A LOGISTIC MODEL OF HIT-AND-RUN CRASHES IN CALGARY

Huafei Sun, Graduate Student
Lina Kattan, Urban Alliance Professor
Richard Tay, AMA Chair in Road Safety
Department of Civil Engineering, University of Calgary

Abstract

Hit-and-run cashes refer to traffic collisions in which at least one driver flees from crash scene without reporting the crash. In the City of Calgary, hit-and-run cashes accounted for 18 percent of total traffic collisions in 2005. The objective of this study is to identify the environment and road characteristics that contribute to the occurrence of hit-and-run crashes in the City of Calgary. A logistic regression model was developed to delineate the likelihood of hit-and-run crashes as opposed to non hit-and-run crashes. The results showed that compared to weekday and daytime collisions, weekend and night time collisions have significantly higher likelihood of hit-and-run. In terms of weather condition, clear weather exhibited the greatest chance of hit-and-run when compared to any other weather conditions. Moreover, hit-and-run crashes are quite likely to occur on undivided one-way roads and the roads with artificial light. As for driver related factors, drivers aged at 55 or above showed the greatest likelihood as compared to other age groups.

Keywords: Hit-and-Run, Logistic Model, Calgary

Introduction

Road crashes are one of the major contributing causes of loss of life, disability, property damage and other social cost. In Canada, there were 2889 road users killed in 2006 and additional 199,337 injured in crashes (Transport Canada, 2006). In Alberta, there were 142,592 motor vehicle collisions in 2006, resulting in 453 road users killed and 25,964 injured (Alberta Transportation, 2006).

Among the different types of crashes, hit-and-run crashes are of special concern to road professionals and law enforcement. A hit-and -run crash is one in which at least one of the drivers leaves the scene of accident without reporting the crash. Leaving the scene without reporting the crash may result in an increase in the severity of the crash since 85 percent of victims in fatal crash would die within one to two hours (Roess et al, 2005; Tay et al, 2008, forthcoming). Hence, hit-and-run is a punishable offence in many countries. In addition, hit-and-run crashes could also increase the investigation cost arising from finding out the fleeing offender and eyewitness of the accident.

The prevalence of hit-and-run behavior varies across jurisdiction depending on the social and road environment as well as law enforcement. For example, hit-and-run constitute only about 1.9% of the total reported crashes in Singapore (Tay et al, 2008) but 8.1% in California, which is the highest among all the states in America (Tay et al, forthcoming). However, in the City of Calgary, hit-and-run crashes constituted 17.8% of total number of crashes in 2005. This very sizeable share of hit-and-run crashes makes Calgary an interesting case study that may advance our knowledge in this area.

The objective of this paper is to examine the impact of road features, environment and driver related factors on the likelihood of occurrence of hit-and-run accidents in the city of Calgary. Using the 2005 Calgary accident data, a logistic regression model is constructed to identify the factors contributing to the likelihood of hit-and-run in the event of a crash.

This paper is organized as follows: a brief review of the literature is first presented, followed by a description of the data and methodology. Then, the results of the analysis are discussed. Concluding comments are presented in the final section.

Literature Review

The topic of hit-and run has attracted a number of researchers from diversified disciplines, namely, engineering, law, as well as medical and social-science. In the area of law and medical sciences, the research is mainly focused on examining the severity and type of victim injury to identify the type of vehicles and infer the collision point on the vehicles (Karger, B. et al, , 2001; Teresinski, G.,2001).

In the area of transportation engineering and planning, a few studies have been conducted to identify the effects of roadway, environment, weather conditions and lighting in influencing the decision of running away from the crash scene. In one of the study, Tay et al (2008) identified three general attributes which have significant effect on hit-and-run crashes in Singapore; namely, night time, lighting condition, low density neighborhoods.

