

## Analysis of left-turn crash using Poisson Regression model at unsignalized intersections

Liyan Qin<sup>1,a</sup>, Yan Lu<sup>2,b</sup>, Shuai Zhang<sup>1,c</sup>

<sup>1</sup>School of Mechanical and Automotive Engineering, Xiamen University of Technology, LR 600, P.R.C.

<sup>2</sup> Department of Transportation Studies, Texas, 3100 Cleburne Street, United States

<sup>a</sup>2013110802@xmut.edu.cn, <sup>b</sup>ye19962013@163.com, <sup>c</sup>163101020@xmut.edu.cn

**Keywords:** unsignalized intersection; left-turn crash; Poisson regression model; short left lane

**Abstract:** The purpose of this study is to explore the safety performance of short left-turn lanes at unsignalized intersections, where it is often impractical to provide the standard lengths either when the available length between two adjacent opening is inadequate or when heavy left-turn volumes leads to demanding lane length. Left-turn crash data of six years from Houston, Texas of the years 2006-2011 were collected. There were thirty-nine lanes shorter than the standard requirements and thirteen lanes meeting the requirements. A Poisson Regression model was developed to relate traffic and geometric attributes to the total count of rear-end, sideswipe, and object-motor vehicle accidents at a left-turn lane. Furthermore, Crash Modification Factors (CMFs) were calculated for future applications in projecting the crash frequency given a specific change of the lane length. The results shows that short left-turn lanes have negative safety impacts. The study finally puts forward the opinion to improve the present conditions at the left-turn unsignalized intersection.

### Introduction

Intersections are among the most dangerous locations of a roadway network. In the state of Texas, 40.1% of fatalities and serious injuries occurred at or were influenced by intersections [1]. In the U.S., there are only about 10% of all intersections are signalized, in 2005[2]. Though Left-turn lane was used to eliminate or decrease crashes occur frequently, they account for a high percentage of total crashes at signalized and unsignalized intersections. In Houston, 58.3% of left-turn crashes involved injury, and there are only around 90% intersections are unsignalized. Therefore, analyzing left-turning traffic at unsignalized intersection is crucial to improve intersection operation and safety.

The left-turn crashes at unsignalized intersections result in a huge cost to society in terms of injury, death, and property damage. From 2006, many crash studies have been conducted. In one of the studies, Wang and Abdel-Aty indicate there are obvious different factors affecting the different left-turn collisions [3]. The NCHRP Report introduced left-turn lanes at unsignalized intersections generally led to a consistent reduction by 50% to 77% [4]. Highway Safety Manual indicate that, on average, left-turn lanes can reduce total crashes by 47% on two-lane streets and 27% on four-lane streets in urban and suburban areas [5]. Fitzpatrick et al. provided equations for predicting total crashes with and without left-turn lanes installed at unsignalized locations [6].

These studies emphasize on the importance of the use of left-turn, ignore how to use the length of the appropriately. The purpose of this study is to investigate of using short left-turn lanes of substandard lengths, based on the historical crashes records. The research results will provide traffic engineers/designers with a better understanding of the safety concerns related to short left-turn lanes at unsignalized intersections and assist them in designing and installing short left-turn lanes appropriately.

## Methodology for safety impact analysis

This research analyzed the three models and made the choosing procedure.

### Model Choice

#### Poisson Regression Model

In a Poisson regression model, the probability of intersection  $i$  having  $y_i$  accidents per year is given by

$$P(Y_i = y_i) = \frac{\text{EXP}(-I_i) I_i^{y_i}}{y_i!} \quad (1)$$

where  $I_i$  is the Poisson parameter for intersection  $i$ , which is equal to the expected number of accidents per year at intersection  $i$ .

