

Identification of environmental factors affecting escape behavior in hit-and-run crash

Wei Bai

College of Transportation and Logistics, Southwest Jiaotong University, Chengdu 611756, China;
359954349@qq.com

Abstract

The hit-and-run crash has caused widespread concern because the crash can't be discovered and get the emergency rescue at the first time which increases the possibility of the crash injury severity. In order to identify the significant influencing factors of road environment characteristics on hit-and-run behavior, the paper employs the binary logistic regression model to examine 9 possible influential factors and partially crossing variables by the formation of dummy variables, based on the 2014 road traffic crash data of Michigan, USA. According to the definition of accident occurrence dominance ratio, the parameter estimation of the model is analyzed to distinct the significant factors. The results indicate that night time, crash at intersection & night time and night time & weekend have significant impacts on the occurrence of hit-and-run behavior, while crash occurred in rural road, speed limit, rural road & poor lighting, rural road & bad weather and rural road & weekend are associated with the less occurrence probability of hit-and-run significantly.

Keywords

Hit-and-run crash; Road environment; Logistic model; Factor identification.

1. Introduction

The hit-and-run behavior is regarded as one criminal action in many countries, it has caused widespread concern and study because the trouble-causing driver left the accident scene without reporting to police and giving the first aid, which caused the crash can't be discovered and get the emergency rescue at the first time after the accident. The hit-and-run behavior increases the possibility of the crash injury severity not only, but also brings the secondary damage easily due to the delay of rescue^[1]. Strengthen the research on hit-and-run crash, especially the study in characteristic factors about the road and environment can help to deepen the understanding of hit-and-run behavior and formulate perfect prevention measures.

At the present stage, more and more studies on the corresponding rules and characteristics of road traffic accidents from the perspective of data mining are gradually increasing at home and abroad^[2], but there are few researches on hit-and-run crash.

Arash M. Roshandeh(2016) analyzed the hit-and-run crash from the angle of driver's concentration, and found that the probability of hit-and-run behavior increasing when the driver in distracted driving^[3]. Chenming Jiang(2016) explored the characteristics of hit-and-run crash occurring in urban river-crossing highway tunnels, and demonstrated the significant accident rules under special road conditions^[4].Meiquan Xie(2018) adopted the real-time data detected to analyze the occurrence and injury severity of hit-and-run crash based on the Bayesian binary logit model with random effects^[5]. Dahianna Lopez(2018) identified factors related to a hit-and-run after a vehicle-bicycle collision with quantitative analysis ^[6].

In conclusion, there are few studies about hit-and-run crash at the present stage, and fewer papers identify the typical characteristic affecting the occurrence of hit-and-run behavior from the perspective of road or meteorological conditions. More quantitative research needed to reveal the intrinsic relationship between the occurrence of hit-and-run behavior and each influencing factors.

2. Methodology

The occurrence of road traffic accident is attributed to the interaction of many influence factors. It is difficult to obtain definite characteristic rules when considering a small number of research samples only. With the help of modern statistical theory and methods, the normal practice of study on road traffic safety is the statistical analysis through a large number of crash data to obtain the corresponding accident conclusions.

The logistic model is one generalized linear regression analysis model which can take the probability of different event results as the dependent variables for regression analysis. This model is suitable for quantitative analysis modeling, data mining, identification and prediction, etc. As one statistical analysis model, the logistic model is used widely in many scientific research. The model is detailed as follows:

$$\text{Logit}\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n \quad (1)$$

where P is the outcome probability of dichotomous variable, x_n is independent variable and β_n is the coefficient of each independent variable, β_0 is the constant of the equation.

It can be obtained from the formula (1):

$$P = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)} \quad (2)$$

$$1 - P = \frac{1}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)} \quad (3)$$

The dependent variable is a binary or polytomous variable in logistic regression. The response variable in this study is hit-and-run or non-hit-and-run crash. Therefore, the binary logistic regression is the most applicable method to use because it is developed to predict a binary dependent variable as a function of predictor variables. In order to clarify the influence of independent variables x_n on dependent variables, assuming independent variable x_n increases by one unit as other factors remain constant, the odds ratio (OR) of each independent variable as follows:

$$OR = \frac{\frac{\varepsilon_1}{1 - \varepsilon_1}}{\frac{\varepsilon_0}{1 - \varepsilon_0}} \quad (4)$$

