Published in final edited form as:

J Surg Res. 2014 April ; 187(2): 371–376. doi:10.1016/j.jss.2013.06.037.

Pre-Hospital Triage of Trauma Patients Using the Random Forest Computer Algorithm

Michelle Scerbo, MD, Hari Radhakrishnan, MD, Bryan Cotton, MD, MPH, Anahita Dua, MD, Deborah Del Junco, PhD, Charles Wade, PhD, and John B. Holcomb, MD

Center for Translational Injury Research (CeTIR), Department of Surgery, University of Texas-Houston, Houston, TX

Abstract

Background—Over-triage not only wastes resources but displaces the patient from their community and causes delay of treatment for the more seriously injured. This study aimed to validate the Random Forest computer model (RFM) as means of better triaging trauma patients to Level I trauma centers.

Methods—Adult trauma patients with “medium activation” presenting via helicopter to a Level I Trauma Center from May 2007 to May 2009 were included. The “medium activation” trauma patient is alert and hemodynamically stable on scene but has either subnormal vital signs or an accumulation of risk factors that may indicate a potentially serious injury. Variables included in the RFM computer analysis including demographics, mechanism of injury, pre-hospital fluid, medications, vitals, and disposition.

Statistical analysis was performed via the Random Forest Algorithm to compare our institutional triage rate to rates determined by the RFM.

Results—A total of 1,653 patients were included in this study of which 496 were used in the testing set of the RFM. In our testing set, 33.8% of patients brought to our Level I trauma center could have been managed at a Level III trauma center and 88% of patients that required a Level I trauma center were identified correctly. In the testing set, there was an over-triage rate of 66% while utilizing the RFM we decreased the over-triage rate to 42% ($p < 0.001$). There was an under-triage rate of 8.3%.

The RFM predicted patient disposition with a sensitivity of 89%, specificity of 42%, negative predictive value of 92% and positive predictive value of 34%.

© 2013 Elsevier Inc. All rights reserved.

Corresponding author: John B. Holcomb, MD, Professor and Vice Chair of the Department of Surgery, Chief of the Division of Acute Care Surgery, Director for CeTIR, University of Texas-Houston, 6413 Fannin Street, MSB 4.170, Houston, TX 77030, Phone: (713) 500-7218, Fax: (713) 500-7213, john.holcomb@uth.tmc.edu.

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

Presented at the 8th Annual Academic Surgical Congress, New Orleans, LA, February 5-7, 2013

No conflicts of interest

Conclusion—While prospective validation is required, it appears that computer modeling potentially could be used to guide triage decisions, allowing both more accurate triage and more efficient use of the trauma system.

Keywords

Random Forest Model; Triage; Trauma; Over-triage; Under-triage; Pre-hospital care

Introduction

Effective triage of trauma patients is critical for efficient utilization of trauma system resources. Over-triage results in the delay of treatment for the more seriously injured, an excessive burden on the trauma center and its staff, an inappropriate use of expensive and limited resources, and the unnecessary displacement of patients from their communities.¹

The quality of pre-hospital care impacts patient outcome.²⁻⁴ This includes not only appropriate management, resuscitation and rapid transport to a hospital, but also transport to the hospital best suited to manage particular injuries. Established in 1999, the accepted over-triage rate of 50% has been re-evaluated but never successfully reduced.³ This high rate of over-triage has led to a crowding of Level I regional trauma centers across the nation at the cost of utilizing expensive and dangerous transport and highly trained staff for patients that do not benefit medically.⁵ Efficient resource management requires emergency medical services (EMS) personnel to correctly triage patients to the appropriate trauma center.

Currently triage is determined based on three domains: physiology, mechanism of injury and anatomical location of injury. These domains are defined during the initial physical exam in the pre-hospital environment and are recorded at intervals throughout transport. None of these domains have been able to accurately predict major trauma, the need for trauma team activation or the necessity of a Level I trauma center, especially in the "medium activation" population.⁶⁻⁹ The "medium activation" trauma patient is alert and hemodynamically stable on scene but has either subnormal vital signs or an accumulation of risk factors that may indicate a potentially serious injury. The criterion used for this classification at our center is outlined in Table I.