In a study on hit-and-run crashes resulting in pedestrian fatalities, Slonick and Hemenway(1995)examined the effects of victim's characteristics, driver's characteristics and circumstances of collision on the decision of running away from the crash scene. Their study showed that a driver is less likely to run if the crash victim is a child or senior pedestrian and if the crash occurred in daylight and if the driver is senior. In a previous study, the same authors found that drunk driving was a significant determinant of hit-and-run behavior (Slonick and Hemenway, 1994).

In another study, Kim et al (2008) used a rough set analysis tools combined with logistic regression to understand the key human factors as well as roadway features associated with hit and run collisions in Hawaii. The authors found that factors such as being a

male, tourist, intoxicated, and driving a stolen vehicle are strong predictors of hit and run crashes.
Finally, Tay et al (forthcoming) showed that roadway functional class, routes, traffic flow, types of roadway section, speed limit, traffic control device, functioning of traffic control device, lighting condition, roadway alignment, roadway profile, weekend and night time are all important determinants of hit-and-run behavior in the event of a crash.

Methodology

In this study, the response variable, hit-and-run, is a binary variable. The logistic model is applicable to such situation in which dependent variable is dichotomous while independent variables can be interval level or categorical. The logistic model is appropriate for this study because the response variable has two possible outcomes:

Y = 1 indicates a hit-and-run crash
Y = 0 indicates a non hit-and-run crash

The formulae of the probability P of a hit-and-run is given by:

$$Prob.(Y_i = 1) = 1/(1 + e^{-\beta X})$$
$$Prob.(Y_i = 0) = e^{-\beta X}/(1 + e^{-\beta X})$$

Where: β is a vector of parameters to be estimated and X is a vector of explanatory variables. The parameter vector β is estimated by SPSS statistical software using maximum likelihood estimator (MLS) technique. The sign and magnitude of β are used to explain the effect of independent variables on likelihood of occurrence.

The logistic model can also be used to estimate the odd ratio of the probability of occurrence of hit-and-run to the probability of not hit-and-run as a function of the dependent variables. In this study, the logit model is the natural logarithm of the probability ratio that the dependent variable is 1 (hit-and-run) relative to 0 (not hit-and-run).

$$Logit(\text{OR}) = \ln\left(\frac{P}{1-P}\right) = \beta X$$

Holding all the other factors constant, the marginal effect arising from a unit change of an independent variable x_i in the logit model is known as the odd ratio, which is quantified by the factor $\exp[\beta_i]$. If $\beta_i > 0$, then OR > 1, indicating that the probability of occurrence of hit-and-run will increase due to a unit change in x_i.

Data

Data on crashes occurring in the City of Calgary in 2005 were extracted from the official crash database maintained by Alberta Transportation. There were a total number of 37,146 reported crashes and 6,636 of them were hit-and-run crashes, comprising about 17.8%. Note that only 36,936 observations were used to estimate the logistic regression model because some observations with missing key data were excluded.

As shown in Table 1, the explanatory variables categories or contributing factors that were examined included: day-of-week, time-of-day, driver gender and age, vehicle's age, total number of vehicles involved, crash severity, weather conditions, as well as road environment characteristic such as road class, lighting condition. Note that most of these factors were recorded using ordered or unordered categories and therefore several dichotomous variables were created in the model to capture each contributing factor. Most of the variables are self-explanatory.

Table1. Descriptive Statistics of Variables

Variables	**Mean**	**Std. Dev.**
Day of week		
Weekday	.69	.462
Weekend	.31	.462
Time of Day		
Day Time	.88	.328
Night Time	.12	.328

Crash Severity		
Injury/Fatality	.10	.296
Non Injury	.90	.298
Road Class		
Undivided one-way	.04	.201
Undivided two-way	.23	.419
Divided with barrier	.16	.368
Divided no barrier	.03	.178
Environment condition		
Clear	.73	.443
Raining	.04	.204
Hail/Sleet/Snow	.04	.200
Others or unknown	.18	.387
Lighting Condition		
No artificial light	.53	.499
Artificial Light	.20	.402
Unknown	.25	.435
Vehicle Age		
Age < 15	.74	.436
Age ≥ 15	.12	.326
Unknown v	.13	.342
Number of Vehicle		
One vehicle	.15	.360
Two or more vehicles	.85	.360
Age of Male Drivers		
Age ≤ 30	.20	.399
30 < Age < 55	.26	.436
Age ≥ 55	.08	.277
Age of Female Drivers		

