Modeling Multiple –Vehicle Property Damage Collisions in Urban Signalized Intersections

Seyed Saber Naseralavi[1], Masoud Ghasemi Noughabi[2], Esmaeel Ayati[3]

Received: 23.08.2012 *Accepted: 25.11.2012*

Abstract:
Development of disaggregate models for estimating different property damage collision type frequencies in urban intersections has rarely been studied, particularly in Iran. It seems very little research work being implemented for studying the effect factors on collision type frequency at intersections. The main objective of this paper is to develop suitable statistical models to predict types of property damage accident frequencies at signalized intersections approaches in the Mashhad of Iran based on geometric, traffic, and regulatory control characteristics. Three negative binomial models are estimated for collisions occurred in four-leg signalized intersections in the city of Mashhad and their results were compared together. These models are total, rear-end and right angel collision models. The goodness of fit was assessed by statistic tests. The Incidence Rate Ratio is used to assess the effectiveness of independent variables on frequency of property damage collision. Validation of models was controlled using paired samples T-test method. Modeling collision types showed a strong relationship between frequency of property damage collision types and independent variables such as road geometry, the type of control system and traffic characteristics. The results of this study revealed that seven of independent variables considerably affect the safety of signalized intersections.

Keywords: Property damage collision, signalized intersection, negative binomial model

Corresponding author E-mail: Saber_naseralavi@uk.ac.ir
1. Assistant Professor, Department of Civil Engineering, Shahid Bahonar University of Kerman, Kerman, Iran
2. MSc. Department of Civil Engineering, Shahid Bahonar University of Kerman, Kerman, Iran
3. Professor, Department of Civil Engineering , Ferdowsi University of Mashhad, Mashad, Iran

1. Introduction

An intersection is defined as the general area where two or more highways join or cross, including the roadway and roadside facilities for traffic movements within the area. Intersections are usually more hazardous than the road sections between them [Lord, 2000]. As the nodes of the roadway network, intersections need more attention for safety analyses than other roadway elements due to the fact that many intersections are found to be relatively collision-prone spots. In urban, the major proportion of total collisions occurred at intersections. This is partially caused by the complicated interactions between roadway users within the influence area of an intersection. The complex vehicle movements at intersections directly lead to the basic problem with these locations, to many conflicts. Usually, once a traffic conflict has not been avoided, a traffic collision will occur. Therefore, intersection safety should be addressed by traffic safety engineers [Pernia et al., 2002].

The usual practice to understand the interaction between geometric and traffic factors with collision causation is to establish a relationship between collision occurrence and intersection characteristics. Many researchers have focused on the development of aggregate collision prediction models, whereby the total expected number of collisions at intersections are predicted by geometric, environmental, and traffic variables [Bauer and Harwood, 2000; Greibe, 2003; Akin and Akbaş, 2010; Elvik, 2011]. Collision prediction models have rarely been developed focusing on predicting different collision types in urban intersection In Iran, very little research has been conducted studying the effective factors in collision type frequency at intersections.

There are two reasons for predicting disaggregate models that estimate and/or explain collision type frequencies as a function of geometric, environmental, and traffic factors. The first is that these models can help us to predict collision frequencies at signalized intersections by collision type and identify sites where these specific collision types occur. A second use of these models is to understand the differing effects of geometric, traffic, and environmental factors on different collision types. Thus, the effective variables on collision occurrence may have different coefficients for the different collision types and to consider a unique coefficient for all of the collision types may be unrealistic.

Because intersection approaches may have one of the collision types as the predominant type, separate models of rear-end, right angel collisions would provide valuable insights into different variables of intersection approach that influence the frequency of these specific types of collisions and countermeasure effectiveness. Therefore, the main objective of this paper is to develop suitable statistical models to predict frequencies of different types of property damage collisions at signalized intersections based on geometric, traffic, and regulatory control characteristics in the city of Mashhad, Iran.

2. Literature Review

Several models used to establish a relationship between collision occurrence and intersection characteristics include the multiple linear regression models, Poisson regression models and negative binomial (NB) regression models. The multiple linear regression models have several limitations to describe adequately the random, discrete, nonnegative and sporadic collision data [Chin and Quddus, 2003]. These include the presence of undesirable statistical properties, such as the possibility of negative collision counts and the lack of distributional properties, such as the condition of normally distributed collision occurrence. It is assumed in these models that collision data follow normal distribution; however, the collision data follow Poisson distribution [Anastasopoulos and Mannering, 2009].

Since collision occurrences are necessarily discrete, often sporadic and more likely random events, the Poisson regression models appear to be more suitable than the multiple linear regression models. A well-known limitation of the Poisson model is that the distribution restricts the mean and the variance to be equal, which seldom holds true with real-life collision data. When variance is greater than mean, the data are said to be over-dispersed. Over-dispersion occurs in practice because there are many factors affecting collision means and not all of them are accounted in the model. Data are said to be under-dispersed when variance is less than mean [Chin and Quddus, 2003]. In a number of recent studies, the collision data were found to be significantly over-dispersed, i.e. the variance is much greater than the mean [Naderan and Shahi, 2010; Vogt and Bared, 1998]. This will result in incorrect estimation of the likelihood of collision occurrence. In overcoming the problem of over-dispersion, several researchers have employed the negative binomial distribution instead of the Poisson [Abdel-Aty and Radwan, 2000; Poch and Mannering, 1996; Naderan and Shahi, 2010]. By relaxing the condition of mean equals to variance, negative binomial regression models are more suitable in describing discrete and nonnegative events.

