Location in Accident Related to Death of Injured Persons in Fatality Traffic Crash

Ding Ding

Introduction

A fatal traffic crash is a collision that involves a motor vehicle traveling on a trafficway customarily open to the public, and must result in the death of an occupant of a vehicle or a non-occupant within 30 days (720 hours) of the crash. On the United States highways, motor vehicle fatality traffic crashes resulted in 35092 lives lost, 45495 injured persons, and billions of dollars in property damage[1] in 2015 alone. For those 80587 injured in fatal crashes, they either succumbed to injuries or survived. The aim of this analysis is to estimate the risk factors related to death of injuried persons in fatal traffic crashes.

Risk factors that may play key roles in death of fatal traffic crashes include the location of the person in the crash (i.e., position inside/outside vehicle)[2], age, blood alcohol level, safety belt use, whether they were ejected and extricated from the vehicle[3], among others. Types of locations include pedestrian, pedalcyclist (e.g., bicyclist), driver and various occupant positions in the vehicle. Blood alcohol content ("BAC") is generally tested on-site with a breathalyzer. Elucidating the relationship of risk factors that ultimately lead to traffic fatalities can be leveraged to decrease the overall mortality in such accidents. Here we performed an analysis to determine if there was a significant association between death of injuried persons in fatality cases and the key variables and confounders involved.

Methods

Data Collection

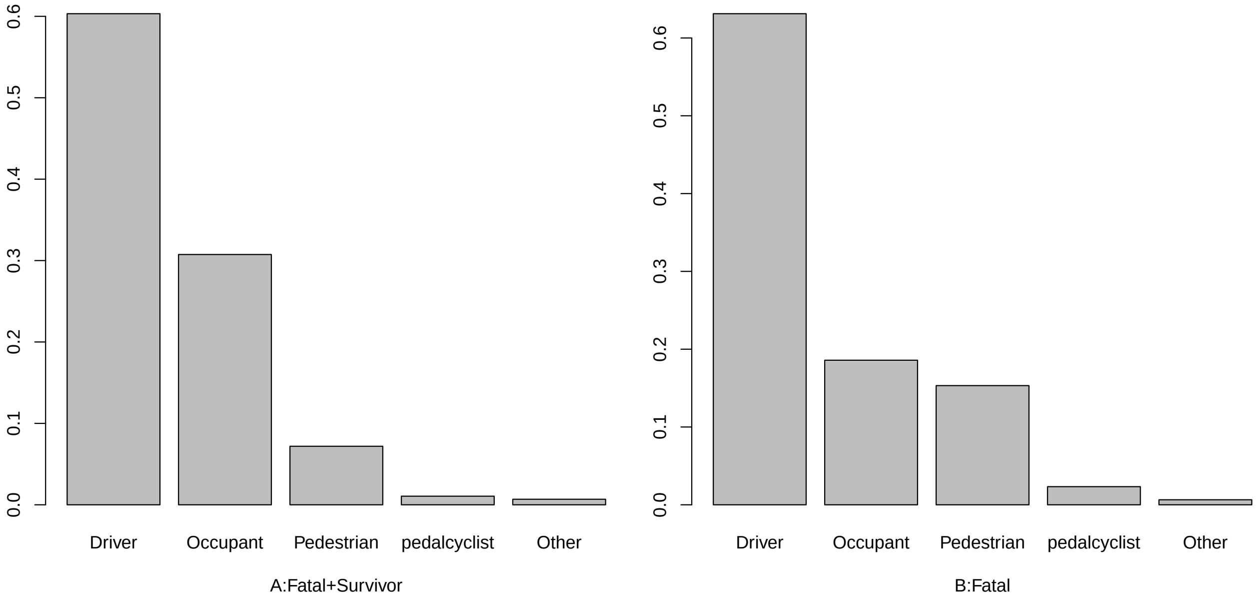
Our analysis based on open data of 2015 from Fatality Analysis Reporting System (FARS), an annual census of all fatal motor vehicle crashes occurring in the United States. Data was downloaded on September 9, 2016 using R programming language[4].

Exploratory Analysis

Exploratory analysis was performed on injured persons' information. We identified all the injuried people in fatal traffic crashes and categorized them as death or survivor, which is a binary outcome for the analysis. Observations that are missing data with respect to key variables (1133 missing in location of person in the crash and 1668 in age, in total 2521 out of 80587) are excluded from the analysis(Supplementary 1).