Age ≤ 30	.11	.313
30 < Age < 55	.15	.359
Age ≥ 55	.05	.209
Unidentified Drivers	.14	.347
Surface condition		
Dry road surface	.60	.490
Wet road surface	.08	.264
Slush/Snow/Ice	.15	.355

Note that to avoid perfect multicollinearity, one category was arbitrarily chosen as the base or reference case and omitted in the estimation. The estimates obtained for the other categories will then be interpreted as the relative effect of the variable compared to the base case. The above model was estimated using SPSS.

Following Kockelman and Kweon (2002), Tay & Rifaat (2007) and Tay et al (2008, forthcoming), some categorical variables that were statistically insignificant were retained in the model as long as one of the categories for the same factor was statistically significant. Hence, a more conservative critical value of $\alpha = 0.10$ was used.

Discussion of Results

The estimation results are reported in Table 2. In general, the model fitted the data quite well, with a large chi-square statistic and very small p-value for the goodness-of-fit.

Table 2: Estimation of Results

Number of Observations: 36,936 Chi-Square Statistics: 181.622 P-value: 0.000		
Variables	**Odd Ratio**	**P-Value**
Day of week (Ref: Weekday)		

Variables	Odd Ratio	P-Value
Weekend	1.148	<0.001
Time of Day (Ref: Daytime)		
Night	1.233	<0.001
Crash Severity (Ref: non injury)		
Injury/Fatality	1.086	0.070
Road Class (Ref: Divided with Barrier)		
Undivided one-way	1.187	0.009
Undivided two-way	1.072	0.035
Divided no barrier	1.096	0.227
Environment condition (Ref: Hail/Sleet/Snow)		
Clear	1.150	0.078
Raining	0.968	0.788
Unknown	1.113	0.219
Lighting Condition (Ref: no artificial light)		
Artificial Light	1.227	<0.001
Unknown	1.057	0.124
Vehicle Age (Ref: age <15)		
Vehicle age≥15	1.058	0.179
Unknown	0.825	0.005
Number of Vehicles (Ref: One Vehicle)		
Two or more vehicles	0.981	0.609
Age of Male Drivers (Ref: 30 < Age < 55)		
Age ≤ 30	1.032	0.402
Age ≥ 55	1.081	0.131
Age of Female Drivers (Ref: 30 < Age < 55)		
Age ≤ 30	1.007	0.873
Age ≥ 55	1.162	0.022
Unidentified Drivers	1.305	<0.001
Surface condition (Ref: Dry)		

Variables	Odd Ratio	P-Value
Wet road surface	1.048	0.532
Slush/Snow/Ice	0.858	0.001
Constant	0.163	<0.001

Occurrence time and day

As shown in Table 2, crashes occurring at night were more likely to be associated with hit-and-hit. The odd ratio for night time was 1.23 which implied that likelihood of hit-and-run crashes at night was 23% greater than that in daytime. This result is expected since drivers might think that they are less likely to be caught because of the relatively low traffic volume and poor visibility at night (Tay et al, 2008, forthcoming; Kim et al, 2008).

Our model also suggested that weekend was associated with a 15% higher likelihood of hit-and-run crashes than weekdays. This result can be explained by the low traffic volume in weekend as compared to weekdays and is consistent with those obtained by Tay et al (forthcoming).

Lighting conditions

Our study also found that lighting condition was a crucial factor for hit-and-run crashes in the City of Calgary. The model indicated that the likelihood of occurrence of hit-and-run crashes at location with artificial light was 23% greater than that that at location with no artificial light. Even though contrary to expectation, this finding is consistent with Tay et al (forthcoming) who found that likelihood of hit-and-run is higher at night on lighted roads compared to unlit roads although the difference is not statistically significant.