This ratio indicates the relative amount by which the odds of the outcome increase or decrease. Where $\frac{\varepsilon_1}{1 - \varepsilon_1}$ stands for the ratio of the occurrence probability of hit-and-run behavior to the probability of non-hit-and-run behavior when $x_n = 1$, $\frac{\varepsilon_0}{1 - \varepsilon_0}$ denotes the same ratio when $x_n = 0$. The occurrence probability of hit-and-run behavior increases as $OR > 1$ and decreases when $OR < 1$ ^[7]. With the aim of identifying the influencing variables which affecting the occurrence of hit-and-run behavior from each characteristic factors, exploring the relevance of each independent variables from Table 1 when the p-value is set 0.5

The paper explores the influencing variables of hit-and-run crash occurrence from road condition, environment and other external factors based on the 2014 road traffic crash data of Michigan, USA. The crash data of Michigan records the whole information of each crash in detail, including the road condition, weather, lighting, time, injury severity and so on.

In 2014, there were 298,699 road traffic accidents occurred in Michigan, including 31,214 hit-and-run crashes and accounting for 10.45% of the total number of accidents. A total of 53,329 accidents resulted in personal injury or even death, of which 4,188 were hit-and-run crash, accounting for 7.85% of all casualty accidents. Excluding the missing data in the data sample, the quantity of complete

study sample is 297,870, including 30,987 hit-and-run crashes which accounting for 10.4% of the total sample.

The dependent variables in the crash data sample is hit-and-run or non-hit-and-run crash based on the research objectives determined in this paper.

The data of the research sample mainly includes the independent variables such as road hierarchy, speed limit, road condition, etc., as well as the meteorological condition, lighting condition, time of accident and other characteristic variables. Each variables are all in the form of dummy variables, which are described in detail as shown in table 1:

Table 1 Classification of virtual variable

Classification	Serial Number	Factors	Description of variables
Road conditions	X1	Road hierarchy	Rural road=1; Urban road=0
	X2	Speed limit	5mph~30mph=1; 31mph~50mph=2; 51mph~120mph=3
	X3	Road surface	Muddy road or other poor condition=1; Road surface is well=0
	X4	Road lane	Crash occurred in the road lane=1; Otherwise=0
	X5	Intersection	Crash occurred at intersection=1; Otherwise=0
Environmental factors	X6	Weather	Bad weather=1; Otherwise=0
	X7	Lighting	Poor lighting=1; Otherwise=0
Time factors	X8	Tome of day	Night time (00: 00~5:00&18:00~23:00)=1; Day time (6:00~17:00)=0
	X9	Weekend	Saturday & Sunday=1; Otherwise=0

3. Discussion

With the aim of significance requirement, there are 3 remarkable influence factors were selected from the 9 independent variables and 297,870 crash data based on the SPSS software platform. The results are shown in table 2:

Table 2 Parameter estimation result of univariate model

Variables	Parameter estimation	P-Value	Odds ratio
Constant	-1.237	<0.001	0.290
Night time	0.615	<0.001	1.850
Rural road	-1.173	<0.001	0.310
Speed limit			
31mph~50mph (1)	-0.943	<0.001	0.389
51mph~120mph (2)	-1.422	<0.001	0.241

There are two influencing variables related to road characteristics in the identified results of three significant univariate factors, rural road and speed limit. Among the two univariate factors, the estimated parameter value of rural road is -1.173, and the corresponding odds ratio is 0.310, which means that the probability of hit-and-run behavior occurrence is fewer in the traffic accidents on rural roads. The independent variable of speed limit need to be calibrated with dummy variables distinctively. Compared with the speed limit of 0mph~30mph, the estimated parameter value of 31mph~50mph and 51mph~120mph are -0.943 and -0.1.422 respectively, and the odds ratio are 0.389 and 0.241in turn. The identified results indicate that the higher speed limit the road is, the lower probability the hit-and-run crash occur, the tendency is more pronounced as the road speed limit

increasing. The estimated parameter value of the night time is 0.615, and the corresponding odds ratio is 1.850, which indicates that the hit-and-run behavior occurred at night is more likely than daytime. Basically, The identification results of univariate factors are consistent with the actual situation. The infrastructure of rural road is more primitive than the urban road and the driving speed of vehicles lower also. Therefore, the possibility of hit-and-run crash occurred in rural road is smaller. The better infrastructure conditions and the more monitoring equipment equipped in the road, the higher the road speed limit is. Moreover, most of the higher speed limit roads are closed passageways, the hit-and-run behavior will be confirmed timely under these circumstances. Therefore, the higher the road speed limit is, the less likely the hit-and-run crash take place. Traffic volume and driver's visibility are smaller at night which make the confirming of the hit-and-run crash difficultly, the probability of hit-and-run behavior occurred is higher than daytime. Consequently, with the aim of preventing the occurrence of hit-and-run behavior after the accident, it is necessary to strengthen the supervision of low speed road and night time and take this as the emphasis of control measures.