Poch and Mannering developed negative binomial models predicting the frequency of total, rear-end, angel and approach-turn collisions using improvement of 63 four-leg intersections in Washington during 1987 and 1997. They concluded 16, 18 and 13 independent variables were involved in total, rear-end, angel and approach turn collisions, respectively. Increment left and right-turn, total approach traffic volume in thousands average daily traffic and number of phases per cycle increase collision frequencies in intersection. The existence of protected left-turn lane decrease collision frequencies in intersection [Poch and Mannering, 1996].

Bauer and Harwood reviewed 1306 urban intersections in the state of California during 1990 and 1992. They used the lognormal regression models to predict the frequency of total, fatal and injury collisions. 19 independent variables were considered in their modeling process resulting in 9 and 8 significant in the prediction of total, fatal and injury collisions, respectively. They also found that an increase in average daily traffic volume of the main road and crossroad and signal timing increase frequency of total collisions. The increasing number of lanes on major and cross road, average lane width on major and cross road, right-turn channelization and access control on major road decrease frequency of total collisions. Furthermore, they observed that an increase in the design speed on major road increases frequency of fatal and injury collisions [Bauer and Harwood, 2000].

Pernia et al. studied 447 signalized intersections in the state of Florida during the period 1990-1997. They applied the random effect negative binomial models to develop prediction models of all, angel, Left-turn and rear-end collisions. They observed that seven of the independent variables affect the safety of signalized intersections: average annual daily traffic, number of lanes on major road, presence of median on major road, surrounding land use (urban or rural), location type (business or other), posted speed on major road and shoulder treatment (paved or other) [Pernia et al., 2002].

Chin and Quddus studied 52 signalized intersections in the southwestern city of Singapore from 1992 to 1999. Applying random effect negative binomial prediction models, they concluded that eleven independent variables affect the safety of signalized intersections. The higher total approach and left-turn volume traffic in thousand, intersection sight distance, number of bus stops surrounding intersection, number of phases per cycle, the existence of median width greater than 2m, uncontrolled right-turn lane and surveillance camera increase total annual collision frequencies. The existence acceleration section on right-turn lane, increasing number of bus bays and Signal control type decreases total annual collision frequencies [Chin and Quddus, 2003].

Greibe studied 250 signalized intersections in Denmark during the period 1991-1998 and considered the influence of eight independent variables on the frequency of collisions. They applied negative binomial prediction models and found four significant variables: motor vehicle traffic flow in primary and secondary direction, number of lanes in primary and secondary direction [Greibe, 2003].

Wong et al. reviewed 262 signalized intersections in Hong Kong during 2002 and 2003. Negative binomial regression model was used to study the influence of sixteen independent variables on the frequency of slight injury collisions. They observed that increasing traffic volume (logarithm of annual average daily traffic), proportion of commercial vehicles, number of pedestrian streams, inverse of the average turning radius, kowloon area and presence of tram stops increase frequency of slight injury collisions. They also found that increasing average lane width decreases the frequency of slight injury collisions [Wong et al., 2007].

3. Methodology

3.1 Model Description

The Poisson regression and negative binomial regression are generally used to estimate collision prediction models [Lord and Mannering, 2010]. These models are suitable for modeling road collision counts that are discrete, nonnegative and sporadic. It is assumed that collisions occurring on a particular intersection are independent of one another. Collisions occurring at an intersection approach per unit time (e.g., year) generally follow the Poisson distribution. The mean number of collisions to be expected at an approach of intersection in a given time period, μ_i , is as follows:

$$E(y_i) = \mu_i = \exp(\beta_0 + \sum_{j=1}^{n} x_{ij}\beta_j) \quad (1)$$

Where $x_{i1}, x_{i2}, ..., x_{in}$ are the values of variables at the approach of intersection number i in a given time period and $\beta_0, \beta_1, ..., \beta_n$ are coefficients to be estimated by the modeling. In the Poisson distribution the variance of collisions at an approach of intersection i is equal to μ_i and the probability of an approach of intersection i having y_i collisions per year is given by:

$$P(y_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!} \quad (2)$$

The coefficients β are estimated by maximizing the log-likelihood function for the $L(\beta)$ Poisson distribution [Kim et al., 2006]:

$$L(\beta) = \sum_i (y_i \log \mu_i - \mu_i - \log y_i) \quad (3)$$

Here $L(\beta)$ is the vector of coefficients and μ_i is given by Eq. (1). The value β of that maximizes Eq. (3) is the estimated coefficient vector β.

A major limitation of the Poisson regression model is that the variance of the dependent variable (annual collision frequency), VAR (y_i), is constrained to be equal to its mean, E(y_i). When the mean and variance of the data are not approximately equal, the variance of the estimated Poisson model coefficients tend to be underestimated and the coefficients themselves are biased. This limitation can be readily overcome by using the negative binomial model [Kim et al., 2006].