Accident Severity

As only a few fatal accidents occurred in 2005, only two severity related variables are presented under this category in which property damage only (PDO) was chosen as the reference variable. Injury/fatality involved accidents showed an OR value of 1.09 along

with its significant level of 0.07. This result revealed that drivers exhibit a higher chance of running away when victim sustained the injury in crash. This finding is consistent with the results obtained by Tay et al (2008).

Driver's Gender and Age

Driver's gender and age was investigated in this study using middle-aged drivers as the reference category. Consistent with previous studies (Slonick and Hemenway, 1995; Tay et al, 2008), drivers aged 30 or below and 55 or above both showed higher chances of running after the crash than their middle age counterpart.

Road Class

Divided road with barrier was selected as the reference variable under this category. Our model revealed that all other types of roads had the higher OR than divided road with barrier. In particular, undivided roads and roads with no barrier were associated with an increased in the probability of running after the crash. These results are consistent with those obtained by Tay et al (forthcoming).

Environmental Condition

Consistent with previous studies (Tay et al, 2008, forthcoming), clear weather showed the higher likelihood of hit-and-run than snowy weather and snowy/icy road conditions are less likely to result in hit-and-run than dry roads.

Vehicle Number and Age

Our study also found that single vehicle crashes have a higher likelihood of hit-and-run. Also, drivers of vehicles that were 15 years or older exhibited a 6% greater chance of running away. This variable may indirectly reflect some socioeconomic characteristics of the drivers such as their income and educational level. Thus, lower income and less educated drivers are more likely to run away from the crash scene without reporting the accident. This behavior might be explained by their perceived inability to pay for the damages.

Conclusion and Recommendation

In this study, a logistic regression model is applied to identify the factors that contributed to the occurrence of hit-and-run crashes relative to non-hit-and-run crashes in the City of Calgary. It should be noted that the conclusions from this study are related to the situation in the City of Calgary although the outcomes from this study are consistent with the other hit-and-run studies conducted in Singapore, Hawaii and California.

Our study identified night time, weekend, aging drivers, crashes involving injury/fatality, clear weather and dry road condition, roads with artificial light, divided roadways and roads with no barriers all contributed to hit-and-run behavior.

References

Karger B, Teige K, Fuchs M & Brinkmann B (2001) Was the pedestrian hit in an erect position before being run over? *Forensic Science International*, 119, 217-220

Kim K, Pant P & Yamashita E (2008) Hit-And-Run Crashes: Using Rough Set Analysis with Logistic Regression to Capture Critical Attributes and Determinants. CD-ROM 87th Annual Meeting of the Transportation Research Board, Washington DC

Kockelman K & Kweon Y (2002) Driver injury severity: an application of ordered probit models, *Accident Analysis and Prevention*, 34, 313–321

Roess P, Prassas E & McShane W (2005) *Traffic Engineering*, Third Edition, Prentice Hall

Solnick S & Hemenway D (1995) The hit-and-run in fatal pedestrian accidents: victims, circumstances and drivers, A*ccident Analysis and Prevention*, 27(5), 643-649

Solnick S & Hemenway D (1994) Hit the bottle and run: the role of alcohol in hit-and-run pedestrian fatalities, Journal of Studies in Alcohol, 55(6), 679-684

Tay R, Barua U & Kattan L (forthcoming) Factors Contributing to Hit-and-Run in Fatal Crashes, *Accident Analysis and Prevention*

Tay R, Rifaat S, Chin H (2008) A logistic model of the effects of roadway, environmental, vehicle, crash, and driver characteristics on hit-and-run crashes, *Accident Analysis and Prevention*, 40(4), 1330 -1336, 2008

Tay R & Rifaat S (2007) Factors Contributing to the Severity of Crashes at Intersections, *Journal of Advanced Transportation*, 41(3), 245-265

Teresinski G & Madro R (2001) Knee joint injuries as a constructive factor in car-to-pedestrian accidents, *Forensic Science International*, 124, 74-82