Computer models can be utilized to assist with medical decision making and are becoming more common in clinical use.¹⁰ The Random Forest Model (RFM) is an ensemble classifier that uses a combination of many decision trees. The decision trees are created using a labeled training set of data associated with each patient. As the RFM receives more information, it creates more trees to avoid overfitting, or the generation of a single decision tree that depends too much on irrelevant features. Class assignment in the testing set is determined by the number of votes from all trees. Each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. As trees become larger, the generalization error for forests converges to a limit. Advantages of RFM include its ability to manage large databases with multiple weak input variables, maintain effectiveness even with large amounts of missing data through accurate estimation and the generation of an internal unbiased estimate of the generalization error as

the forest building progresses.¹¹ These properties permit the RFM to function as a “learning algorithm.”

This study aimed to create and validate the RFM as a tool to triage minimally injured trauma patients away from Level I centers using pre-hospital variables.

Materials and Methods

Adult trauma patients with “medium activation” presenting via helicopter to a single Level I Trauma Center within a 3-tiered triage system from May 2007 to May 2009 were included. Transferred patients, patients with burns as a major complaint and patients under 18 years old were excluded. Patients that arrived with the “highest activation”, classified as top tier in our triage system, were not included in this analysis. Variables included in the RFM were demographics, mechanism of injury, pre-hospital fluid, medications, vitals, and disposition. The selection of patients for study is displayed in Figure 1.

Pre-hospital data was collected from a hand-written “run-sheet” used by Life Flight either by accessing the physical paper files or viewing a scanned PDF of the paper file on Sovera Health Information Management (Healthcare Solutions Group, a division of CGI Group, Inc. Montreal, Quebec). The ED data was collected from the ED Nursing Record, accessed via Care4 (PowerChart 2010.11.1.30, Cerner Corporation. Kansas, City, Missouri), the electronic medical record.

The vital sign variables included in the RFM were systolic blood pressure (SBP), diastolic blood pressure (DBP), heart rate (HR), peripheral oxygen saturation (SpO₂), respiratory rate (RR), and Glasgow Coma Scale, which was collected as the individual components of the scale: Eye (GCS-E), Voice (GCS-V) and Motor (GCS-M), as well as the total of all three components (GCS). Other variables that were retrospectively calculated included mean arterial pressure (MAP), calculated as the sum of two-thirds DBP and one-third SBP, pulse pressure (PP), the difference between SBP and DBP, and shock index (SI), the ratio between HR and SBP. Other pre-hospital variables included in the model were analgesic utilization and crystalloid administration.

Patients were grouped by their disposition: ED discharge/lower level admission or upper level admission. “Over-triage” was defined as patients that were either discharged from the ED or admitted to a lower level unit, as these patients did not utilize the resources unique to a Level I trauma center. Actual patient disposition was compared to the predictive disposition of the RFM to determine sensitivity, specificity, positive predicted value, and negative predicted value of the RFM.

Statistical analysis of the data was performed using Microsoft Excel 2007 and Weka 3²⁰ in Windows 7 and Linux 10.10, respectively. The null hypothesis was rejected at $p < 0.05$. Univariate analysis was performed using a two-tailed unpaired Student's t-test on each collected variable within the two groups.

Random Forest Model

Our patient population was classified into two categories of admission: upper level admission (patients that required level I trauma center care) and ED discharge/lower level admission (patients that did not require level I trauma center care) based on admission to one of seventeen places at our institution or discharge (Table II).

The classifier ED discharge/lower level admission included 83 attributes relating to demographics, injury data such as mechanism of injury, incident details, triage accuracy, patient management characteristics (analgesia, crystalloid administration), bleeding status, pulse character, anatomical site of injury, type of injury at that anatomical site, range of vital signs as expressed by minimum and maximum values and intensive/intermediate care unit disposition. Sixteen hundred and fifty three patients were included in the RFM, 70% (1157) were used for training and 30% (496) were used for testing the model.

One input variable was used to determine the decision at a node of the tree and the depth for each tree was set at five. The Random Forest Algorithm selected seven variables for each tree, creating 399 additional trees, each with unique attributes. Two thirds of the training set was passed through each tree and was internally validated using the remaining one third of the training set. The final model was externally validated using the same set of variables on a naïve data set.

The patient disposition prediction from the computer model was compared against the actual disposition in order to determine the accuracy of this algorithm.

Results

Demographic and Injury Data

A total of 1,653 patients were included in our study to train and test the RFM after excluding 453 patients (patients < 18 years, burn injury necessitating transfer to burn center, missing records, transfers and non-trauma patients). From this cohort, 999 were admitted to hospital of which 459 (28%) were classified as upper level and 1194 (72%) were classified as lower level (540) or discharged (654) (Figure I).

