Risk Score for Predicting Mortality in Flail Chest

Meaghan Zehr, MS, Neil Klar, PhD, and Richard A. Malthaner, MD, MS

Departments of Epidemiology and Biostatistics and Surgery, Division of Thoracic Surgery, Western University, London, Ontario, Canada

Background. **Flail chest injuries are associated with high mortality and morbidity. Despite evidence that operative repair of flail chest is beneficial, it is rarely done. We sought to create a simple risk score using available preoperative covariates to calculate individual risk of mortality in flail chest.**

Methods. **A logistic regression model was trained on Ontario Trauma Registry data to generate a mortality risk score. The final model was validated for calibration and discrimination and corrected for optimism.**

Results. **The model uses five risk factors that are readily obtained during the initial assessment of the trauma patient: age, Glasgow Coma Score, ventilation, cardiopulmonary resuscitation, and number of comorbidities. It was determined that less than 6 points is consistent with 1% observed mortality, 6 to 10 points predicts 5% mortality, 11 to 15 points predicts 22% mortality, and 16 or more points predicts 46% mortality.**

Conclusions. **We have developed a simple model that can be easily applied at bedside to predict mortality in patients with flail chest by accessing a spreadsheet program in an application or other handheld computer device. This model has the potential to be a useful tool for surgeons considering operative repair of flail chest.**

(Ann Thorac Surg 2015;■:■–■)

Flail chest is a severe type of rib fracture with an estimated mortality rate of 10% to 36% of cases [1–4]. Operative repair of flail chest has been found to reduce length of stay and complications in the critical care unit and to improve long-term quality of life when compared with standard methods of care [5–12]. Despite proven benefits, surgical chest wall stabilization continues to be underused [8–10]. There is uncertainty about the optimal technique and its indications. Several studies have attempted to characterize mortality risk factors in flail chest, but the findings are conflicting [13–19]. The cut points vary, parameters are not available before surgery, terms are too broadly defined, and the clinical applications are unclear.

A risk score that could be calculated before operative repair is needed to quantify individual risk of mortality in flail chest. Patients at low to medium risk of mortality would then be good candidates for operative repair. Traditionally, logistic regression has been used to predict binary outcomes such as mortality and to generate scoring systems that can be applied at bedside by using a calculator, application, or computer. The objective of this study was to create a simple risk score using available preoperative covariates to calculate individual risk of mortality.

Accepted for publication March 25, 2015.

Presented at the Poster Session of the Fifty-first Annual Meeting of The Society of Thoracic Surgeons, San Diego, CA, Jan 24–28, 2015.

Address correspondence to Dr Malthaner, Divisions of Thoracic Surgery and Surgical Oncology, London Health Sciences Center, Victoria Hospital, 800 Commissioners Rd E, Ste E2-124, London, ON N6A 5W9, Canada; e-mail: richard.malthaner@lhsc.on.ca.

Material and Methods

Data were available for 1,190 adult flail chest patients admitted to lead trauma hospitals from January 1, 1999, to March 31, 2009, and recorded as part of the Ontario Trauma Registry (OTR). The OTR was established in 1992, and its direct sources are the 11 lead trauma facilities in Ontario [20]. Each hospital is mandated to report demographic, prehospital and hospital care, and patient outcomes on all adult hospitalizations due to major trauma [20]. The Canadian Institute for Health Information maintains the OTR and deidentifies all patient and institution data [21]. According to Canadian and provincial legislation, it may disclose data without the consent of the individual patient for its mandated purposes, including statistical analyses and reporting [21]. Therefore, ethical approval was not necessary for this study. To be included in our study, patients must have survived at least 24 hours after admission to a lead trauma hospital, leaving 1,082 patients from the original total. There were 41 cases of operative repair of flail chest that were included and adjusted for in this analysis.

Model predictors were selected by testing 15 potential mortality risk factors for significance (ie, demographics, physiology, injury, procedures) through independent t tests for continuous variables and Fisher's exact test for categorical variables, adjusting for multiple testing using the Bonferroni correction (p value was calculated as $p < 0.003$ after $p < 0.05$ was divided by 15 variable comparisons). The definitive airway variable was defined as oral or nasal intubation or tracheotomy. Mechanical ventilation included both invasive cases (intermittent positive pressure ventilation with endotracheal intubation) and noninvasive cases (continuous positive airway pressure

