SuperLearner versus Clinicians to Prioritise Trauma Patients (working title)

Draft version 1.0.0

Ludvig Wärnberg Gerdin[1], Name2 Surname[2], Name3 Surname[2], Name4 Surname[2], Name5 Surname[2], Name6 Surname[2], Martin Gerdin Wärnberg[1*¤]

1 Global Health: Health Systems and Policy, Department of Public Health Sciences, Karolinska Institutet, Stockholm, Sweden
2 Affiliation Dept/Program/Center, Institution Name, City, State, Country

¤Current Address: Martin Gerdin Wärnberg, Department of Public Health Sciences, Karolinska Institutet, 171 77 Stockholm, Sweden
* martin.gerdin@ki.se

Abstract

Remains to be written.

Author Summary

Remains to be written.

Introduction

Trauma is a major threat to population health globally [1, 2]. Every year about 4.6 million people die because of trauma - a number that exceeds the total number of yearly deaths from HIV/AIDS, malaria and tuberculosis combined. The most common cause of trauma is road injuries and in 2016 an estimated 1.3 million people died from such trauma alone [2]. Global actors have vowed to try to halve the number of deaths from road trauma by 2020, but this sustainable development goal is far from being realised [3]. This situation calls for not only more action, but also more research on what works and does not work in trauma care.

Trauma care is highly time sensitive [4]. Early identification of potentially fatal injuries and conditions is crucial for survival. A key component of trauma care is therefore the process of prioritising patients to match level of care with clinical acuity [5, 6]. The literature on how to prioritise trauma patients focuses largely on two issues. First, in the prehospital setting the main focus has been to idenfity patients who should be transferred to a trauma centre [7]. In the hospital setting a substantial body of research has focused on when to activate the trauma team [10, 12].

Although both these issues are important, clinicians all over the world are on a daily basis faced with the more complex problem of how to decide in what order to assess and treat trauma patients that arrive to the emergency department (ED) without any pre-notification. In health systems with formalised criteria for prioritising ED patients, all patients are assigned a priority coupled with a target time to treat. These priorities are commonly coded using numbers or colors, for example red, orange, yellow and green, with red being assigned to the most urgent patients and green to the least urgent [13].

In contrast to the trauma centre transfer or trauma team activation issues the "how" in how to prioritise among trauma patients in the ED has received little attention from the research community. Framed as a classification problem this "how" can be approached using a statistical learner. Logistic

or proportional hazards models are common classification learners whereas more modern alternatives include random forests or convolutional neural networks. These learners all exist along the machine learning spectrum governed by their relative "human-to-machine decision-making-effort", with regression learners in the more-human-than-machine (MHTM) end and networks in the other [14].

The application of more-machine-than-human (MMTH) learners to solve classification problems in medicine is not new [15], but the uptake and use of such learners in trauma research has been slow [16]. Some studies have approached the trauma centre transfer and trauma team activation issues using MMTH learners, and the results are conflicting with regards to the superiority of such learners over MHTM learners or standard criteria [17–20]. One very recent study used a random forest learner to assign priority to patients in a general ED population, and found a slight performance improvement using this MMTH learner compared to the standard criteria [21].

Thus, there seems to be a void of reseach on how to leverage machine learning to prioritise among trauma patients in the ED. Therefore, we set out to conduct a benchmark study, in which we attempted to improve on what we considered the two most important limitations of previous related research, namely the use of retrospective data and the focus on one specific MHTM or MMTH learner. Hence, we aimed to compare the performance of an ensemble machine learning methodology called SuperLearner to that of clinicians based on patients' clinical gestal. Our hypothesis was that the performance of the SuperLearner would be non-inferior to that of clinicians.

Materials and Methods

Study Design

We used data from an ongoing prospective cohort at three publichospitals in urban India. Our analysis is an adjunct to a registered observational study to compare the performance of clinical prediction models with clinicians (ClinicalTrials.gov identifier NCT02838459).