The demographics and types of injuries of the study cohort are displayed in Table II. Patients admitted to a lower level unit were younger (37 ± 15 versus 43 ± 16 , $p < 0.001$). There was a lower percentage of patients greater than 65 years of age in the lower level admission group (5% versus 13%, $p < 0.001$).

Out of our entire patient cohort, only 459 (28%) were found to require an upper level admission, with the remaining 1194 (72%) either being discharged or admitted to a lower level unit. This is 22% higher than the accepted 50% over-triage rate put forth by the ACS-COT.³

Vital Signs

There were no differences noted in the measured pre-hospital values of DBP, HR, SpO₂, RR, GCS-E, GCS-V, GCS-M, GCS, MAP, PP, SI and while patients discharged from the

ED or admitted to a lower level unit had a higher average SBP during their pre-hospital course (138 ± 0.9 versus 134 ± 0.5); this SBP was well within normal range hence deciphering which patients require higher levels of care is difficult.

Random Forest Modeling

Out of the included patients of 1653, the RFM used 70% for the training set and 30% as the testing set. Of the patient subset utilized for testing purposes of the RFM (496 patients – 168 upper level, 328 ED discharge/lower level), there was an over-triage rate of 66%.

In our testing patient cohort, 33.8% of patients brought to our level I center could have been managed at a level III and 88% of patients that required a level I center were identified correctly. Our actual over-triage rate in the testing population was 66% and utilization of the RFM would have reduced this rate to 42%. There was an under-triage rate of 8.3% using the RFM as 14 patients were incorrectly classified as not requiring a level I trauma center for care.

The RFM predicted patient disposition with a sensitivity of 89%, specificity of 42%, negative predictive value of 92% and positive predictive value of 34%.

In the testing set, there was an over-triage rate of 66% while utilizing the RFM we decreased the over-triage rate to 42% ($p < 0.001$). Overall, in our patient group of 1653 patients, we had an over-triage rate of 72% while the RFM had an over-triage rate of 50%.

Discussion

Appropriate triage of trauma patients is vital for efficient utilization of trauma system resources and deliverance of appropriate care. While under-triage can have devastating consequences, over-triage can be equally problematic by forcing patients out of their community unnecessarily, wasting resources, and delays in treatment for those critically injured. Accurately triaging patients who are not critically injured is difficult based on the limited set of data that is collected and the fact that typical identifiers of critically injured patients that assist with deciphering which patients require level I care are not different enough between groups to assist with decision making.

Prior attempts have been made to predict hospital admission from pre-hospital data.¹²⁻¹⁴ The majority of these attempts were in the highest level of injured trauma patients, which is not applicable to our patient population. In this minimally injured population, typical deciphering variables such as vitals, type of injury, and anatomical injury location are not significantly different; our data on pre-hospital vitals showed no significant differences between vital signs and the small difference between SBP (138 versus 134) was negligible given that both those values are well within normal range and hence not helpful for triage.

Patients that arrive with “medium activation” have presumed moderate injuries, and are categorized based on demographics, moderately stable vital signs, anatomical location and mechanism of injury. These patients were in a significant enough accident, such as a motor vehicle crash that was responsible for the death of other occupants or a fall from a significant height, which could potentially cause more severe injuries. It is the accumulation

of risk factors that even makes these patients eligible to be seen at a Level I, but these risk factors are neither accurate nor precise enough predictors of severe injury.

The decision to evaluate our data set using a RFM was based on several factors. Our data set was dense, involving nearly one million variables. It was difficult to interpret the value of the results that we generated using logistic regression modeling. Furthermore, while logistic regression modeling is successful in discerning differences between groups, it cannot adapt or be manipulated to error on the side of caution, for example, over-triage instead of under-triage patients. This ability to learn/train based on multiple variable input is unique to the RFM.

As a result our institution had an overall over-triage rate of 72% and within our testing set an over-triage rate of 66%. The RFM had an over-triage rate of 42% in the testing set and 50% overall both better than our current practice. This was a significant difference; utilization of the RFM could potentially decrease our over-triage rate by 24%. The RFM had an under-triage rate of 8% as 14 patients in the testing set were misclassified as not needing a level I. However, 216 patients avoided being misclassified as requiring the level I center and potentially could have not only saved resources but decreased hospital cost significantly if appropriately triaged. Prospective, continuous collection of a larger cohort will be necessary to formulate more acceptable models of triage decision-making.

