Factors Associated with Mortality Following Admission to a Hospital for a Gunshot Wound

Corey Evans

Masters Student in Biostatistics

University of Vermont, Burlington, Vermont

**Abstract**

Each year in the United States there are approximately 30,000 deaths that occur due to gunshot wounds. While this number is approximately the same as the number of deaths each year due to car accidents and drug overdoses, obtaining data related to this is difficult due to Federal funding restrictions. Examining the risk factors associated with gunshot wound mortality is critical in determining the best approach to take in reducing gunshot wound death, and potentially reducing gunshot wounds as a whole. This paper investigates several risk factors believed to be involved in gunshot wound mortality using a sample of the National Inpatient Survey (NIS) data specific to gunshot wounds. The results indicate that higher TMPM scores when shock is not present is a major risk factor for death in gunshot related hospital admissions. The model developed from this data correctly classifies with an overall rate of 93%. This model may be useful in clinical settings in order to assess the chance of survival of a patient, based on the covariates identified as significant.

**Introduction**

Gunshot wounds are endemic in the United States. Despite this, due to Federal legislation in 1994 blocking federal funding to the CDC for any data collection on gunshot wounds it is an incredibly understudied area. The National Inpatient Survey is an extremely large data set that contains data on patients from many hospitals, such as why they were admitted, among many other factors. From this large dataset gunshot specific data was extracted. The purpose of this paper is to use a bootstrapped sample of this data to examine the risk factors associated with gunshot wound mortality. The focus is to develop a logistic regression model that accurately assesses whether or not the outcome will be observed in the patient based upon the risk factors that will be discussed later in this paper.

**Methods**

Logistic Regression Analysis was performed by purposeful selection using R (version 3.3.3). Fractional polynomial analysis was performed using R (package: mfp). Fit diagnostics were performed using R (package: LogisticDx)

**Results**

*Initial Analysis of Covariates*

The univariate analysis of each covariate revealed that on their own, there was a significant relationship between the subject dying in the hospital (DIED) and each covariate examined (Appendix Table 1.). Multivariable logistic regression of subject dying in hospital with all the covariates revealed several non-significant (p>0.01) factor levels for covariates, as well as several non-significant covariates (Appendix Table 2). Covariates with that had only one significant factor level were re-coded in order to reduce the number of parameters in the model that were not significant. Refitting the model resulted in covariates that were all significant, except for Gunshot wound intent (GSW\_INT) (Appendix Table 3). Gunshot wound intent was not restructured as two levels were significant, while two were not. A test of the original model against the new smaller model revealed that the new model was not good as the old one (p=0.003). While this appears to be problematic, closer examination of the model reveals that this low p-value is likely due to the restructuring of the factor level covariates for RACE, INS\_TYP, YEAR, and GUN\_TYP. The variables that were not restructured changed very little in their estimated coefficients. The estimated coefficients for the compressed factor variables did change by more than 20%, but this is a byproduct of combining several levels of a factor into one. Based on this rationale, model selection continued, using this reduced model.

*Fractional Polynomial Analysis*

Analysis (R package: mfp) revealed that transformations were required for both of the continuous covariates present in the model; Age and TMPM (Appendix Table 4). Refitting of the model with these transformed variables was performed and it was determined that FEMALE and ELECTIVE were no longer significant predictors in the model (p>0.01) and were removed. The model was then refit without these two covariates (Appendix Table 5).

*Testing of Interactions*

All potential interactions that made clinical sense were tested. Interactions for both Age and TMPM were only tested on one of the fractional polynomial variables, reducing the total number of potential interactions to 48. Of those 48, only 4 were significant (Appendix Table 6). When these interactions were all added to the model, only two were significant and were retained. This led to our preliminary final model, upon which fit diagnostics and classification analysis was performed (Appendix Table 7).

*Fit Diagnostics*

The Hosmer-Lemeshow Decile of Risk Tests was performed on the preliminary final model (Appendix Table 8). While the p-value (0.16) indicated that the model was a good fit, several plots were created to assess the influence of the individual observations in the model (Appendix Figures 1-3). From these graphs a total of 9 outliers were identified (Appendix Table 9). These observations were deleted, one at a time, in order to assess their individual effect on the coefficients. Two observations produced 33% changes in GSW\_INT4 (Gunshot wound Intent, Law Enforcement). Deletion of all other observations produced changes in the covariates of less than 15%. The outlying observations were then grouped by poorest fit and largest influence, and then these groups were deleted. Additionally, all 9 points were deleted to assess how the model fit (Appendix Table 10). Deleting the points with the poorest fit produced a 32% change in AgeFp1 and a -60% change in GSW\_INT4. The majority of this variation can be accounted for by 4 of the 7 poorest fit points. Observations 11307 and 25092 account for 25% of the 32% change in AgeFp1, while observations 42869 and 23234 account for almost all the change in GSW\_INT4. Deletion of the points with the largest influence did not produce any extremely large changes in the covariates. Deletion of all nine points once again produced large changes in the AgeFp1 and GSW\_INT4 covariates. This large change is again, due to the 4 points mentioned previously. Upon closer examination of the two points with the highest influence, it appears this high influence is due to each individual’s high TMPM scores. Based on examination of the outlying data points, they all seemed to be reasonable , and thus, none were deleted from the model.

*Classification Analysis*

Classification Analysis revealed an area under the Receiver Operating Curve of 94.6% (Appendix Figure 4). Further analysis revealed an optimum cut point at 0.85, where sensitivity was 93.3% and specificity was 88%, with an overall correct classification rate of 93%.

**Discussion**

Considering both the results of the diagnostic tests, and the classification analysis, the model developed here appears to be excellent at classifying and is a good fit for the data that was examined. Implications of this model includes the ability to rapidly predict survival for a patient who comes in as a result of a gunshot wound, with a relatively high probability that the correct outcome will be predicted. Odds ratios of the covariates indicate that the odds of death were 4.14 times greater if suicide was the intent compared to accident as the intent, holding all other covariates constant (Appendix Table 11). Among individuals who were not in shock upon admission, the odds of death were 12.16 times greater for a 1 unit increase in TMPM. In contrast, among individuals not in shock upon admission, the odds of death were only 2.76 times greater for a 1 unit increase in TMPM. This seems to indicate that SHOCK is an extremely important factor in whether or not a patient will live. Coupled with their TMPM score, it seems doctors may be able to make a somewhat accurate prediction as to that individual’s odds of survival. Odds ratios of a 10 year change in Age for 3 different TMPM scores shows very little difference in the odds ratios for the 3 different scores (Appendix Table 12). None of the odds ratios for any other covariates stood out, indicating that while they may be important in the fit of the model, only a few of the variables drive the predictive power of this model. One potential shortfall of this model may be that the state the hospital in was not used, as these results were not significant in the univariate analysis. However, grouping the states into region could potentially be useful, and may provide significant covariates. However, this would add more covariates to a model that already has fourteen in its current state. It is also likely that the collapsing of certain factors into one another, as they were not significant has led to a loss of interpretability in the model, as some of the new covariates are a conglomerate of old factor levels. Despite these potential issues, the model does show strong predictive power and could possibly be useful in assessment of a patient’s odds in a clinical setting. Additionally, the analysis has revealed that the major risk factors for death are TMPM scores when Shock is not present at admission, and when suicide is the intent of the gunshot wound.