The negative binomial regression model includes a quadratic term in the variance to reflect over-dispersion in the model variance [Kim et al., 2006]. The negative binomial regression model is represented by:

$$P(y_i) = \frac{(y_i + \alpha - 1)!}{y_i!(\alpha - 1)!} \frac{\mu_i^{y_i}}{(1 + \mu_i)^{y_i + \alpha}} \quad (4)$$

Where α over dispersion parameter and the variance is:

$$Var(y_i) = \mu_i + \alpha(\mu_i)^2 \quad (5)$$

As pointed out by Vogt and Bared the negative binomial allows for extra-Poisson variation due to other variables not included in the model [Vogt and Bared, 1998]. If the $\alpha = 0$ negative binomial reduces to the Poisson model. For the negative binomial distribution the estimated coefficients vector is obtained by maximizing $L(\beta, \alpha)$.

$$L(\beta, \alpha) = \sum_i \left[\left(\sum_{j=0}^{y_i} \log(1 + \alpha j) - \log(1 + \alpha y_i)\right)\right] + \sum_i \left[y_i \log \mu_i - (y_i + \frac{1}{\alpha}) \log(1 + \alpha\mu_i) - \log(y_i!) \right] \quad (6)$$

3.2 Model Evaluation

3.2.1 Over-Dispersion

A decision about whether the Poisson or Negative binomial model is appropriate can be based on deviance or Pearson chi-square statistic. The deviance of a model is defined as:

$$D^m = 2(L^f - L^m) \quad (7)$$

Where L^f is the log-likelihood function (Eq. (3)) that would be achieved if the model gave a perfect fit ($\mu_i = y_i$ for each i, $\alpha = 0$) and L^m is the log-likelihood (Eq. (3) or Eq. (6)) of the model under consideration. If the latter model is correct, D^m is approximately a chi-squared random variable with degrees of freedom equal to the number of observations (n) minus the number of parameters (p). Value of the deviance greatly in excess of n-p suggests that the model is over-dispersed due to missing variables and/or non-Poisson form. Thus when deviance divided by degrees of freedom is significantly larger than 1, over-dispersion is indicated [Vogt and Bared, 1998].

3.2.2 Goodness of Fit

Similar to the R^2 in linear regression, a measure based on the standardized residuals, Pearson's R^2, can be calculated for each generalized linear model to give some indication of the goodness-of-fit [Vogt and Bared, 1998].

$$R_p^2 = 1 - \frac{\sum_{i=1}^{n} \frac{\left(y_i - \hat{y}_i\right)^2}{\hat{y}_i}}{\sum_{i=1}^{n} \frac{\left(y_i - \bar{y}\right)^2}{\bar{y}}} = 1 - \frac{Pearson's\chi^2}{\frac{n}{\bar{y}} \times Var[y_i]} \quad (8)$$

Where, R_p^2 = Pearson's R-square statistic; y_i =Observed number of collision at ith approach of intersection during a time period; $\hat{y}_i$ = Estimated number of collisions during a time period; $\overline{y}$ = Average collision counts at all intersections of interest, n = number of observation.

3.3 Model Interpretation

The Incidence rate ratio (IRR), i.e. exp(β) were computed to facilitate interpretation of the variables included in the model. If IRR of a given variable is much less than 1.0, then an increase in value of the variable is associated with a significant improvement in safety. Conversely, if IRR is much greater than 1.0, an increase in the value of the variable is associated with a significant decline in safety. Otherwise, the variable has no effect on safety [Chin and Quddus, 2003].

3.4 Data

To establish a suitable statistical model that examines the relationship between property damage collisions frequency and the geometric, traffic and regulatory control characteristics of signalized intersections, a total of 50 four-legged signalized intersections in the city of Mashhad, Iran were used. The number of intersections may appear small but it covers quite a large area of city accounting for more than 45% of such intersections. Collision severity is classified into three categories based on the level of injury sustained in the collision: fatal, injury, and property-damage-only. This study focuses on property damage collisions only. The collision data were collected from the forms filled by police officer at collision scenes. Each four-leg urban signalized intersection was divided into four separate approaches and collision data were taken at each approach for 75 meter distance from the center of the intersection. In the police report in Iran, the collision occurred between 75 m to the center of the intersection is labeled ‘at intersection’ or ‘influenced by intersection’. In the event that the collision occurred in the center of the intersection, collisions were assigned to the approach of the faulty vehicle. A total of 1532 property damage collision data were gathered for 2007. To obtain a dependable model, it is necessary that the intersection and its approaches be considered by independent variables of real geometric, traffic and control situations. Most of geometry related variables were taken from 1:2000 scale Mashhad map, and traffic related variables were obtained from Mashhad transportation and traffic organization. The principle used to select explanatory variables was to include as many useful variables as possible based on the data available and engineering judgment. According to these criteria, for each approach, a total of 23 possible explanatory variables were considered. A sample summary statistics of explanatory variables is presented in Table 1.