Limitations

Our modeled relied on manually recorded data at inconsistent intervals that is written down in a less than ideal, sub-austere environment. Consequently, the validity of such data cannot be verified. It would be optimal to have continuously, electronically recorded data. The medical monitors that are currently used by our Life Flight helicopter service in the pre-hospital environment have the ability to display and also record data continuously, but this feature is not utilized. This would result in a higher resolution of the data that is already available and could potentially assist to interpret autonomic variability not representative of the true trajectory of the patient, for instance as influenced by pain with movement of an affected limb or analgesic administration, a hurdle to interpreting the hemodynamic variability of a patient who is also being treated and manipulated. Second, it reflects a population of one urban area by a single method of transport. Other factors influence the modality of transport used such as availability of vehicles, traffic and geographic location of the patient. We recognize that our EMS personnel may collect different pre-hospital variables from other populations and that our population may have a different distribution of injury mechanism from other areas. Our decision to only include helicopter transport was based on several factors. Our over-triage rate increased during this time period as a consequence of a trauma center 50 miles away closing after a national disaster. The majority of this influx of patients arrived via helicopter as they came from areas where ground transport would be inappropriate. Using a standard pre-hospital record also decreased inconsistencies in data collection and decreased the likelihood for incomplete data collection. Helicopter transport also possesses the ability to rapidly change destination without impacting transport time, which would be difficult for a vehicle on the ground to

accomplish. If validated, utilization of the RFM could change destination en route to improve triage.

The third major limitation is that the RFM did under-triage 14 patients. Our true under-triage rates were not assessed in this study, making it difficult to comment on the ability of RFM to improve or worsen our actual under-triage rate.

Another potential limitation is the time period in which this study was undertaken. In 2008, because of Hurricane Ike, our institution, (one of the two remaining adult Level I academic trauma centers in the region), experienced a peak trauma diversion rate of 40% and a 12 month average of 15%. Hence it is difficult to determine whether the RFM should be utilized only in diversion/crisis situations on a daily basis to prevent over-triage. This could also be strength of our model given that in times of crisis, it would be utilized to ensure that the patients who genuinely require level I care are given priority based on a triage method that is valid and accurate.

Conclusion

Pre-hospital data can be used to appropriately triage trauma patients using the RFM. There is a significant decrease in over-triage with utilization of the RFM. While prospective validation is required, it appears that computer modeling can be used to guide triage decisions allowing both more accurate triage and more efficient use of the trauma system, without negatively affecting patient outcomes.

Acknowledgments

Funding: Alpha Omega Alpha 2010 Carolyn L. Kuckein Student Research Fellowship

References

1. Mackersie RC. History of Trauma Field Triage Development and the American College of Surgeons Criteria. *Prehosp Emerg Care*. 2006; 10:287–94. [PubMed: 16801263]
2. Hannan EL, Farrell LS, Cooper A, Henry M, Simon B. Physiologic Trauma Triage Criteria in Adult Trauma Patients: Are they effective in saving lives by transporting patients to trauma centers? *J Am Coll Surg*. 2005; 200:584–92. [PubMed: 15804473]
3. American College of Surgeons Committee on Trauma. Resources for Optimal Care of the Injured Patient. ACS; Chicago: 2006.
4. MacKenzie EJ, Rivara FP, Jurkovich GJ, Nathens AB, Frey KP, Egleston BL, Salkaever DS, Scharfstein DO. A National Evaluation of the Effect of Trauma-Center Care on Mortality. *NEJM*. 2006; 354:366–78. [PubMed: 16436768]
5. Chen L, Reisner AT, Gribok A, Reifman J. Exploration of Prehospital Vital Sign Trends for the Prediction of Trauma Outcomes. *Prehosp Emerg Care*. 2009; 13:286–94. [PubMed: 19499463]
6. Boyle MJ. Is mechanism of injury alone in the prehospital setting a predictor of major trauma - a review of the literature. *J Trauma Manag Outcomes*. 2007; 1:4. [PubMed: 18271994]
7. Newgard CD, Hedges JR, Diggs B, Mullins RJ. Establishing the need for trauma center care: anatomic injury or resource use? *Prehosp Emerg Care*. 2008; 12:451–8. [PubMed: 18924008]
8. Holcomb JB, Salinas J, McManus JM, Miller CC, Cooke WH, Convertino V A. Manual vital signs reliably predict need for life-saving interventions in trauma patients. *J Trauma*. 2005; 59:821–8. [PubMed: 16374268]

