Forensic Science International 281 (2017) 92–97

Contents lists available at ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

Prediction of mortality risk in victims of violent crimes

Rolf Gedeborg[a,b], Bodil Svennblad[b], Liisa Byberg[c], Karl Michaëlsson[c], Ingemar Thiblin[d,*]

[a] Dept. of Surgical Sciences, Anesthesiology and Intensive Care, Uppsala University, Uppsala, Sweden
[b] Uppsala Clinical Research Center, Uppsala University, Uppsala, Sweden
[c] Dept. of Surgical Sciences, Orthopedics, Uppsala University, Uppsala, Sweden
[d] Dept. of Surgical Sciences, Forensic Medicine, Uppsala University, Uppsala, Sweden

ARTICLE INFO

Article history:
Received 15 February 2017
Received in revised form 7 June 2017
Accepted 11 October 2017
Available online 20 October 2017

Keywords:
Bayesian inference
Mortality
Violent crime
Forensic medicine

ABSTRACT

Background: To predict mortality risk in victims of violent crimes based on individual injury diagnoses and other information available in health care registries.
Methods: Data from the Swedish hospital discharge registry and the cause of death registry were combined to identify 15,000 hospitalisations or prehospital deaths related to violent crimes. The ability of patient characteristics, injury type and severity, and cause of injury to predict death was modelled using conventional, Lasso, or Bayesian logistic regression in a development dataset and evaluated in a validation dataset.
Results: Of 14,470 injury events severe enough to cause death or hospitalization 3.7% (556) died before hospital admission and 0.5% (71) during the hospital stay. The majority (76%) of hospital survivors had minor injury severity and most (67%) were discharged from hospital within 1 day. A multivariable model with age, sex, the ICD-10 based injury severity score (ICISS), cause of injury, and major injury region provided predictions with very good discrimination (C-index = 0.99) and calibration. Adding information on major injury interactions further improved model performance. Modeling individual injury diagnoses did not improve predictions over the combined ICISS score.
Conclusions: Mortality risk after violent crimes can be accurately estimated using administrative data. The use of Bayesian regression models provides meaningful risk assessment with more straightforward interpretation of uncertainty of the prediction, potentially also on the individual level. This can aid estimation of incidence trends over time and comparisons of outcome of violent crimes for injury surveillance and in forensic medicine.

1. Introduction

Injuries resulting from violent crimes are responsible for more than one million deaths annually, and for each death many more suffer non-fatal injuries [1]. Violence is one of the most common causes of death among young males worldwide [2]. Due to the circumstances, physical injuries are often complicated not only by severe disability but also long-standing mental health problems [3].

While a downward trend in homicide rates since 1990 has been reported [4] there is a general notion of increased brutality in violent crimes over the same time period and reports of increased incidence of violent crimes [5]. One explanation for this apparent paradox is the potential impact by changes over time in treatment success after severe injury [6].

Another potential application for injury severity estimation is the assessment required in the legal procedures following violent crimes. These could be made more objective using a reliable prediction model with estimates of precision that have a reasonably meaningful interpretation of probability.

The aim of the present study was to derive a prediction model for mortality risk after violent crimes using data from Swedish national hospital discharge and/or cause of death registries. A specific goal was to see if a model based on individual ICD-10 diagnoses and measures of interaction further improves accuracy compared to the ICD-based injury severity score (ICISS).

2. Material and methods

2.1. Data sources and study population

The Swedish National Patient Registry (NPR) [7] and the Swedish Cause of Death Registry (CDR) were linked using the unique personal identification number that is given to all Swedish

* Corresponding author.
E-mail address: ingemar.thiblin@surgsci.uu.se (I. Thiblin).

https://doi.org/10.1016/j.forsciint.2017.10.015
0379-0738/

citizens [8,9]. Injury hospitalizations from 1998 to 2004 were defined as hospital admissions with a principal diagnosis S00–T80 but excluding T78 as listed in ICD-10 [10]. Readmissions were excluded on the basis of a validated prediction model [11]. In 99.5% (624/627) of the deaths the cause of death was determined after autopsy. ICD-10 cause of injury codes were present in 96% of these incident injury hospital admissions. They were categorized using the cause of injury matrix [12,13]. Only the first homicide/assault event for each individual was included in the study population. The regional Human Ethics Committee approved the study.

2.2. Injury severity estimates

The International Classification of Diseases Injury Severity Score (ICISS) has been shown to perform well compared to other injury severity scores [14–17]. It is calculated from diagnosis-specific survival probabilities (DSPs) for individual injury ICD codes. This ratio represents the proportion of patients with a specific injury code who survived until hospital discharge. For evaluation of the predictive ability of ICISS, it was calculated using DSPs based on all injury hospitalizations and prehospital injury deaths during 1998–2002. The ICISS score for the individual patient was calculated as the product of each of the DSPs corresponding to the patient's injuries (i.e. the product of the probabilities of surviving each of their injuries individually).