$$I_i = \text{EXP}(b X_i) \quad (2)$$

where  $X_i$  is a vector of explanatory variables and  $b$  is a vector of estimable parameters. In this formulation, the expected number of events per period is given by

$$E[Y_i] = I_i = \text{EXP}(b X_i) \quad (3)$$

#### Negative Binomial (NB) Regression Model

When in incident frequency data is overdispersed, the Poisson regression model cannot be employed. Then a negative binomial regression model is commonly used. The negative binomial model is derived by rewriting Equation 2 such that, for each observation  $i$ ,

$$I_i = \text{EXP}(b X_i + e_i) \quad (4)$$

where  $\text{EXP}(e_i)$  is a gamma-distributed error term with mean 1 and variance  $a^2$ . The addition of this term allows the variance to differ from the mean as below:

$$\text{VAR}[Y_i] = E[Y_i][1 + a E[Y_i]] = E[Y_i] + a E[Y_i]^2 \quad (5)$$

The negative binomial distribution has the form:

$$P(Y_i = y_i) = \frac{\Gamma((\frac{1}{a}) + y_i)}{\Gamma(\frac{1}{a}) y_i!} \left(\frac{1/a}{(\frac{1}{a}) + I_i}\right)^{1/a} \left(\frac{I_i}{(\frac{1}{a}) + I_i}\right)^{y_i} \quad (6)$$

#### Zero-Inflated Poisson and Negative Binomial Models

By investigating incident count data, it was found that for some roadway segments no incident occurred in many days. To handle the zero-inflated count data, two zero-inflated models, i.e., the zero-inflated Poisson (ZIP) and negative binomial (ZINB) model, are developed.

The ZIP model assumes that the events,  $Y = (y_1, y_2, \dots, y_n)$ , are independent and the model is

$$\begin{aligned} y_i = 0 & \text{ with probability } p_i + (1 - p_i) \text{EXP}(-I_i) \\ y_i = y & \text{ with probability } \frac{(1 - p_i) \text{EXP}(-I_i) I_i^y}{y!} \end{aligned} \quad (7)$$

where  $y$  is the number of events per period.

The ZINB regression model follows a similar formulation with events,  $Y = (y_1, y_2, \dots, y_n)$ , independent and the model is

$$y_i = 0 \text{ with probability } p_i + (1 - p_i) \left[ \frac{a^{-1}}{a^{-1} + I_i} \right]^{a^{-1}}$$

$y_i = y$  with probability

$$(1 - p_i) \left[ \frac{\Gamma(a^{-1} + y) u_i^{a^{-1}} (1 - u_i)^y}{\Gamma I^{-1} y!} \right], y = 1, 2, 3, \dots \quad (8)$$

Zero-inflated models imply that the underlying data-generating process has a splitting regime that provides for two types of zeros. Splitting model  $p_i = \text{Logistic}(tb'X_i)$ . Which is a logistic cumulative distribution function. The maximum likelihood estimates can be used to estimate the parameters of the zero-inflated models, and confidence intervals can be constructed by likelihood ratio tests.

The testing of whether a zero-inflated incident state is more appropriate than the non-zero-inflated incident state is complicated by the fact that the zero-inflated model is not nested within either the Poisson or the negative binomial models. The Vuong's statistic for testing the non-nested hypothesis of a zero-inflated model versus a traditional model is [9]:

$$v = \frac{\sqrt{n} \left[ (1/n) \sum_{i=1}^n m_i \right]}{\sqrt{(1/n) \sum_{i=1}^n (m_i - \bar{m})^2}} = \frac{\sqrt{n}(\bar{m})}{S_m} \quad (9)$$

$$\text{Where: } m_i = \log \left( \frac{f_1(y_i | X_i)}{f_2(y_i | X_i)} \right),$$

$f_1(y_i | X_i)$  = the probability density function of the zero-inflated model,

$f_2(y_{ij} | X_{ij})$  = the probability density function of either the Poisson or negative Binomial distribution,

$\bar{m}$  = the mean of  $m_i$ ,  $S_m$  = the standard deviation of  $m_i$ ,  $n$  = the sample size.