We plotted and tables all the potental risk factors and confounders (e.g. Person Type showed in Figure1A &1B). We found that age interval and individuals position/location with respect to the accidents have a great impact on fatality rate.

Figure1: Histogram of Injured and dead persons in Fatality crashes



Statistical Modeling

To estimate the risk factors related to death of injuried persons in fatal traffic crashes, we fit the data to a logistic regression model[5]. Model selection was performed on the basis of our exploratory analysis, the type of response(binary) and prior knowledge of the relationship between the variables and response. Models are estimated with cross-validation[6], Receiver Operating Characteristic (ROC) curve(Supplementary Fig.1) and Area Under Curve(AUC).

Results

The fatal traffic crash data used in this analysis contains information measured as the Person injury Type(Survivor,1-Fatal,0-Survivor), location of the person with the vehicle(Position,includes Non-Motor Vehicle Occupant and Motor Vehicle Occupant in different positions), age intervals(Age), Restraint Use(Restraint), Non-Motorist Location(Intersection, within an intersection or not), if equipment or other force was used to remove this person from the vehicle(Extricate) and blood alcohol test result(Alcohol).

Based on exploratory analysis, we fit the first logistic regession relating fatal injury(Survivor=0) to only position in the accidents(Model1), both position in the accidents age interval(Model2).

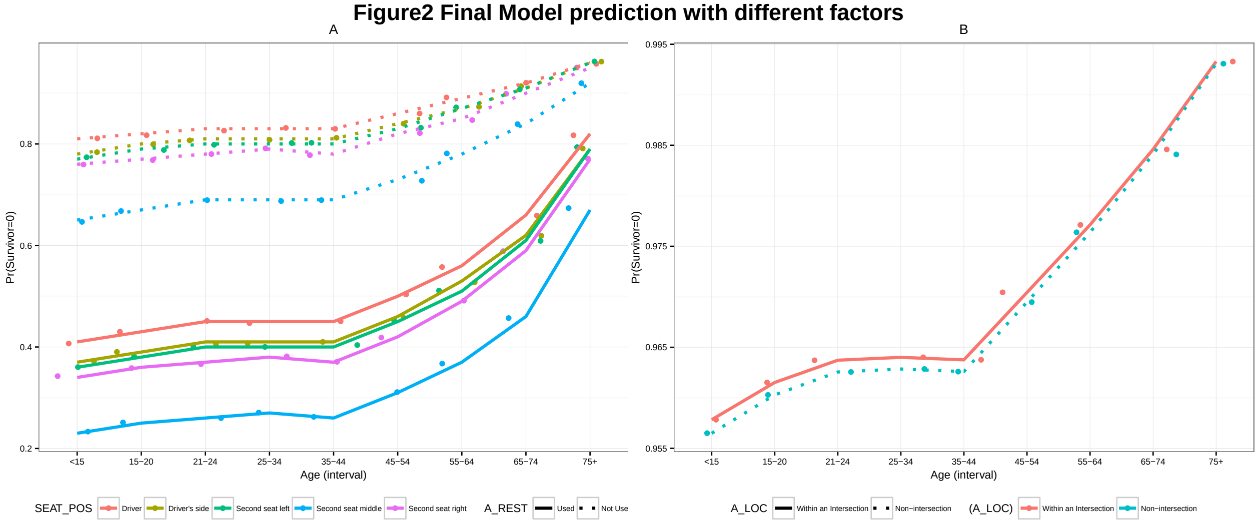
$$\text{logit}Pr(Survivor_i = 0) = \beta_0 + \boldsymbol{\beta_1} f(Position_i) + \boldsymbol{\beta_2} Age_i.$$

We evaluated the model by cross-validation[4] with AUC 0.65(Supplementary 2). Then we aimed to improve prediction by adjusting models to include potential confounders. Our final model(Model3) for fatal injuries is

$$\text{logit}Pr(Survivor_i = 0) = \beta_0 + \boldsymbol{\beta_1} Position_i + \boldsymbol{\beta_2} Age_i + \boldsymbol{\beta_3} Extricate_i + \boldsymbol{\beta_4} Restraint_i + \boldsymbol{\beta_5} Alcohol_i + \boldsymbol{\beta_6} Intersection_i$$

where all the coviariants are category treated as factors. And estimation with cross-validation comes out with AUC 0.86 (Supplementary 3). Figure2A shows both the observed and fitted probabilities of fatal injuries by age, postion in the vehicle and restraint use for the Motor Vehicle Occupant group with Extricate=FALSE, Alcohol=FALSE; Figure2B shows the observed and fitted probabilities of fatal injuries by age, Intersection for Non-Motor Vehicle Occupan group with Extricate=FALSE, Alcohol=not tested.