To prepare for model development, it is appropriate to ask what variables correlate strongly with collision counts. Thus, an analysis of correlation coefficients between collision types and intersection variables for the signalized intersections was carried out using Probability values (P-value) to gain an insight of the effective variables. A small P-value indicates that a correlation is significant; a large one indicates that no particular significance can be attached to it. The positive and negative significant correlation means p-value is less than 0.1 and insignificant p-value is in excess of 0.1. The results showed the correlations between number of through traffic lanes in each approach, width of approach, protected exclusive left turn phase, existence of exclusive left and/or right turn lane(s) and its (their) number, width and length, median width in approach, right and/or Left turn, total and/or through traffic volume of approach (logarithm of average daily traffic) correlate positively with each of three collision types. Skew angel, one-way or two-way the approaches correlate negatively with each of three collision types. Exclusive right turn lane is insignificant in rear-end and right angel collisions. The distance of bus-stop in arrival direction to intersection is insignificant in rear-end collisions. The distance of bus-stop in exit direction to intersection and number of phases in each cycle are insignificant in right angel collisions. Existence of control camera is insignificant in total and right angel collisions. Type of intersection control system is insignificant in each of three collision types.

The correlation coefficients between independent variables are considered and found that all of correlation coefficients are under 0.5; therefore, correlation is small.

4. Property Damage Collisions Types Frequency Models

The objective of this study is to develop statistical models of the property damage collisions types frequency on individual intersection approaches (i.e. a four-legged intersection oriented north-south, east-west would have four approaches: northbound, southbound, eastbound and westbound). In this study has been tried to develop below models to predict property damage collision type frequency in signalized four-leg intersection: (1) total, (2) rear-end, (3) right angel collision frequency predicted models. The relative frequency (percent) of property damage collisions of right angel, rear-end, sideswipe, head-on, other cases and rear-side is 55.9%, 23.7%, 11.4%, 3.9%, 3.5% and 1.5%, respectively. Therefore, right angel and rear-end property damage collisions are major collision in signalized intersection. The modeling was done using statistical analysis software (SAS). In all cases the dependent variable (annual collision frequency) will be a non-negative integer. A summary statistics of dependent variables is presented on Table 2. As table 2 reveals the variance to mean ratio is greater than one that represents the collision data is over-dispersed. The collision data frequency distribution is presented in figure 1.

Figure 1 highlights that the shapes of collision frequencies follow negative binomial distribution. Negative binomial regression was used because of the collision data were over-dispersed. For negative binomial regression, the regression parameters were estimated by maximum likelihood method with GENMOD procedure in SAS.

The SAS procedure GENMOD software fits a generalized linear model to the data by maximum likelihood estimation of the parameter vector β .There is, in general, no closed form solution for the maximum likelihood estimates of the parameters. The GENMOD procedure estimates the parameters of the model numerically through an iterative fitting process.

The dispersion parameter α is also estimated by maximum likelihood or, optionally, by the residual deviance or by Pearson's chi-square divided by the degrees of freedom. Covariances, standard errors, and p-values are computed for the estimated parameters based on the asymptotic normality of maximum likelihood estimators. Using GENMOD procedure and a backward elimination modeling approach, models were developed for collision types. In each case, an initial "full" model was developed that included all variables. Initial problems, such as multicollinearity, were addressed and affected variables were removed as appropriate. The resulting "full" model most completely explains the effects of the variables on intersection safety. Though not all variables are statistically significant in the initial model, many displayed practical significance and would likely become statistically significant if the sample size were increased. In the next step, variables were removed from the initial model based upon p-values. After removing the variable in the model with the highest p-value, the coefficients and p-values of the remaining variables were examined for changes due to multicollinearity. Models were reduced until all variables had p-values of 0.10 or less to arrive at the final "reduced" model. In each step of modeling, the correlation matrix was studied and if two variables were correlated strongly with each other, one variable was excluded from the model on the condition that the model fit did not suffer significantly.

The Multicollinearity describes the strength of an association between variables. An association between variables means that the value of one variable can be predicted, to some extent, by the value of the other. Variance inflation factor (VIF) is common way for detecting multicollinearity. VIF is defined by equation (9):

$$VIF_k = \frac{1}{1 - R_k^2} \tag{9}$$

Where R_k^2 is the R^2-value obtained by regressing the k^{th} predictor on the remaining predictors. Note that VIF exists for each of the k predictors in a multiple regression model. We can decide to throw out which variable by examining the size of VIF. A general rule is that the VIF should not exceed 5 [Belsley et al., 1980].

4.1 Total Property Damage Collisions Frequency Predicted Model

Total property damage collisions occurred in intersection include head-on, rear-end, right angel, rear-side and sideswipe collisions Negative binomial estimation results of total annual property damage collisions frequency at intersection approaches are presented in table 3. This table includes the explanatory variables, degree