Asymptotically, the Vuong's statistic denoted as  $v_{ZIP}$  and  $v_{ZINB}$  for ZIP and ZINB models, respectively. The Vuong's statistic is distributed as standard normal, so its value can be compared to the critical value of the standard normal distribution, e.g., 1.96. The test is directional, i.e., values larger than 1.96 favor the zero-inflated model while values less than -1.96 favor the Poisson or negative Binomial regression models.

### Safety analysis model

Based on the introduction to these models, a property frequency model can be chosen by following process. The model choosing starts with estimation of a NB model. A decision is made on whether to fit a ZINB regression model or a ZIP model based on a t-test on  $a$ , as expressed in Equation 6. If the t-test statistic is greater than a threshold, e.g. 1.65 with 0.1 significant level, a ZINB model needs to be fitted. Following this point, a test based on the Vuong's statistic needs to be conducted. If the Vuong's statistic is greater than a chosen threshold, an additional test needs to be performed based on  $t_{a_{ZINB}}$ . Otherwise, it can be determined that NB is the right model for the incident frequency data. If  $t_{a_{ZINB}}$  is greater than a given threshold, it is the ZINB model that needs to be fitted for the incident frequency data. Otherwise, the ZIP model should be chosen for the data. Return to the start point, the fitted NB model, if the t-statistic  $t_{a_{NB}}$  is smaller than the chosen threshold, a ZIP model needs to be fitted. Then, a test based on Vuong's statistic needs to be performed. If it is greater than a given threshold, the final chosen model should be ZIP. Otherwise, it should be a Poisson model. In summary, there are two paths to select the ZIP model, one is on the left through the fitting of a ZINB model, and the other is on the right through the fitting of a ZIP model. As far as Poisson and NB models are concerned, there is only one path to reach them.

### Crash Modification Factor

A crash modification factor (CMF) is used to compute the expected number of crashes after implementing a given change at a specific site. A CMF greater than 1.0 indicates an expected increase in crashes, while a value less than 1.0 indicates an expected reduction in crashes after implementation of a given countermeasure. For example, a CMF of 0.8 indicates an expected safety benefit, a 20 percent expected reduction in crashes.

In this study, a crash modification factor (CMF) was developed for the total number of related crashes (i.e., rear-end, sideswipe, and OMV crashes) at median left-turn lanes. As an indicator of crash potential, the mathematical expectation (i.e., mean value) given a specific lane length was used to formulate the CMF as Equation 8. In the calculation, the base case represented a lane that is equal to the Green Book length.

$$CMF(L_i = x) = \frac{E(y_i | L_i = x)}{E(y_i | L_i = 0)} = \frac{\sum_k k \cdot \Pr(y_i = k | L_i = x)}{\sum_k k \cdot \Pr(y_i = k | L_i = 0)} \quad (10)$$

### Calculation result

#### Data preparation

In this research, a historical crash data analysis was conducted from January 2006 to December 2011. Historical crash records were obtained from Texas Crash Record Information System (CRIS). Fifty-two median left-turn lanes were selected in Houston. The lengths of the median left-turn lanes studied spanned from 140 feet to 450 feet, all located at four-leg unsignalized median openings.

Fig. 1 presents the locations of the studied lanes. For each of the lanes, the Green Book standard was used to calculate the required length, given the observed left-turn volume. Thirty-nine of the lanes studied are shorter than the requirements, while thirteen lanes meet the requirements. Among the studied locations, the posted speed limits ranged from 30 to 40 mph. The left-turn volumes were observed for PM peak hours on weekdays during April to June 2013, and the peak-hour left-turn volumes spanned from 2 to 162 vph with percentages of heavy vehicles ranging from 0-25%.

Average daily traffic counts were retrieved from records available at the City of Houston and the TxDOT. It should be noted that, once a left-turning vehicle departs from a median left-turn lane, the lane finishes serving its purpose, and thus, the opposing traffic volume should be excluded from analyzing those crashes attributed to a short left-turn lane (i.e., three types of crashes identified in the following section). Therefore, traffic volumes (expressed in vehicles per day per lane or vpdpl) were only prepared for the direction, in which the left-turns travel at the studied left-turn lanes.