Considering the requirements of crash prevention, 9 independent variables including road infrastructure, meteorological conditions and time factors were cross-analyzed. A total of 5 cross-variables have significant influence on the hit-and-run behavior were identified, the results are shown in table 3:

Table 3 Parameter estimation result of cross variable model

Variables	Parameter estimation	P-Value	Odds ratio
Constant	-2.061	<0.001	0.127
Crash at Intersection & Night time	0.409	<0.001	1.506
Rural road & Poor lighting	-0.993	0.038	0.370
Rural road & Bad weather	-0.884	<0.001	0.413
Rural road & Weekend	-0.584	0.001	0.557
Night time & Weekend	0.709	0.001	2.032

As shown in table 3, Among the 5 influencing cross-variables, the estimated parameter value of crash at intersection & night time and night time & weekend are 0.409 and 0.709, the corresponding odds ratio are 1.506 and 2.032. The identification indicate that the hit-and-run crash is more likely to take place at intersection and weekend during the night time, the occurrence probability at weekend night is higher than intersection under night time. The cross-variables of rural & poor lighting, rural & bad weather, rural road & weekend reduce the hit-and-run crash occurring probability, each cross-variable's estimated parameter value is -0.993, -0.884 and -0.584 and odds ratio is 0.370, 0.413 and 0.557 in turn. That means the hit-and-run behavior is less likely to happen on weekend, bad weather and poor lighting during the rural road. The identification result s of cross-variables show are corresponding to the univariate factor, rural road and the night time are the significant variables of hit-and-run crash, which demonstrates that urban roads and night time are the key words of prevention in hit-and-run crash.

4. Conclusion

(1) In order to identify the influencing variables of hit-and-run crash from the road infrastructure, meteorological condition, etc., the logistic regression model was adopted to mine the crash data. The identification results show that the significant variables of hit-and-run crash are night time, rural road, speed limit, crash at intersection & night time, night time & weekend, rural road & poor lighting, rural road & bad weather and rural road & weekend. The influence effects of each significant variables and cross-variables are clarified by the analyzing in ratio odds.

(2) With the aim of reducing the occurrence of hit-and-run crash, lowering the crash injury severity and possibility of the secondary damage effectively, it is necessary to make the urban road, road with low speed limit and night time as the key emphasis in work for accident prevention.

(3) The paper identified the significant factors briefly which affecting the occurrence probability of hit-and-run crash from road infrastructure, meteorological conditions and time factors, the internal

influence mechanism of each factors were not discussed further. In the future, it is necessary to strengthen the study and exploration of hit-and-run crash from the perspective of data mining.

References

- [1] Traffic Management Bureau of Ministry of public security. Compilation of Road Traffic Safety Regulations (People's Public Security University Press of China, China 2009), p.473-492. (In Chinese)
- [2] Zhi Ye, Wang Dashan, Cong Haozhe, Rao Zhongbo. Road Traffic Accident Data Analyzing and Its Application: Example of Shenzhen, Urban Transport of China, Vol. 16(2018) No. 3, p.28-32. (In Chinese)
- [3] Arash M. Roshandeh, Bei Zhou, Ali Behnood. Comparison of contributing factors in hit-and-run crashes with distracted and non-distracted drivers, Transportation Research Part F, Vol. 38(2016), p. 22-28.
- [4] Chenming Jiang, Linjun Lu, Shengdi Chen, Jian John Lu. Hit-and-run crashes in urban river-crossing road tunnels, Accident Analysis and Prevention, Vol. 95(2016), p. 373-380.
- [5] Meiquan Xie, Wen Cheng, Gurdiljit Singh Gill et, al. Investigation of hit-and-run crash occurrence and severity using real-time loop detector data and hierarchical Bayesian Binary Logit model with random effects, Traffic Injury Prevention, Vol. 19(2016)No. 2, p. 207-213.
- [6] Dahianna Lopez, Mark E. Glickman, Stephen B. Soumerai et, al. Identifying factors related to a hit-and-run after a vehicle-bicycle collision, Journal of Transport & Health, Vol. 8(2018), p. 299-306.
- [7] Li Ming, Wang Yonggang, Zhang Wei, Wang Ying. Identifying significant factors influencing occurrence of fatal crash on mountainous freeway, China Safety Science Journal, Vol. 25(2016) No. 5, p. 126-130. (In Chinese)