Table 1. Summary statistics of the explanatory variables in the study

Symbol	Explanatory Variable	Min	Max	Mean	Standard deviation (S.D.)
X_{NL}	Number of through traffic lanes in each approach	2	9	5.89	1.3
X_{W}	Width of approach (in meter)	7.64	34.5	20.87	4.62
X_{LT}	Exclusive left turn lane (1 if exclusive left turn lane, 0 otherwise)	0	1	0.46	0.50
X_{PLT}	Protected exclusive left turn phase (1 if protected left turn phase, 0 otherwise)	0	1	0.03	0.17
X_{NLT}	Number of exclusive left turn lanes	0	1	0.46	0.50
X_{WLT}	Width of exclusive left turn lane (in meter)	0	4	1.16	1.35
X_{LLT}	Length of exclusive left turn lane (in meter)	0	72.13	12.9	15.67
X_{RT}	Exclusive right turn lane (1 if exclusive right turn lane, 0 otherwise)	0	1	0.46	0.50
X_{NRT}	Number of exclusive right turn lanes	0	3	0.87	0.96
X_{WRT}	Width of exclusive right turn lane (in meter)	0	12	3.12	3.43
X_{LRT}	Length of exclusive right turn lane (in meter)	0	60	9.81	12.58
X_{MW}	Median width in approach (in meter)	0	8.88	2.41	2.22
X_{DBSA}	The distance of bus-stop in arrival direction to intersection (1 if distance bus stop greater than 50 m, 0 otherwise)	0	1	0.81	0.39
X_{DBSE}	The distance of bus-stop in exit direction to intersection (1 if distance bus stop greater than 50 m, 0 otherwise)	0	1	0.76	0.42
X_{DM}	One-way or two-way the approaches (1 if one-way approach, 0 otherwise)	0	1	0.1	0.31
X_{SA}	skew angel (The angle between major and minor approaches, in degrees)	0	126.26	44.64	45.8
X_{NPH}	Number of phases in each cycle	2	4	2.38	0.56
X_{SC}	Existence of control camera (1 if a control camera exist in an intersection, 0 otherwise)	0	1	0.14	0.35
X_{TCS}	Type of intersection control system (1 if adaptive control system, 0 otherwise)	0	1	0.44	0.50
X_{THTV}	Through traffic volume (in logarithm)	3.94	4.88	4.427	0.179
X_{RTV}	Right turn traffic volume (in logarithm)	0	4.11	3.072	0.995
X_{LTV}	Left turn traffic volume(in logarithm)	0	4.12	3.111	1.024
X_{TTV}	Total traffic volume of approach (in logarithm)	3.94	4.93	4.551	0.168

Table 2. Summary statistics of the dependent variables in the study

Property damage collision type	Symbol	Min	Max	Mean	Standard deviation (S.D.)	Variance	Variance to mean ratio
Total collision	y_{TDA}	0	38	7.66	6.12	37.45	4.89
Rear-End collision	y_{RDA}	0	13	1.84	2.42	5.86	3.18
Right angel collisions	y_{SDA}	0	30	4.21	4.09	16.73	3.97

of freedom, estimated coefficients, p-value, incidence rate ratio (IRR), variance inflation factor (VIF) and dispersion parameter. Goodness of fit tests are also shown in table 4. In table 3, the variables with a positive sign increasing collision frequency and a negative sign decreasing collision frequency. The variables included in this model (and rear-end and right angel property damage collisions predicted models) are those that resulted in the lowest p-value (after a systematic evaluation of all variables) and were selected from possible exploratory variables available. IRR was also calculated to facilitate interpretation of the variables included in this model. As shown in table 3, the existence protected exclusive left turn phase, the increasing the median width,

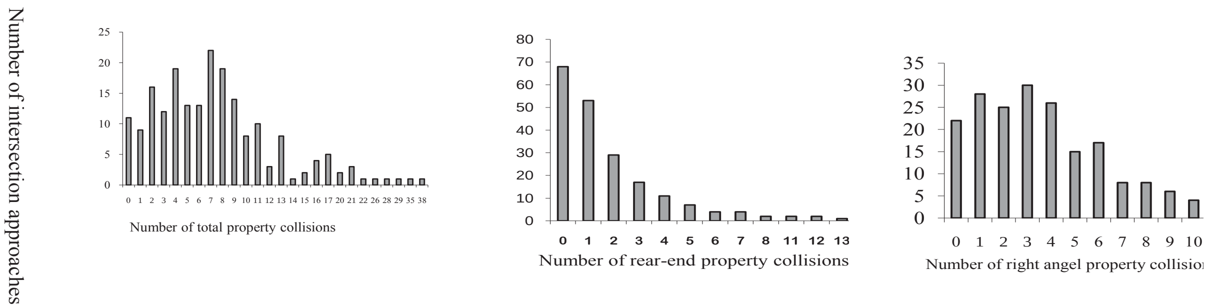

Figure 1. Total, rear-end and right angel property damage collisions frequency distribution at four-leg urban signalized intersection approaches

number of phases per cycle, through, right and left turn traffic volume of approach increase total property damage collisions frequency and the increasing skew angel decreases total property damage collisions frequency.

Total property damage collisions frequency predicted model as follows:

$$Ln(y_{TDA}) = -3.9034 + 0.3207x_{PLT} + 0.1166x_{MW} - 0.003x_{SA} + 0.1882x_{NPH} + 0.922x_{THTV} + 0.1126x_{RTV} + 0.2222x_{LTV} \quad (10)$$

The parameters of above model are described in table 3. The coefficient variable of protected exclusive left-turn phase shows the existence of protected exclusive left turn phase tend to increase rear-end collisions. This can be explained by considering the fact that in above mentioned intersections approaches which have a protected exclusive left turn phase, left-turn traffic volume is high; however, there is only one exclusive left turn lane. In such intersections, the drivers that aim to turn left use straight lanes because of the number of exclusive left turn lanes are not proportional to left-turn traffic volume. When a vehicle stops to complete left turn maneuver in straight lanes, it conflicts to vehicles that intend to move straight direction. Therefore, the probability of rear-end collisions during the phase change periods increases.