Study Setting

Data anlysed for this study came from patients enrolled between 28 July 2016 and XX YYYY ZZZZ at the three hospitals Khershedji Behramji Bhabha hospital (KBBH) in Mumbai, Lok Nayak Hospital of Maulana Azad Medical College (MAMC) in Delhi, and the Institute of Post-Graduate Medical Education and Research and Seth Sukhlal Karnani Memorial Hospital (SSKM) in Kolkata. KBBH is a community hospital with XX inpatient beds. There are departments of surgery, orthopedics and anesthesia, but not emergency medicine. It has a general ED where all patients are seen. Most patients present directly and are not transferred from another health centre. Plain X-rays and ultrasonography are available around the clock but computed tomography (CT) is only available in-house during day-time. During evenings and nights patients in need of a CT are referred elsewhere. MAMC and SSKM are both university and tertiary referral hospitals. This means that all specialities and imaging facilities relevant to trauma care, except emergency medicine, is available in-house around the clock. MAMC has approximately XX inpatient beds and SSKM has around XX inpatient beds. MAMC has a general ED whereas SSKM has two EDs, one where patients with suspected or confirmed neurosurgical conditions are seen and one where patients with other acute conditions are seen. The rationale for this setup is that SSKM is the only referral centre for neurosurgical care in the Kolkata metropolitan area, which has a population of close to 15 million people. Because both MAMC and SSKM are tertiary referral hospitals a majority of patients arriving at their EDs are transferred from other health facilities. Prehospital care is rudimentary in all three cities, with no organised emergency medical services. Ambulances are predominatly used for inter-hospital transfers and most patients who arrive directly from the scene of the incident are brought by the police or in private vehicles. Patients arriving to the ED are at all centres first seen be a casualty medical officer on a largely first come first served basis. There is no formalised system for prioritising ED patients at any of the centres.

Data Collection

Data was collected by one dedicated project officer at each site. The project officers all had a masters degree in life sciences. They worked five eight hour shifts per week so that mornings, evenings and nights were covered according to a rotating schedule. Project officers spent approximately six out of the eight hours collecting data in the ED and the remaining two following up patients. They first collected data on paper forms and then transferred this data to a digital database.

Participants

Eligibility criteria

Any person aged $\geq$ 18 years or older and who presented alive to the emergency department (ED) of participating sites with history of trauma was included. The age cutoff was chosen to align with Indian laws on research ethics and informed consent. We defined history of trauma as having any of the external causes of morbidity and mortality listed in block V01-Y36, chapter XX of the International Classification of Disease version 10 (ICD-10) codebook as primary complaint, with some exclusions (Supplementary material). These causes were excluded because they are not considered trauma at the participating centres.

Source and methods of selection of participants and follow up

The project officers enrolled the ten first consecutive patients who presented to the ED during each shift. The number of patients to enrol was set to ten to make follow up feasible. A follow-up was completed by the project officer 30 days after participant arrived at participating hospital. The follow-up was completed in person or per phone, depending on if the patient was still hospitalised or if the patient had been discharged. Phone numbers of one or more contact persons, e.g. relatives, were collected on enrollment and contacted if the participant did not reply on follow up. Only if neither the participant nor the contact person answered any of three repeated phone calls was the outcome recorded as missing.

Variables, Data Sources and Measurement

Patient characteristics and SuperLearner variables

The dependent variable, or label, used to train the SuperLearner was all-cause 30 day mortality, defined as death from any cause within 30 days of arrival to a participating centre. This data was extracted from patient records if the patient was still in hospital 30 days after arrival, or collected by calling the patient or a patient representative if the patient was not in hospital.

The independent variables, or features, included patient age in years, sex, mechanism of injury, type of injury, mode of transport, transfer status, time from injury to arrival in hours. The project officers collected data on these features by asking the patient, a patient representative, or by extracting the data from the patient's file. Sex was coded as male or female. Mechanism of injury was coded by the project officers using ICD-10 after completing the World Health Organization's (WHO) electronic ICD-10-training tool [22]. The levels of mechanism of injury was collapsed for analysis into transport accident (codes V00-V99), falls (W00-W19), burns (X00-X19), intentional self harm (X60-X84), assault (X85-X99 and Y00-Y09), and other mechanism (W20-99, X20-59 and Y10-36). Type of injury was coded as blunt, penetrating, or both blunt and penetrating. Mode of transport was coded as ambulance, police, private vehicle, or arrived walking. Transfer status was a binary feature indicating if the patient was transferred from another health facility or not.

The features also included vital signs measured on arrival to the ED at participating centres. The project officers recorded all vital signs using hand held equipment, i.e. these were not extracted from patient records, after receiving two days of training and yearly refreshers. Only if the hand held equipment failed to record a value did the project officers extract data from other attached monitoring equipment, if available. Systolic and diastolic blood pressure (SBP and DBP) were

measured using an automatic blood pressure monitor (OMRON HEM-7130-L). Heart rate (HR) and peripheral capillary oxygen saturation (SpO^2) were measured using a portable non-invasive fingertip pulse oximeter (ChoiceMMed MD300 C2D). Respiratory rate (RR) was measured manually by counting the number of breaths during one minute. Level of consciousness was measured using both the Glasgow coma scale (GCS) and the Alert, Voice, Pain, and Unresponsive scale (AVPU). GCS has three components, called the eye, verbal, and motor components. Each component indicates the response of the patient to no, voice or painful stimuli. The eye component ranges from one to four, where four indicates that the patient opens his or her eyes spontaneously (best response) whereas one indicates that the patient does not open eyes regardless of stimuli (worst response). The verbal and motor responses are graded similarly, but ranges between one to five and one to six respectively. The eye and verbal components also include a non-testable level. The eye component is coded non-testable if for example there is so much facial swelling that the patient cannot open his or her eyes. The verbal component is coded non-testable if for example the patient is intubated, and because of this cannot talk. In assigning GCS the project officers used the official Glasgow Coma Scale Assessment Aid [23]. AVPU simply indicates whether the patient is alert, responds to voice stimuli, painful stimuli, or does not respond at all.