Short Left-Turn Lanes				Full-Length Left-Turn Lanes			
Number of Sites	Street	Number of Lanes	Speed Limit, mph	Number of Sites	Street	Number of Lanes	Speed Limit, mph
6	Bellaire Blvd	6-lane	35	1	Hillcroft St	8-lane	35
4	Kirby Dr	4-lane	35	8	Westheimer Rd	8-lane	35
3	Kirby Dr	6-lane	35	1	S Main St	8-lane	40
4	Richmond Ave	6-lane	35	1	Holcombe Blvd	6-lane	30
3	Gulfton Dr	4-lane	30	2	Old Spanish Trail	6-lane	35
3	Renwick Dr	4-lane	35				
4	Blodgett St	4-lane	35				
11	Westheimer Rd	8-lane	35				
1	Beechnut St	6-lane	35				
Total Number of Sites = 39				Total Number of Sites = 13			

**Fig. 1. Study locations in Houston**

For each crash record, the data specified the location (in a format of GIS coordinates and street numbers), severity (e.g., fatalities, injuries, and property damage), crash type (e.g., the relative position, angle of involved vehicles, and contributing factors), and other information (e.g., time, weather, lighting conditions, condition of the surface of the road, and traffic control).

Due to short left-turn lanes, crashes may happen for the following reasons: (1) an unfavorably large speed differential between a turning vehicle and the follow-up vehicle (i.e., either a through or a turning vehicle), (2) a deceleration length insufficient for a left-turning vehicle to stop, or (3) overflowed turning vehicles stacking in through-traffic lanes. Thus, relating to the lengths of left-turn lanes, three types of crashes were identified and analyzed

#### Model calculation results

Using the data acquired from the 52 studied lanes, a series of preliminary tests were performed by fitting the data into 1) Poisson regression model, 2) zero-inflated Poisson (ZIP) regression model, 3) negative binomial (NB) regression model, and 4) zero-inflated negative binomial (ZINB) regression model, respectively. Following a sequential procedure of model choice presented in Chapter 3, the preliminary tests evidenced that a Poisson regression model should be selected over the other options in representing the relationship between the attributes and the crash count for a median left-turn lane. In addition, overdispersion tests indicated that we could not reject the null hypothesis of equidispersion at a confidence level of 95%, which further justified the use of Poisson regression modeling approach.

Using the fifty-two data samples, maximum-likelihood estimation (MLE) was used to estimate the coefficients  $b$  in the model, and the outcomes are presented in Table 1.

**Table 1 Calibrated coefficients for the model**

Parameter	Coefficient $b_j$	Standard Error	Z-Statistics	p-value	95% Confidence Interval	
Intercept	-2.9155	0.7437	-3.92	.0001	-4.373	-1.458
Directional ADT volume per lane ( $v_i$ )	0.2208	0.0784	2.82	.0048	0.067	0.374
Relative lane length ( $L_i$ )	-4.1993	1.3227	-3.17	.0015	-6.792	-1.607

Therefore, the final model can be expressed as:

$$E(y_i|X) = e^{(-2.9155 + 0.2208 \times V_i - 4.1993 \times L_i)} \quad (11)$$

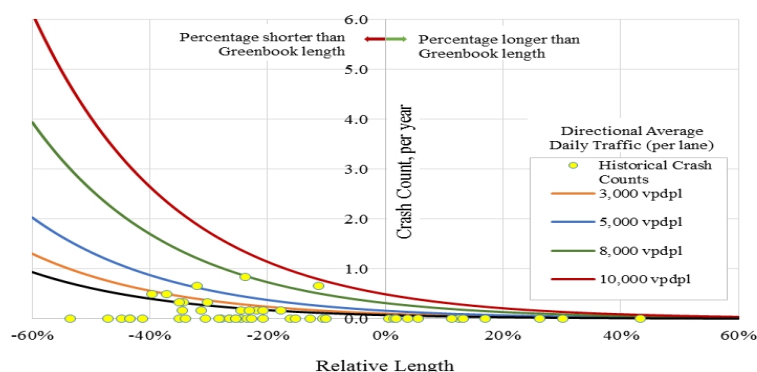


Fig.2. Fitness of Historical Crash Counts into Crashes Predicted by Poisson Regression