9. Lehmann RK, Arthurs ZM, Cuadrado DG, Casey LE, Beekley AC, Martin MJ. Trauma team activation: simplified criteria safely reduces overtriage. *Am J Surg.* May. 2007; 193(5):630–4. discussion 634–5. [PubMed: 17434371]
10. Fernández-Blanco E, Aguiar-Pulido V, Munteanu CR, Dorado J. Random Forest classification based on star graph topological indices for antioxidant proteins. *J Theor Biol.* Jan 21. 2013; 317:331–7. [PubMed: 23116665]
11. Breiman L. *Forests Random. Machine Learning.* 2001; 45:5–32.
12. Champion HR, Sacco WJ, Carnazzo AJ, Copes W, Fouty WJ. Trauma score. *Crit Care Med.* 1981; 9:672–6. [PubMed: 7273818]
13. Champion HR, Sacco WJ, Copes WS, Gann DS, Gennarelli TA, Flanagan ME. A revision of the Trauma Score. *J Trauma.* 1989; 29:623–9. [PubMed: 2657085]
14. Koehler JJ, Baer LJ, Malafa SA, et al. Prehospital Index: a scoring system for field triage of trauma victims. *Ann Emerg Med.* 1986; 15:178. [PubMed: 3946860]

Figure I.
Breakdown of patients based on classification

Table I

Criteria for “highest activation” or “medium activation” at our Level I trauma center.

	“Highest Activation”	“Medium Activation”
Physiologic Criteria		
GCS	10	> 10, 14
Heart Rate	> 120	110-120
Systolic Pressure	90	> 90
Respiratory Rate	< 10, > 29	Not specified
Intubated	Yes	No
Anatomic Criteria		
Penetrating Injury	Any to torso, groin, head or neck	To extremity
Amputation	Proximal to ankle or wrist	None Specified
Sensory Deficit	Paraplegia, Quadriplegia	None Specified
Hemorrhage	Uncontrolled external	None Specified
Fracture	Pelvic, Two or more long bone	None Specified
Trauma with Burns	20% Body Surface Area	10-20% Body Surface Area
Risk Factors		
Extrication	None Specified	Any patient requiring extrication
Intrusion Depth	None Specified	Into a passenger space of a motor vehicle of > 12 inches
Ejection	None Specified	From an enclosed vehicle, or from motorcycle >20 MPH
Pregnancy	None Specified	> 20 weeks
Death of occupants	None Specified	In the same motor vehicle
Auto vs. pedestrian	None Specified	Any injury
Age	None Specified	> 65 years
Fall	None Specified	> 15 feet
Transfer	None Specified	Receiving blood to maintain vital signs
Respiratory	None Specified	Compromise/obstruction

Table II

Admission Categories.

Upper level Admission	ED Discharge/Lower Level Admission
Neurosurgery Trauma Intensive Care Unit	Surgical Intermediate Care Unit
Neurosurgery Intermediate Care Unit	Medical Intermediate Care Unit
Shock Trauma Intensive Care Unit	Orthopedics Floor
Neighboring Level I Trauma Center	Surgical Floor
Medical Intensive Care Unit	Clinical Observation Unit
	Cardiac Care Unit
	Burn Unit *
	Neurology Floor
	Medicine Telemetry Unit
	Geriatrics Medicine Floor
	Medicine Floor
	Discharge from Emergency Department
	Left Against Medical Advice

* Patients admitted to the burn unit as overflow from surgical floor or surgical intermediate care unit

Table III

Demographics of study cohort.

Variable	Discharge/Lower Level Admission	Upper Level Admission	p value
Age	37.5 ± 15.6	43.4 ± 17.9	<0.001
>65 years of age (%)	5.44	13.9	<0.001
Gender			
Male (%)	70.0	69.3	0.83
Female (%)	31.3	30.8	
Mechanism of Injury			
Blunt Injury (%)	95.0	96.5	0.18
Penetrating Injury (%)	2.76	1.52	0.08
Head Injury (%)	1.92	1.96	0.96
Type of Injury			
MVC (%)	50.1	51.2	0.69
MCC (%)	15.5	18.3	0.16
Fall (%)	15.6	15.5	0.96
Autoped (%)	3.77	3.92	0.88
Other (%) [*]	15.0	11.1	0.05
Minimally injured (%)	71.9	80.4	<0.001
# Minimally injured criteria met	1.11 ± 0.923	1.36 ± 0.982	<0.001
Documented LOC (%)	20.3	9.73	<0.01

Values are expressed as average ± standard deviation unless specified as percentage of a cohort.

^{*} includes: Assault, Stab/Impalement, GSW, Crush, unclassified