The rationale for including these specific features were that they can be reasonably exptected to be available when a trauma patient arrives to the ED. They, or some variation thereof, represent standard variables collected in more or less all health systems. They are also included in the most well know clinical prediction models designed to predict trauma mortality [24].

Clinicians' priorities

Clinicians were instructed by the project officers to assign a priority to each patient. The priority levels were color coded. Red was assigned to the most serious patients that should be treated first. Green was assigned to the least serious patients that should be treated last. Orange and yellow were intermediate levels, where orange patients were less serious than red but more serious than yellow and green whereas yellow patients were less serious than red and orange patients but more serious than green patients. The clinicians were allowed to use all information available at the time when they assigned these variables, which was as soon as they had first seen the patient.

Bias

Remains to be written.

Quantitative Variables

All quantitative features (age, SBP, DBP, HR, SpO^2, and RR) were treated as continuous.

Qualitative Variables

The levels of all qualitative variables (sex, mechanism of injury, type of injury, mode of transport, transfer status, and GCS components) were treated as buckets (dummy variables).

Statistical Methods

We used R for all analyses [25]. We first made a non-random temporal split of the complete data set into a training and test set. The split was made to achieve an adequate sample size in each set as detailed in the study size section below. We then calculated descriptive statistics of all variables, using medians and interquartile ranges (IQR) for continuous variables and counts and percentages for qualitative variables.

We then developed our SuperLearner in the training set using the SuperLearner R package [26]. SuperLearner is an ensemble machine learning algorithm, meaning that it uses a library of techniques or specific learners, in principle any technique or learner that the analyst wants, to come up with an

"optimal learner". Our library included techniques suitable for predicting a binary outcome such as all cause 30-day mortality (Table To Be Defined). The SuperLearner was trained using ten fold cross validation. This procedure is implemented by default in the SuperLearner package and entails splitting the development data in ten mutually exclusive parts of approximately the same size. All learners included in the library are then fitted using the combined data of nine of these parts and evaluated in the tenth. This procedure is then repeated ten times, i.e. each part is used once as the evaluation data, and is intended to limit overfitting and reduce optimism.

The SuperLearner was then used to assign levels of priority to the patients in the training set. This was done by bucketing the SuperLearner prediction into four buckets by its values at the 25^{th}, 50^{th}, and 75^{th} percentiles. These buckets corresponded to the green, yellow, orange, and red levels of priority assigned by the clinicians'. The performance of both the continuous and bucketed SuperLearner predictions in the training set was then evaluated using the area under the receiver operating characteristics curve (AUROCC).

We then used the SuperLearner to predict the outcomes of the patients in the test set and used the cutoff values from the training set to assign a level of priority to each patient in this set. The perfomance of the continous and bucketed SuperLearner predictions, as well as the clinicians, was then evaluated by estimating their AUROCC. The levels of priority assigned by the SuperLearner and clinicians respectively were then compared using the net reclassification improvement, in events (patient with the outcome, i.e. died within 30-days from arrival) and non-events (patient without the outcome) respectively. The net reclassification improvement in events was defined as the difference between the proportion of events assigned a higher priority by the SuperLearner than the clinicians and the proportion of events assigned a lower priority by the SuperLearner than the clinicians. Conversely, the net reclassification in non-events was defined as the difference between th proportion of non-events assigned to a lower priority by the SuperLearner than the clinicians and the proportion of non-events assigned a higher priority by the SuperLearner than the clinicians.

We used an emperical bootstrap with 1000 draws of the same size as the original set to estimate 95% confidence interval (CI) around point estimates and differences. We concluded that the SuperLearner was non-inferior to clinicians if the 95% CI of the net reclassification in events did not exceed a pre-specified level of -0.05, indicating that clinicians correctly classified 5 in 100 events more than the SuperLearner.

Study Size

Remains to be written.

Results

Remains to be written.

Discussion

Remains to be written.

Conclusion

Remains to be written.

Acknowledgments

Remains to be written.

References

[1] Brohi K, Schreiber M. The new survivors and a new era for trauma research. PLoS Med. 2017;14(7):3–5.

[2] GBD 2016 Causes of Death Collaborators Global, regional, and national age-sex specific mortality for 264 causes of death, 1980 – 2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2017 September 16;390:1151–210.