The final model included relative length of left-turn lane as a statistically significant predictor ( $p$ -value = 0.0015). Generally, the extent to which a median left-turn lane follows Greenbook recommendations had significant effects on safety performance at unsignalized median openings, i.e., longer lanes that better follows the recommendation generally led to better safety performance. In addition, the final model included the directional average daily traffic volume of the street (vpdpl) in the direction that the studied left-turns traveled. Given the same lane length, higher volumes were associated with more interactions between through traffic and left-turning vehicles that decelerated in preparation for entering median left-turn lanes, which led to a higher crash potential.

Fig.2 shows the fitness of historical data into the likelihood of crashes calculated by Equation 11.

## Summary and discussion

This paper presents a series of crash models to investigate the safety impacts of short left-turn lanes. By using 52 sample lanes, the main conclusions are that: First, Poisson regression model is the better model to analyze the accident occurred in left-turn lanes. Second, short left-turn lanes do have negative safety impacts. When it is impractical to provide the recommended length, short left-turn lanes might be acceptable in some particular cases, in which engineers' judgments are needed for a trade-off decision accounting for traffic volume, crash potential, mobility, accessibility, and economic and social impacts in determining whether a short left-turn lane is appropriate. In addition directional daily traffic volumes have significant effects on the accident potential; Third, CMFs can be used to predict accidents frequency if shorten or lengthen left-turn lane. The study indicates that left-turn lane with shorter length, the higher crash potential the left-turn lane would be.

Although the outcomes of this study provided important understanding of the safety performance of short left-turn lanes at unsignalized intersections, the results may be limited in scope and applicability due to the limited sample size involved.

## Acknowledgements

This research was financially supported by Ximan University of Technology Program of Natural science foundation (Grant No:YKJ13020R), Science & Technology Innovation Project of Fujian Province, China ( Grant No:2016H2003 ) .

## References

- [1] Texas Department of Transportation. Roadway Design Manual, March 2011
- [2] Abdel-Aty, M. Temporal and spatial analyses of rear-end crashes at signalized intersections. *Accident Analysis and Prevention* 38 (6), 1137–1150 ( 2006).
- [3] XuesongWang\*, Mohamed Abdel-Aty. Analysis of left-turn crash injury severity by conflicting pattern using partial proportional odds models. *Accident Analysis and Prevention* 40(January (1)), 1674-1682 (2008).
- [4] Gluck, J., H.S. Levinson, and V. Stover. NCHRP Report 420: Impacts of Access Management Techniques. Transportation Research Board, National Research Council (1999).
- [5] Highway Safety Manual, First Edition. American Association of State Highway and Transportation Officials. Washington, D.C.(2010).
- [6] Fitzpatrick, K., M.A. Brewer, J.S. Gluck, W.L. Eisele, Y. Zhang, H.S. Levinson, W. Zharen, M.R. Lorenz, V. Iragavarapu, E.S. Park. Development of Left-Turn Lane Warrants for Unsignalized Intersections, NCHRP Web-Only Document 193 - Final Report for NCHRP Project 03-91 (2010).
- [7] Wang, X., Abdel-Aty, M. Right-angle crash occurrence at signalized intersections. In: *Transportation Research Record: Journal of the Transportation Research Board* No. 2019. TRB, National Research Council, Washington, DC, pp.156–168 ( 2007).  
DC, pp.156–168.