The increasing median width of approaches intersection increases the total property damage collisions frequency. Wider median widths usually come from larger intersections and they allow greater degrees of spatial freedom for left-turning vehicles. Near the stop line, wider median widths may also create more conflicts as

Table 3. Negative binomial estimation results for total, rear-end and right angel annual property damage collisions frequency

Variable	Variable symbol	Total annual property damage collisions frequency				Rear-end annual property damage collisions frequency				Right angel annual property damage collisions frequency			
		Coefficient estimate	P-value	Incidence rate ratio	Variance inflation factor	Coefficient estimate	P-value	Incidence rate ratio	Variance inflation factor	Coefficient estimate	P-value	Incidence rate ratio	Variance inflation factor
Intercept	β_0	-3.9034	0.0012	-[a]	-[b]	-11.463	<0.0001	-	-	-	-	-	-
Protected exclusive left turn phase (1 if protected left turn phase, 0 otherwise)	X_{PLT}	0.3207	0.10	1.3781	1.11	0.6302	0.0424	1.878	1.109	-	-	-	-
Median width in approach (in meter)	x_{MW}	0.1166	<0.0001	1.1237	1.185	0.1301	0.0002	1.138	1.106	0.1764	<0.0001	1.1929	1.131
Skew angel (in degrees)	x_{SA}	-0.003	0.0037	0.997	1.329	-[c]	-	-	-	-0.0038	0.0012	0.9962	1.073
Number of phases in each cycle	x_{NPH}	0.1882	0.0092	1.2071	1.124	0.3463	0.006	1.414	1.089	-	-	-	-
Through traffic volume (in logarithm)	x_{THTV}	0.922	0.0006	2.5143	1.3464	2.1612	<0.0001	8.68	1.091	-	-	-	-
Right turn traffic volume (in logarithm)	X_{RTV}	0.1126	0.0427	1.1192	1.542	-	-	-	-	-	-	-	-
Left turn traffic volume (in logarithm)	X_{LTV}	0.2222	<0.0001	1.2488	1.502	0.3082	0.0037	1.361	1.091	0.3269	<0.0001	1.3867	1.070
Negative binomial dispersion parameter, α ,	α	0.1454	-[d]	-	-	0.3388	-	-	-	0.3048	-	-	-

a. (-) dash in the column of incidence rate ratio means the value is not defined for intercept.
b. (-) dash in the column of variance inflation factor means the value is not defined for intercept.
c. (-) dash in the column of coefficient estimate means the variable is not included the model.
d. (-) dash in the column of P-value means the value is not defined for negative binomial dispersion parameter.

the number of conflict points is higher and movements of through vehicles are less channelized.

In this study, the skew angle for an intersection was defined as the angel between major and minor approaches. The coefficient of skew angel is tiny that shown obtuse approach angle reduces the property damage collisions occurrence very small amount. This geometry facilitates an easy maneuvering for vehicles turning right from major to minor street as well as for vehicles turning left from minor road to major road.

The coefficient variable of number phases per cycle shows having a higher number of phases per cycle may increase the number of collisions. This is not surprising since most collisions occur during the phase change periods. The high volume and high congestion intersections usually have the greater number of phases per cycle. When number of phases increased, drivers might get more nervous due to driver frustration and might try to complete the maneuver quickly, which may lead to severe injury and fatal collisions.

The high through traffic volume on the approach increases collision likelihoods. This may be due to the increment in the exposure to conflicts. As traffic volume increases, there are fewer available gaps for the left-turning opposing maneuver as well as right-turning merging maneuver. As a result of fewer turning opportunities, drivers may be more willing to take risks when making the turn.

The Increasing right turn traffic volume increases the likelihood of collisions occur. To maneuver around to the right, it is necessary that vehicles reduce their speed. The speed difference with the vehicle moving directly cause the collision in intersection where the exclusive right turn lane wasn't provided because lack of sufficient space in urban areas.

The coefficient variable of left turn traffic volume shows increasing left turn traffic volume increases the total property damage collisions frequency. The left turn movement has always propounded one of concerned problems in intersections. The increasing left turn traffic volume increases conflict points between vehicles moving left turn and through. The left turn traffic volume is one of the effective factors in signal timing, taking up valuable cycle time. If aren't provided the exclusive left turn lane in approaches intersection, occur collision and decrease level of service; because, in most cases, it requires crossing the path of opposing traffic.

In table 4, the deviance to degree of ratio is 1.1325 that represent the collision data is over-dispersed. Table 3 also shows the negative binomial dispersion parameter, α, is 0.1454 that the use of the negative binomial model is justified by the highly significant value of . The Pearson R-square value is equal to 0.80 representing that the model has a satisfactory ability in explaining the variation of the data.