For comparison individual ICD-10 injury codes were also used in the regression models, along with possible two-way interactions. ICISS strata for descriptive purposes were defined as critical (0–0.219), severe (0.220–0.354), serious (0.355–0.664), moderate (0.665–0.940) or minor (0.941–1.0) [18].

2.3. Other potential predictor variables

Age and sex were used in the basic reference model and were always kept in all other models. After inspection of a smoothed function of age against mortality using a general additive model (GAM, mgcv package in R version 2.11.1) [19] age was found to be non-linear to the logit of the outcome and therefore used after log transformation in the models.

Cause of injury was modeled using dummy variables to represent eight categories from the cause of injury matrix referenced above — cut/pierce, fall, fire/flame, firearm, poisoning, suffocation, other specified, not specified. The ninth main cause of injury, struck by/against, served as the reference category.

2.4. Outcome

Hospital deaths were identified as deaths during a hospital stay with injury as the main diagnosis. Autopsied deaths with an underlying cause of death in the range of V01–Y36 but with no associated recording of an injury hospital admission in the NPR were considered as prehospital injury deaths [20]. Prehospital and hospital deaths were combined to generate the mortality outcome variable.

2.5. Statistics

The statistical packages SAS version 9.2 (SAS Institute Inc., Cary, NC, USA) and R version 2.11.1 (R Foundation for Statistical Computing, Vienna, Austria) were used for data management and statistical analyses as further specified below. Prediction models for binary outcomes are commonly developed using logistic regression with Maximum Likelihood (ML) estimates. For our purposes, there are two important disadvantages with this approach: Firstly, estimates become increasingly unstable as the number of predictors increase. Basing injury severity estimation on ICD-10 codes means in this case to make predictions from up to 1188 individual ICD-10 injury diagnoses, which would require extremely large datasets using standard logistic regression. Secondly, standard logistic regression is a population-averaged model, meaning that predictions and their confidence intervals must be applied to individuals with caution and their interpretation is not straightforward. Bayesian models generating credibility intervals for the predicted probabilities for specific covariate patterns have a more intuitive interpretation.

The study population was divided into a training dataset consisting of all injury events 1998 through 2002 (n = 10,260), and a validation dataset (n = 4210) using all the injury events from 2003 through 2004. Several models, summarized in Table 2, were estimated with standard logistic, Lasso and/or Bayesian logistic regression in the training dataset with injury death as outcome. The estimated models were then applied to patients in the validation dataset to obtain individual probabilities for their outcome. To validate the models predictive ability, Harrell's R package 'rms' was used to generate estimates of the area under the receiver operating characteristic curve (C-index), Brier Score, intercept and slope from a new logistic regression estimating the relationship between the estimated probability and the observed outcome (model calibration) [21]. We also calculated the Hosmer Lemeshow's goodness of fit statistic using deciles of predicted risk and the scaled Brier score [22].

For models not including an excessive number of predictors (e.g. not incorporating the individual ICD-10 codes) standard logistic regression was used as the reference method. ICISS was modelled both as a linear effect and after logit transformation [23]. The impact of transformation was assessed using GAM plots and comparing predictive performance of the models.

When a prediction model is derived and validated in the same dataset, the validation tends to be overoptimistic as to the models performance when applied to a new external dataset. There will be a tendency for low predictions to be too low and high predictions to be too high [24]. More accurate performance in external data may be achieved by shrinking the maximum likelihood estimates of coefficients provided by the standard logistic regression towards zero. One such shrinkage technique is Lasso (Least Absolute Shrinkage and Selection Operator) regression [25]. This method shrinks some of the coefficients to exactly zero, thereby serving not only as a technique for shrinkage but also resulting in variable selection. The method can therefore be useful in a situation with many potential predictors and small datasets. Lasso regression was applied using the R package Penalized [26].

When introducing the Lasso method Tibshirani also gave it a Bayesian interpretation [25]. Assuming the coefficients are independently distributed according to a Laplace distribution centered in 0, the Lasso coefficients are represented by the maximum posterior mode (MAP) from the resulting marginal posterior distributions. This is a property of the MAP estimator [27], and not a fully Bayesian approach but offers a way to obtain Lasso estimates in practice.

Using Bayesian logistic regression generally we have several options to obtain point estimates of the coefficients, e.g. (a) the MAP as suggested by Tibshirani or (b) the posterior mean from the marginal distributions. For prediction of p, the probability of the outcome, based on a new observation of the predictors, we can use the point estimates (a) or (b) of the coefficients resulting in a point estimate of p, or we can (c) estimate the posterior distribution of p and as point estimate use e.g. the mean. Using the full distribution of p also gives an opportunity to calculate a credibility interval for the probability of outcome given the new observation.

Applying Lasso and Bayesian logistic regression with MAP estimates provides the opportunity to model a large number of predictors, e.g. all individual ICD-10 injury codes with all possible