Figure2 Final Model prediction with different factors



We observed a strong significant positive correlation of crash fatality with both the persons' position in the crash and age. Here we summarize the probability of fatal injuries(Pr(Y>0) for some Position-Age strata(fixing Extricate=FALSE, Alcohol=FALSE, Intersection=vehicle Occupants).

Table 1a: Estimated Death Probability with Certain Factors

	Restraint Used			Restraint Not Used		
Age	Pr(Y>0,Driver)	Pr(Y>0,Driver's Side)	Pr(Y>0,Seat Behind Driver)	Pr(Y>0,Driver)	Pr(Y>0,Driver's Side)	Pr(Y>0, Seat Behind Driver)
21-24	0.45	0.41	0.4	0.83	0.81	0.8
25-34	0.45	0.41	0.4	0.83	0.81	0.8
35-44	0.45	0.41	0.4	0.83	0.81	0.8
45-54	0.5	0.46	0.45	0.86	0.84	0.83

Table 1b: Estimated Coefficient with Standard Error

	Compare to Non-Motor Vehicle Occupant				Compare to Age<15		
Position	Coefficient	Standard Error	P-value	Age	Coefficient	Standard Error	
Driver	-3.003	0.345	0	25-34	0.156	0.056	0.005
Driver's right	-3.93	0.396	0	45-54	0.36	0.057	0
Second row left	-3.131	0.345	0	55-64	0.633	0.058	0
Second row right	-3.796	0.356	0	65-74	1.048	0.062	0
Cargo Area	-4.922	0.451	0	75+	1.866	0.065	0

Conclusions

Our analysis suggests that fatal injury is significantly correlated with the age and position of injuried persons in the crash. An injuried Non-Motor Vehicle Occupant has a higher fatality than Motor Vehicle Occupant, and injuried person in the second row middle seat has a lower fatality rate than other seats(e.g. The odds ratio of death in driver's right seat is 0.02 (95% confidence interval[0.009,0.043]) times of Non-Motorist. The chance of fatality tends to increase with the age of injured persons(e.g. The odds ratio of death for 45-54 years old persons in 1.434 (95% confidence interval[1.282,1.604]) times of persons < 15., with other factors held consistent. We also included many other confounders like belt usage and extrication – mainly for vehicle occupant, within intercross mainly for non-vehicle occupant and alcohol test, which improve the model fit and prediction, but do not remove the significant relationsip between the variables of interest.

While we get compelling results from the analysis, our analysis is limited by the personal information provided by FARS. Other potentially important factors not included like general health condition, the type medical aid administered to rescuscitate, etc. can also contribute to the result of death from injuries. Despite these limitations, the factors of interest and confounders includeded still provide a good model fit and prediction. Our model may be of interest to people seeking to improve the survival of injured persons and to those estimating the level of severity and medical help needed (e.g., first responders) for a potentially fatal vehicle crash.

Reference

[1] U.S Departmetn of Transportation “2015” Page.URL:https://www.transportation.gov/fastlane/2015-traffic-fatalities-data-has-just-been-released-call-action-download-and-analyze

[2] Smith KM, Cummings P. Passenger seating position and the risk of passenger death in traffic crashes: a matched cohort study. Injury Prevention. 2006;12(2):83-86. doi:10.1136/ip.2005.010306.

[3] World Health Organization,Road traffic injuries Page.URL:http://www.who.int/mediacentre/factsheets/fs358/en/

[4] R Core Team (2012). ”R: A language and environment for statistical computing.” URL: http://www.R-project.org

[5] David A. Freedman (2009). Statistical Models: Theory and Practice. Cambridge University Press. p. 128.

[6] Geisser, Seymour (1993). Predictive Inference. New York, NY: Chapman and Hall. ISBN 0-412-03471-9.