0003-4975/$36.00
http://dx.doi.org/10.1016/j.athoracsur.2015.03.090

Abbreviations and Acronyms

BAC = blood alcohol concentration
CPAP = continuous positive airway pressure
CPR = cardiopulmonary resuscitation
GCS = Glasgow Coma Scale
IPPV = intermittent positive pressure ventilation
ISS = injury severity score
OR = odds ratio
OTR = Ontario Trauma Registry
SAS = statistical analysis system
SD = standard deviation
TRISS = trauma injury severity score

given through a face mask). To avoid counting definitive airway and mechanical ventilation twice by the model (as a definitive airway is often required to administer ventilation), we chose to only include mechanical ventilation as a model predictor. Year of hospital admission and operative repair were investigated as potential confounders. We used the first recorded Glasgow Coma Scale (GCS) score at scene of injury or in hospital; that was the only variable that contained missing observations, and there were 132 (12.2%). Logistic regression was, therefore, modeled three ways to account for possible biases due to missing observations, including selection bias, confounding, and lack of generalizability, respectively: (1) a model excluding missing data (ie, complete case analysis); (2) a model excluding GCS score as a predictor; and (3) a model including multiple imputation for missing observations of GCS. Numerous imputations were performed using a fully conditional specification approach, which has been shown to provide good coverage even with nonnormal factors such as GCS score. Forty imputations were used, as has been recommended to prevent power falloff [22, 23]. The endpoint was all-cause inhospital mortality.

The study followed the suggested modeling strategy of Harrell and colleagues [24], using the entire dataset to train the model. Component plus residual plots were evaluated to determine whether the model met linearity assumptions. Predetermined interactions included in the model were age by GCS and age by ventilation. These interaction terms were chosen because age is known to be a strong risk factor for mortality, and there was enough variation in GCS score and ventilation to make a study of these possible interactions relevant. A prespecified subset of predictors was tested for significance in the logistic model, including the two interaction terms and two possible confounders (age by GCS, age by ventilation, year of hospital admission, and whether operative repair was performed). If a predictor was not significant but had an odds ratio (OR) greater than or equal to 1.2, it was allowed to remain in the model because these predictors contribute to model performance [25].

The final model was validated for calibration and discrimination. Calibration refers to how well the observed outcome and predicted outcome agree and can be assessed with the Hosmer-Lemeshow goodness of fit test [26]. Discrimination refers to how well the model can separate patients at high risk of death and patients at low risk of death and is given by area under the receiver-operating characteristic curve (c index) [26, 27]. A good or excellent predictive model would have a c index above 0.8, a moderately discriminating model would have a c index between 0.7 and 0.8, and a low discriminating model would have a c index between 0.6 and 0.7 [26].

Predictive models tend to perform better in the training dataset than in new datasets; therefore, to determine the expected model performance in new patients at similar risk of mortality, the optimism-corrected c index was calculated [24]. The c index was calculated for the final model in the original sample and was compared with the c indexes calculated in bootstrap samples. Bootstrapping was performed by generating 400 random samples of equal size to the original sample with replacement from the original sample. The average difference between the c index calculated using the original sample and the c indexes calculated using bootstrap samples represents the optimism of the model [24]. All analysis was conducted using SAS software, version 9.4 (SAS Institute, Cary, NC).

The ORs and c index of the final model 1 were compared with the other two models used to address the missing values. A plot displaying the predicted and observed probability of risk across the deciles of risk for model 1 is provided in Figure 1.

Risk scores for mortality were calculated using the method developed to produce the Charlson Comorbidity Index, where a whole number point is assigned based on the OR of a risk factor [28]. In this method, if the OR is equal to or greater than 1.2 but less than 1.5, 1 point is assigned; for ORs equal to or greater than 1.5 but less than 2.5, 2 points are assigned; for ORs equal to or greater than 2.5 but less than 3.5, 3 points are assigned, and so forth. For ORs less than 1, the inverse was taken so that positive point values were always assigned. A plot displaying the number of points and observed mortality is provided in Figure 2.

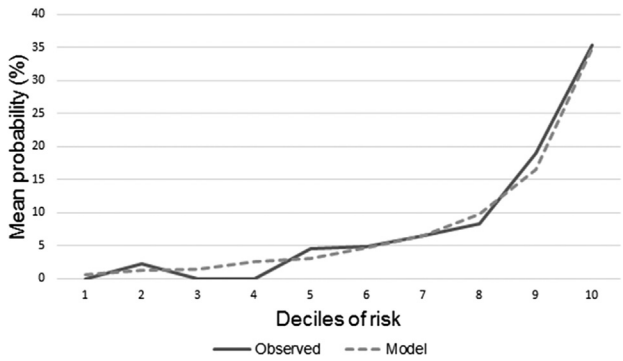

Fig 1. The observed (solid line) and predicted probabilities of death across deciles of risk for model 1 (dashed line) tend to increase with higher rankings of predicted risk of death.