[3] United Nations, Division for Sustainable Development Sustainable development goal 3. Ensure healthy lives and promote well-being for all at all ages [Cited 4 May 2018]. Available from: `https://sustainabledevelopment.un.org/sdg3`

[4] Fitzgerald M, Cameron P, Mackenzie C, Farrow N, Scicluna P, Gocentas R, et al. Trauma Resuscitation Errors and Computer-Assisted Decision Support. Arch Surg. 2011;146(2):218–25.

[5] Eastern Association for the Surgery of Trauma (EAST). Practice Management Guidelines for the Appropriate Triage of the Victim of Trauma. EAST. 2010.

[6] National Institute for Health and Care Excellence (NICE) Major trauma: service delivery NICE. 2016.

[7] 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 April;153(4):322–7.

[8] Granström A, Strömmer L, Schandl A, Östlund A. A criteria-directed protocol for in-hospital triage of trauma patients. Eur J Emerg Med. 2018;25(1):25–31.

[9] Committee on Trauma American College of Surgeons. Resources for the optimal care of the injured patient. Chicago; 2014.

[10] Tignanelli CJ, Vander Kolk WE, Mikhail JN, Delano MJ, Hemmila MR. Noncompliance with American College of Surgeons Committee on Trauma recommended criteria for full trauma team activation is associated with undertriage deaths. J Trauma Acute Care Surg. 2018;84(2):287–94.

[11] Benjamin ER, Khor D, Cho J, Biswas S, Inaba K, Demetriades D. The Age of Undertriage: Current Trauma Triage Criteria Underestimate The Role of Age and Comorbidities in Early Mortality. J Emerg Med. 2018;(April):1–10.

[12] van Rein EAJ, van der Sluijs R, Houwert RM, Gunning AC, Lichtveld RA, Leenen LPH, et al. Effectiveness of prehospital trauma triage systems in selecting severely injured patients: Is comparative analysis possible?. Am J Emerg Med. 2018.

[13] South African Triage Group. The South African Triage Scale Training Manual 2012. Western Cape Government. 2012.

[14] Beam AL, Kohane IS. Big Data and Machine Learning in Health Care. Jama. 2018 March 12;E1–2.

[15] Nevin L. Human Intelligence & Artificial Intelligence in Medicine: A day with the Stanford Presence Center. PLOS Blogs. 2018 [cited 2018 May 5]. Available from: http://blogs.plos.org/speakingofmedicine/2018/04/24/human-intelligence-artificial-intelligence-in-medicine-a-day-with-the-stanford-presence-center/

[16] Liu NT, Salinas J. Machine Learning for Predicting Outcomes in Trauma. Shock. 2017;48(5):504–10.

[17] Talbert S, Talbert DA. A comparison of a decision tree induction algorithm with the ACS guidelines for trauma triage. AMIA Annu Symp Proc. 2007;1127.

[18] Pearl A, Bar-Or R, Bar-Or D. An artificial neural network derived trauma outcome prediction score as an aid to triage for non-clinicians. Stud Heal Technol Informatics. 2008;136:253–8.

[19] Scerbo M, Radhakrishnan H, Cotton B, Dua A, Del Junco D, Wade C, et al. Prehospital triage of trauma patients using the Random Forest computer algorithm. J Surg Res. 2014;187(2):371–6.

[20] Follin A, Jacqmin S, Chhor V, Bellenfant F, Robin S, Guinvarc'h A, et al. Tree-based algorithm for prehospital triage of polytrauma patients. Injury. 2016;47(7):1555–61.

[21] Levin S, Toerper M, Hamrock E, Hinson JS, Barnes S, Gardner H, et al. Machine-Learning-Based Electronic Triage More Accurately Differentiates Patients With Respect to Clinical Outcomes Compared With the Emergency Severity Index. Ann Emerg Med. 2018;71(5):565–574.e2.

[22] World Health Organization ICD-10 Interactive Self Learning Tool. WHO [cited 2018 May 7]. Available from: http://apps.who.int/classifications/apps/icd/icd10training/

[23] glasgowcomascale.org GLASGOW COMA SCALE: Do it this way. glasgowcomascale.org [cited 2018 May 7]. Available from: http://www.glasgowcomascale.org/downloads/GCS-Assessment-Aid-English.pdf?v=3

[24] Rehn M, Perel P, Blackhall K, Lossius HM. Prognostic models for the early care of trauma patients: a systematic review.. Scand J Trauma Resusc Emerg Med. 2011;19(17).

[25] R Core Team R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2017. Available from https://www.R-project.org/.

[26] Polley E, LeDell E, Kennedy C, van der Laan M. SuperLearner: Super Learner Prediction. R package version 2.0-23. 2018. Available from: https://CRAN.R-project.org/package=SuperLearner