jclinepi.2020.03.005](https://doi.org/10.1016/j.jclinepi.2020.03.005)] [Medline: [32201256](https://pubmed.ncbi.nlm.nih.gov/32201256/)]
38. Larsson A, Berg J, Gellerfors M, Gerdin Wärnberg M. The advanced machine learner XGBoost did not reduce prehospital trauma mistriage compared with logistic regression: a simulation study. *BMC Med Inform Decis Mak*. 2021;21(1):192. [FREE Full text] [doi: [10.1186/s12911-021-01558-y](https://doi.org/10.1186/s12911-021-01558-y)] [Medline: [34148560](https://pubmed.ncbi.nlm.nih.gov/34148560/)]
39. Soto JM, Zhang Y, Huang JH, Feng DX. An overview of the American trauma system. *Chin J Traumatol*. 2018;21(2):77-79. [FREE Full text] [doi: [10.1016/j.cjte.2018.01.003](https://doi.org/10.1016/j.cjte.2018.01.003)] [Medline: [29605432](https://pubmed.ncbi.nlm.nih.gov/29605432/)]
40. van Rein EAJ, van der Sluijs R, Voskens FJ, Lansink KWW, Houwert RM, Lichtveld RA, et al. Development and validation of a prediction model for prehospital triage of trauma patients. *JAMA Surg*. 2019;154(5):421-429. [FREE Full text] [doi: [10.1001/jamasurg.2018.4752](https://doi.org/10.1001/jamasurg.2018.4752)] [Medline: [30725101](https://pubmed.ncbi.nlm.nih.gov/30725101/)]

## Abbreviations

**ACS-COT:** American College of Surgeons Committee on Trauma

**AUC:** area under the curve

**ED:** emergency department

**EMS:** emergency medical service

**GCS:** Glasgow Coma Score

**ISS:** Injury Severity Score

**ML:** machine learning

**NPV:** negative predictive value

**NTDB:** National Trauma Data Bank

**PHI:** Prehospital Index

**PPV:** positive predictive value

**pTEST:** prehospital triage model to predict severe trauma

**RTS:** revised trauma score

**SHAP:** Shapley additive explanation analysis

**TRIPOD:** Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis

**XGBoost:** extreme gradient boosting model

*Edited by T de Azevedo Cardoso; submitted 24.03.24; peer-reviewed by L Zhang, D Peng; comments to author 19.04.24; revised version received 11.05.24; accepted 07.08.24; published 30.09.24*

*Please cite as:*

*Chen Q, Qin Y, Jin Z, Zhao X, He J, Wu C, Tang B*

*Enhancing Performance of the National Field Triage Guidelines Using Machine Learning: Development of a Prehospital Triage Model to Predict Severe Trauma*

*J Med Internet Res 2024;26:e58740*

URL: <https://www.jmir.org/2024/1/e58740>

doi: [10.2196/58740](https://doi.org/10.2196/58740)

PMID:

©Qi Chen, Yuchen Qin, Zhichao Jin, Xinxin Zhao, Jia He, Cheng Wu, Bihan Tang. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org/>), 30.09.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.