4.2 Rear-End Property Damage Collisions Frequency Predicted Model

Negative binomial estimation results of rear-end annual property damage collisions frequency at intersection approaches are presented in table 3. Goodness of

Table 4. Goodness of fit test statistics of negative binomial model for total, rear-end and right angel annual property damage collisions frequency

Item	Value		
	Total collisions	Rear-end collisions	Right angel collisions
Number of observations، (n)	200	200	200
Number of variables included the model، (p) ،	8	6	3
Degree of freedom، $(n - p)$ ،	192	194	197
Log likelihood at convergence, $L(B)$	1853.3131	-24.2644	523.6928
Deviance	217.4436	210.8196	213.4101
Deviance/Degree of freedom	1.1325	1.0867	1.0833
Pearson chi-square	192.5035	197.1595	197.7024
Pearson chi-square/Degree of freedom	1.0026	1.0163	1.0036
Pearson's R-square	0.80	0.69	0.751

fit tests are also shown in table 4. As shown in table 3, the increasing the protected exclusive left-turn phase, median width in approach, through and left turn traffic volume of approach increase rear-end property damage collisions frequency. Rear-end property damage collisions frequency predicted model as follows:

$$Ln(y_{RDA}) = -11.463 + 0.6302x_{PLT} + 0.1301x_{MW} + 0.3463x_{NPH} + 2.1612x_{THTV} + 0.3082x_{LTV} \quad (11)$$

In table 4, the deviance to degree of freedom ratio is 1.0867 that represent the collision data is over-dispersed. Table 4 also shows the Negative binomial dispersion parameter, α , is 0.3388 that the use of the negative binomial model is justified by the highly significant value of α. The Pearson R-square value is equal to 0.69 representing that the predictive ability of rear-end collision frequency by variables included in this model is 0.69.

4.3 Right Angel Property Damage Collisions Frequency Predicted Model

Negative binomial estimation results of right angel annual property damage collisions frequency at intersection approaches are presented in table 3. Goodness of fit tests is also shown in table 4. As shown in table 4, the variable of skew angel decrease right angel collisions frequency. The increasing median widths and Left turn traffic volume increase right angel property damage collisions frequency. Right angel property damage collisions frequency predicted model as follows:

$$Ln(y_{SDA}) = 0.1764x_{MW} - 0.0038x_{SA} + 0.3269x_{LTV} \quad (12)$$

The variables included in this model follow the aforementioned explanation for total and rear-end collision models.

In table 4, the deviance to degree of freedom is 1.0833 that represent the collision data is over-dispersed. Table 3 also shows the Negative binomial dispersion parameter, α , is 0.3048 that the use of the negative binomial model is justified by the highly significant value of α. The Pearson R-square value is equal to 0.751 representing that the model has a satisfactory ability in explaining the variation of the data.

An important assumption of analysis is that all of the predictor variables are statistically independent. Multicollinearity refers to the violation of this assumption, and describes a situation in which the possible correlations between predictor variables are significant. To eliminate this problem, correlation analysis is conducted and found correlation between each pair of variables is insignificant, because the coefficient correlation is less than 0.5. The Variance inflation factors are computed to ensure the correlation between an independent variable with other independent variables. As shown the table 3, all of the variance inflation factors are under 5; therefore, multicollinearity is low.

4.4 Validation of Models

Validation of models is one of the major steps to develop models. The paired samples T-test are used to verify if the differences are systematic or caused by mere chance [Montgomery, 2004]. To use this method, several intersections which are not involved in development modeling are selected and were predicted the collisions frequency in their approaches by developed model. The paired sample T-test compares the mean of observed collisions frequency with predicted collisions frequency that was computed based on developed models. The paired samples T-test procedure in this study was done using SAS software. Table 5 shows the descriptive statistics for both observed and predicted collisions frequency. The most relevant statistics for our purposes are the two means. Remember, this test is based on the difference between the two variables. As shown table 5, the significant value is greater than 0.05 for property damage collisions types frequency models. If the significance value is greater than 0.05, cannot reject the null hypothesis of no difference between the mean of observed and predicted collisions frequency of models.

5. Conclusions

As stated previously, one of the justifications for modeling collision type is to identify which variables contribute to certain types of collisions and to compare how different significant variables affect safety for different collision types. The results show that a handful of the available roadway geometric, traffic volume and regulatory control variables affect the safety of four-leg urban signalized intersections. The mentioned variables are as follow: through traffic volume, right turn traffic volume, left turn traffic volume, median width in approach, skew angel, number of phases in each cycle,

Table 5. Descriptive statistics for both observed and predicted collisions frequency

Model	Predicted and observed collisions	Number of observation	Mean	Standard deviation	Standard error mean	T- statistic	Degree of freedom	p-value
Total property damage collision	Predicted	36	6.28	4.779	0.796	-0.757	35	0.254
	observed	36	5.94	5.534	0.922			
Rear-end property damage collision	Predicted	36	1.58	1.811	0.302	0.552	35	0.585
	observed	36	1.42	3.307	0.551			
Right angel property damage collision	Predicted	36	3.31	2.109	0.351	-1.276	35	0.210
	observed	36	2.75	3.027	0.505			

protected exclusive left turn phase.

