Development and validation of a local trauma severity score for adult trauma patients in urban India

Study plan

Siddarth David

Martin Gerdin Wärnberg

# Background

Trauma accounts for one-tenth of all deaths and disability-adjusted life-years (DALYs) [@chandran2010; @toroyan2013; @haagsma2016]. Nearly 90% of trauma-related deaths occur in low-and-middle-income countries (LMICs), and improving trauma care in these settings can save nearly 2 million lives each year (4,5). Standardized quantification of variables is important to assess patient prognosis and can improve the quality of trauma care (6,7) These variables can be used as predictors to develop prediction models to estimate the probability of defined outcomes (8,9). This can play a crucial role in over-burdened settings with limited pre-hospital care or triage (10,11).

Prediction models should be objective, replicable in different settings, less resource-intensive, and revised over time (12–14). There are multiple variables that affect trauma outcomes but trauma severity is the most important variable for predicting trauma outcomes (15,16). Severity informs clinical practice at different stages such as pre-hospital triage, in-hospital decision-making, and patient outcomes (12,17). There are several trauma scoring systems designed to quantify trauma severity. They predict mortality using different physiological, anatomical parameters, injury features, and patient characteristics to predict patient prognosis, specifically patient mortality (18–28).

The most widely used trauma severity prediction model is the Trauma and Injury Severity Score (TRISS) (29). This model was developed using a large population sample from the United States and Canada, and it predicts mortality using age, physiological status, anatomical severity of the injury, and the nature of the injury (21,30). Despite subsequent revisions and additions (31,32), TRISS continues to have considerable limitations. There are methodological issues such as the nature of the variables, high propensity for misclassification, and limited external validation (12,33–41). TRISS has also been criticized for its poor predictions in different settings, especially in LMICs which have limited resources to capture advanced or accurate data (6,42–52).

Machine learning algorithms are increasingly being used in medicine, including trauma research and care, to accurately predict complex outcomes across different settings (53–55). Ensemble machine learning algorithms combine several different statistical techniques to build an optimal prediction model rather than relying on a single technique (56). Thus, they are flexible and can be used to capture the complex relationships in trauma data. The aim of this study is to develop a local trauma severity model using an ensemble machine learning algorithm and to compare this model with TRISS.

# Methodology

## Study Design

This is a retrospective analysis of a prospectively collected multi-center observational cohort from three public hospitals in urban India between August 2016 to December 2019. The study will be reported using the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) guidelines (9).

## Setting

The three hospitals participating in this study are Maulana Azad Medical College (MAMC), New Delhi; KB Bhabha Hospital (KBBH), Mumbai; and the Institute of Post-Graduate Medical Education and Seth Sukhlal Karnani Memorial Hospital (IPGMER & SSKM), Kolkata. They are part of an on-going study Trauma Triage Study (TTRIS). Each of the three hospitals have trauma units that receive patients from across the city. These are public hospital with free or nominal fees, providing access to low socio-economic groups.

## Eligibility Criteria

All adult patients (age 18 years of age) who were admitted to participating centers with a history of trauma—Chapter XX, block V01-Y36, in the International Classification of Disease 10th-revision (ICD 10) (57). Patients who were dead on arrival were excluded.

## Variables

The primary outcome variable will be all-cause mortality within 30-days of arrival at the participating center. Predictor variables were selected based on clinically relevant variables from literature as well as consultations with trauma surgeons experience in working in the urban Indian setting. These included physiological measures: systolic blood pressure (SBP), respiratory rate (RR), heart rate (HR), oxygen saturation, and Glasgow Coma Scale (GCS); injury etiology: time of injury, mode of transport, type of injury, mechanism of injury, and injury severity scores (ISS); and demographic measures: age and sex.

## Data

Each participating center has a dedicated project officer collecting data in 8-hour shifts per day by prospectively enrolling patients. On-arrival of the patients the project officer would also measure SBP, HR, RR, and oxygen saturation independently but would not be part of patient care. The project officer would follow-up with the patients in the hospital and if discharged telephonically to record the mortality at 24-hours, 30-days and 6-months. The shifts would alternate between morning, evening, and night in rotation. The project officers have continuous training and supervision throughout the study period. The collected data is uploaded to a central database and each week reviewed by the research team. Based on injury details, ISS was computed for each participant by accredited coders.

## Developing Trauma Severity Models

For the local model we constructed a priori consisting of the predictor variables mentioned above. For the TRISS model, we calculated the Revised Trauma Score (RTS) based on GCS, SBP and RR. Age, type of injury (blunt or penetrating) RTS and ISS were used to calculate the probability of survival (P) ranging from 0 to 1, where 0 corresponds to 0% and 1 to 100% probability of survival (30,58).

## Analyses and Statistical Methods

The ensemble machine learning procedure SuperLearner will be used in the study (59). It will combine different statistical techniques (also called machine learning algorithms) such as generalized linear and additive models, random forests, etc. to create a local model that best fits with the data. SuperLearner uses cross-validation to optimize prediction by separating the data into a “training sample” which will be used to generate a model and then using another “test sample” for validating model. We will use the training set to build the ensemble learner and to update TRISS. We will use the test set to estimate the performance of each of the ensemble learner, the original TRISS, and the updated TRISS.

Based on a recent review of machine learning (ML) algorithms for predicting outcomes in trauma patients (40), we selected seven most commonly used machine learning (ML) algorithms for developing the model: support vector machines (SVM), artificial neural networks (ANN), decision trees (DT), Bayes classification (BC), k-nearest neighbor (KNN), random forest (RF), and logistic regression (LR). We will then compare the performance of all models in a pair-wise fashion. Performance and differences in performance will be estimated as medians across imputations and 95% confidence intervals will be estimated using bootstrapping.

## Sample Size

To develop a prediction model with a binary outcome, the current recommendation, is to include at least ten events, i.e. participants with the outcome, and at least as many non-events per free parameter in the model (60). Depending on the data structure as many as 25 events and non-events or more per free parameter may be required to obtain stable estimates (61). These recommendations are however mainly for logistic regression, whereas no recommendations exist for ensemble learners except that more data is likely needed (62). We will therefore include at least 25 events and non-events per free parameter in the training sample. With 12 predictor variables and around 40 free parameters, an estimated sample of at least 2000 will required. The training sample constitute 80% of the total sample. The remaining 20% of the cohort will be used as the test sample.

## Missing Data

We will use multiple imputation using chained equations to handle missing data (63).

## Model Assessment

We will assess the local model and TRISS model for overall performance, discrimination, and calibration (64). Discrimination, if the higher scores correspond to higher mortality, will be measured using sensitivity, specificity, and the cross-validated area under the receiver-operating characteristic curve (AUROC), reported with corresponding 95% confidence interval (95% CI). Calibration, if the predicted mortality coincides well with the observed mortality, will be assessed by either Hosmer-Lemeshow statistic or Cox calibration test.

## Ethical Considerations

The institutional ethics committee of each participating center has individually approved the collation and analysis of the TTRIS dataset. The reference numbers are: Maulana Azad Medical College, New Delhi (F.1/IEC/MAMC/(53/2/2016/No.97); KB Bhabha Hospital, Mumbai (HO/4882/KBBH of 3/8/2016); and Institute of Post-Graduate Medical Education and Seth Sukhlal Karnani Memorial Hospital (IPGMER & SSKM), Kolkata (Inst/IEC/2016/328).

# References