HOSTED BY

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

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

Original Research Paper

Estimating likelihood of future crashes for crash-prone drivers

Subasish Das[*], *Xiaoduan Sun*[*], *Fan Wang*, *Charles Leboeuf*

Department of Civil Engineering, University of Louisiana at Lafayette, Lafayette, LA 70504, USA

ARTICLE INFO

Article history:
Available online 31 March 2015

Keywords:
Roadway safety
Crash-prone drivers
Crash risk
Logistic regression
Sensitivity

ABSTRACT

At-fault crash-prone drivers are usually considered as the high risk group for possible future incidents or crashes. In Louisiana, 34% of crashes are repeatedly committed by the at-fault crash-prone drivers who represent only 5% of the total licensed drivers in the state. This research has conducted an exploratory data analysis based on the driver faultiness and proneness. The objective of this study is to develop a crash prediction model to estimate the likelihood of future crashes for the at-fault drivers. The logistic regression method is used by employing eight years' traffic crash data (2004–2011) in Louisiana. Crash predictors such as the driver's crash involvement, crash and road characteristics, human factors, collision type, and environmental factors are considered in the model. The at-fault and not-at-fault status of the crashes are used as the response variable. The developed model has identified a few important variables, and is used to correctly classify at-fault crashes up to 62.40% with a specificity of 77.25%. This model can identify as many as 62.40% of the crash incidence of at-fault drivers in the upcoming year. Traffic agencies can use the model for monitoring the performance of an at-fault crash-prone drivers and making roadway improvements meant to reduce crash proneness. From the findings, it is recommended that crash-prone drivers should be targeted for special safety programs regularly through education and regulations.

1. Introduction

Based on highway crash reports, conservatively speaking, more than 50% of crashes each year are caused by human errors. Engineers are always trying to make roadways more forgiving and vehicles more crashworthy, which has made considerable impact on highway safety, in order to account for human error. Due to the persistent effort put forth by engineers, highway fatal crashes in the U.S. have finally reached the lowest number since 1960. Much of this effort has been spent on implementing crash countermeasures on highway facilities by enhancing the safety on roadway geometric features and traffic control devices. Safety education and enforcement, the other two elements in the 4E approach (emergency service is the fourth), also made strides in

* *Corresponding authors*. Tel.: +1 225 288 9875.
E-mail addresses: subasishsn@gmail.com (S. Das), xsun@louisiana.edu (X. Sun).
Peer review under responsibility of Periodical Offices of Chang'an University.
http://dx.doi.org/10.1016/j.jtte.2015.03.003

educating the general public on various safety risks and enforcing safety traffic laws.

In order to fulfill the hefty goal established by the American Association of State Highway and Transportation Officials (AASHTO) Highway Safety Strategy to cut traffic fatalities in half by 2020 and by Louisiana Strategic Highway Safety Plan for 'Destination Zero Deaths', it is important to have effective safety education and regulation programs while continually improving the highway infrastructure's safety. Since crash-prone drivers present a big adverse effect on highway safety, they should be effectively targeted in various safety education and enforcement programs. It is widely known that very young and very old drivers have the highest fatal crash rates, but this does not mean that these two groups commit the most crashes. People with similar personal traits could have very different levels of crash risk. Identifying high risk drivers and studying their characteristics are critical in order to further reduce the number of crashes through targeted safety education and enforcement programs.

Thus, a study was conducted at the University of Louisiana to study the impact of crash-prone drivers on safety and to predict how a driver's crash history could affect his/her crash occurrence(s) in the upcoming year. Logistic regression methods were used for the drivers with repeated crash involvements in previous years in order to establish relationships between driver responsibility and potential crash predictors. More importantly, the study was done to provide evidence for developing better and more efficient safety education programs and supporting targeted traffic laws or programs based on these crash over-involved drivers.

This paper begins with the review of earlier studies that have attempted to relate various variables to develop models for crash-prone drivers. This review is followed by a description of available data of nearly 2.08 million crash records for eight years' crash data. The next section provides discussion of model estimation results and its validation. In conclusion, an overall summary of findings on the model and their implications is given and some recommendations and direction for future research are provided.

2. Literature review

Investigating crash-prone drivers' characteristics, exploring the relationship between drivers' past crash/citation history and their crash risk, and predicting drivers' future crash occurrences from their previous crash history were the points focused on in many past studies.

The existence of crash-prone drivers was first recognized by Greenwood and Yule (1920). In their published paper, crash-prone drivers are defined as the drivers with a number of crashes higher than expected. In a study of Blasco et al. (2003), crash-prone drivers are described as drivers with recurring crashes that are caused by human error, not by coincidence. A study conducted by Peck et al. (1971), concludes that it is quite difficult to accurately identify which driver will or will not cause crashes because of the statistical nature of crash frequencies. After analyzing five years' crash data (1993–1997) in Kentucky, Stamatiadis et al. (1999) found that about 2.1% of licensed drivers who were charged with six or more points in the past 2 years accounted for nearly 5.3% of all crashes.

Predicting a driver's crash risk based on his/her past crash and traffic offence history is the topic of many investigations. The predictability of future crashes in terms of past violations or past crashes was investigated by Stewart and Campbell (1972). This study observed a four-year history of crash and violation records of North Carolina drivers to predict the future crashes. Through examining older drivers' previous conviction records and crash data, Daigneault et al. (2002) concluded that prior crashes would be a better predictor for crash risk than prior convictions. In a published study, Hauer et al. (1991) determined that the performance of their multivariate model for a crash would be improved by making right use of the driver's past crash records. A logistic regression model was developed by Chen et al. (1995) to identify crash-prone drivers based on their records prior to their at-fault crash involvements, which discovered that a model using prior at-fault crash data can recognize up to 23% more drivers who will have one or more at-fault crash involvements in the next 2 years than a model that uses the conviction information. After studying 17 logistic regression models, Gebers (1999) concluded that his models could correctly classify crash-involved drivers up to 27.6%. By deploying canonical correlation techniques in a subsequent research effort, Gebers and Peck (2003) achieved an accuracy level up to 27.2% from their best model to identify crash-prone driver. Chandraratna et al. (2006) studied Kentucky drivers to develop a crash prediction model that can be used to estimate the likelihood of a driver being at fault for a near future crash occurrence by using logistic regression technique. Although no model can be considered perfect, the modeling progress can be seen in research, especially in research from the Californian studies (Chen et al., 1995; Gebers, 1999). However, some researchers have voiced their skepticism over predicting crash-prone drivers (Gebers, 1999; Peck et al., 1971). In the recent years, research on at-fault drivers is becoming popular among researchers (Brar, 2014; Chandraratna and Stamatiadis, 2009; Currya et al., 2014; Goh et al., 2014; Greer et al., 2014; Harootunian et al., 2014; Karacasu and Er, 2011; Moghaddam and Ayati, 2014; Lee et al., 2014; Tseng, 2012; Yannis et al., 2005; Zhang et al., 2014).

In contrast with the previously published works focusing on human factors for the risk analysis of crash-prone drivers, this research also takes into account roadway and crash variables in order to get a better insight on the risky drivers' crash proneness.

3. Methodology

3.1. *Dataset*

The preliminary dataset was prepared from eight years (2004–2011) of crash data from Louisiana. It was arranged by merging three different tables (crash table, roadway table, and vehicle table) from the microsoft access dataset. For an individual crash record, a total of 371 crash attributes (possible explanatory variables) were collected. A total number of 2,076,009 crash records remained after deleting the records