Among traffic variables, traffic volume in general increases collisions frequency in each three models. As traffic volume increases, exposure to risk (at the site) increases. The traffic volume is generally not viewed as a controllable factor but instead an important predictor of collisions, since controlling total traffic volume is generally not an option to engineers or planners. As the Incidence rate ratio (IRR) in table 3 shows increasing through traffic volume at one unit increases total and rear-end collisions frequency per year about 2.5 and 8.68 times, respectively. The increment of right turn traffic volume increases rear-end annual property damage collisions frequency by 11.9%. The increasing left turn traffic volume increase total, rear-end and right angel collisions frequency by 24.9%, 36.1% and 38.7%, respectively.

The influential geometric variables which affect the type of collisions occurrences are: median width in approach, skew angel, number of phases in each cycle, protected exclusive left turn phase. The effect of these variables on collisions frequency in urban signalized intersection is summarized below:

• The existence of protected exclusive left turn phase increases total and rear-end collisions frequency at 37.8%, and 87.8%, respectively.

• The increment of median width in approach increases total, rear-end and right angel collisions frequency at 12.4%, 13.8% and 19.3%, respectively.

• The increment of skew angel decreases total and right angel collisions frequency at 0.3% and 0.4%, respectively.

• The increment of number of phases in each cycle increases total and rear-end collisions frequency at 20.7%, and 41.4%, respectively.

The modeling of collision type frequencies clearly demonstrates, at least statistically, that collision types are correlated with set of predictors to different coefficients. The estimation of collision type models may lead to insights as to the relative effectiveness of various countermeasures and/or predictive variables.

6. Recommendation

That is not possible to predict models of other collision types because the data collection was complete. If the data collection of the type of collisions occurrence in intersection is completely available over a period of several years, it will be possible to estimate sideswipe, rear-side and other property damage collision models in intersection. In addition, other variables entering the modeling which indicated the road pavement and weather conditions may cause the predictability of models will improve but were unavailable for this study. Finally, it is recommended that collision type models be estimated more routinely in conjunction and as complements to total collision models to identify different effective factors on property damage collision types and select feasible countermeasure effectiveness.

7. References

- Abdel-Aty, M. and Radwan E. (2000) Modeling traffic accident occurrence and involvement, Journal of Accident Analysis and Prevention, Vol. 32, No. 5, pp. 633-642.

- Akin, D. and Akbaş, B. (2010) “A neural network (NN) model to predict intersection crashes based upon driver, vehicle and roadway surface characteristics”, Journal of Scientific Research and Essays, Vol. 5, No. 19, pp. 2837-2847.

- Anastasopoulos, P. and Mannering, F.L. (2009) "A note on modeling vehicle accident frequencies with random-parameters count models", Journal of Accident Analysis and Prevention, Vol. 41, No. 1, pp. 153-159.

- Bauer, K. and Harwood, D. (2000) "Statistical models of at-grade intersection accidents-addendum, Federal Highway Administration, Washington, D.C., Report FHWA-RD-99-094.

- Belsley, D.A., Kuh, E. and Welsch, R.E. (1980) "Regression diagnostics", Wiley, New York.

- Chin, H.C. and Quddus, M.A. (2003) "Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections", Journal of Accident Analysis and Prevention, Vol. 35, No. 2, pp. 153–159.

- Elvik, R. (2011) "Assessing causality in multivariate accident models", Journal of Accident Analysis and Prevention, Vol. 43, No. 1, pp. 253-264.

- Greibe, P. (2003) "Accident prediction models for urban roads", Journal of Accident Analysis and Prevention, Vol. 35, No. 2, pp. 173–185.

- Kim, D.G., Washington, S. and Oh, J. (2006) "Modeling crash types: new insights into the effects of covariates on crashes in intersection", Journal of Transportation Engineering, Vol. 132, No. 4, pp. 282–292.

- Lord, D. and Mannering, F. L. (2010) "The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives", Journal of Transportation Research Part A: Policy and Practice, Vol. 44, No. 5, pp. 291-305.

- Montgomery, D.C. (2004) "Design and analysis of experiments", 6nd ed, Wiley, New York.

- Naderan, A. and Shahi, J. (2010) "Aggregate crash prediction models: Introducing crash generation concept", Journal of Accident Analysis and Prevention, Vol. 42, No. 1, pp. 339–346.

- Pernia, J., Lu, J. J., Xie, X., Weng, M. and Snider, D. (2002) "Development of models to quantify the impacts of signalization on intersection crashes", Presented at the 81th Annual Meeting of the Transportation Research Board, Washington, D.C., pp. 1-22.

- Poch, M. and Mannering, F.L. (1996) "Negative binomial analysis of intersection accident frequencies", Journal of Transportation Engineering, Vol. 122, No.2, pp. 105–113.

- Vogt, A. and Bared J. (1998) "Accident models for two-lane rural segments and intersections", Federal Highway Administration, Washington, D.C., Report FHWA-RD-98-133.

- Wong, S. C., Sze, N. N. and Li, Y.C. (2007) "Contributory factors to traffic crashes at signalized intersections in Hong Kong", Journal of Accident Analysis and Prevention, Vol. 39, No. 6, pp. 1107-1113.