Modeling Multiple –Vehicle Property Damage Collisions in Urban Signalized Intersections

Seyed Saber Naseralavi¹, Masoud Ghasemi Noughabi², Esmacel Ayati³

Received: 2012.08.23   Accepted: 2012.11.25

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 that very little research work has been implemented for studying the factors affecting the 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 City of Mashhad, 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. 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. Results 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- M.Sc., Department of Civil Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
3- Professor, Department of Civil Engineering, Ferdowsi University of Mashhad, Mashhad, 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 compared to other roadway elements due to the fact that many intersections are found to be relatively collision-prone spots. In urban areas, 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. When reaching an intersection, roadway users are confronted with a lot of choices to make, such as either to stop or keep going, or go either left, right or straight. The complex movements of vehicle at intersections make 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 intersections in Iran, very little research has been conducted to study 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 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 angle 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 Model. The multiple linear regression models have several limitations to describe adequately the random, discrete non-negative 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 oc-
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 non-negative events. Poch and Mannering developed negative binomial models predicting the frequency of total, rear-end, angle and approach-turn collisions using improvement of 63 four-leg intersections in Washington during 1987 and 1997. They concluded that 16, 18 and 13 independent variables were involved in total, rear-end, angle 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 and colleagues 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, angle, 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 and colleagues 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 the 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, non-negative 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
3.1. Model Description

Here, distribution. When the mean and variance of the data are not approximately equal, the variance of the Poisson distribution is constrained to be equal to its mean, $E(y_i) = \mu_i$. The estimated coefficient vector $\beta$ is equal to

$$P(y_i) = \frac{y_i^{\mu_i}}{y_i!}$$

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

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

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

A major limitation of the Poisson regression model is that the variance of the dependent variable (annual collision frequency), $VAR(y)$, is constrained to be equal to its mean, $E(y)$. 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:

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

where $\alpha$ over dispersion parameter and the variance is:

$$VAR(y_i) = \mu_i + \alpha (\mu_i)^2$$

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 $\alpha = 0$, the negative binomial reduces to the Poisson model. For the negative binomial distribution the estimated coefficients vector is obtained by maximizing $L(\beta, \alpha)$.

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

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:

$$D^* = 2(L' - L^*)$$

where $L'$ is the log-likelihood function [Eq. (3)] that would be achieved if the model gave a perfect fit ($\mu_i = y_i$ for each $i$ and $\alpha = 0$) and $L^*$ is the log-likelihood (Eq. (3) or Eq. (6)) of the model under consideration. If the latter model is correct, $D^*$ 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 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].

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

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

3.3. Model Interpretation

The Incidence Rate Ratio (IRR), i.e. $\exp(\beta)$ was com-
Seyed Saber Naseralavi, Masoud Ghasemi Noughabi, Esmaeel Ayati

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 angle, 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 angle 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 angle collisions. Existence of control camera is insignificant in total and right angle 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 it has been tried to develop the following models to predict property damage collision type frequency in signalized four-leg intersection: (1) total, (2) rear-end, (3) right angle collision frequency predicted models. The rela-
Modeling Multiple–Vehicle Property Damage Collisions in Urban ...

The SAS procedure GENMOD software fits a generalized linear model to the data by maximum likelihood estimation of the parameter vector $\beta$. 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 $\alpha$ is also estimated by maximum likelihood or, optionally, by the residual deviance or by Pearson’s chi-square divided by the degrees of freedom. Co-variances, 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 multi-collinearity, were addressed and affected variables were removed as appropriate. The resulting

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

“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 multi-collinearity. 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 multi-collinearity 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 multi-collinearity. VIF is defined by equation (9):

\[ VIF_k = \frac{1}{1 - R^2_k} \]  

where: \( R^2_k \) is the R2-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 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 existent protected exclusive left turn phase, the increasing the median width, number of phases per cycle, through right and left turn traffic volume of approach increase total property damage collisions frequency and the increasing skew angle decreases total property damage collisions frequency.

Total property damage collisions frequency predicted the model as follows:

\[ \ln(y_{p}) = -3.9034 + 0.3207 x_{TV} + 0.1166 x_{TVL} - 0.003 x_{TVR} + 0.1882 x_{TH} + 0.922 x_{THR} + 0.1126 x_{THL} + 0.2222 x_{THR} \]  

---

Seyed Saber Naseralavi, Masoud Ghasemi Noughabi, Esmaeel Ayati
Modeling Multiple –Vehicle Property Damage Collisions in Urban ...

The parameters of the above mentioned 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 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 and the left turning vehicle occupy more than one lane during red period which block the movement of through vehicles and result in more rear-end collisions frequency. 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 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 angle between major and minor approaches. The coefficient of skew angle is tiny, that shows obtuse approach angle reduces the property damage collisions occurrence to very small amount. This geometry facilitates an easy maneuvering for vehicles turning right from major to minor streets as well as for vehicles turning left from minor road to major road.

The coefficient variable of number phases per cycle shows that more 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 occurrence. 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 of the lack of sufficient space in urban areas.

The coefficient variable of left turn traffic volume shows that by increasing left turn traffic volume, the total property damage collisions frequency increases. 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 the exclusive left turns lane in approaches are not provided, the probability of collisions increase and leads to lower level of service; because, in most cases, it requires crossing the path of the opposing traffic.

In table 4, the deviance to degree of ratio is 1.1325 that represents the collision data is over-dispersed. Table 3 also shows the negative binomial dispersion parameter, $\alpha$, is 0.1454 that the use of the negative binomial model is justified by the highly significant value of $\alpha$.

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 fit tests are also shown in table 4. As shown in table 3, the existence of protected exclusive left-turn phase, increasing median width of approach, and increasing through and left turn traffic volume of approach, result in increasing of rear-end prop-
Table 3. Negative binomial estimation results for total, rear-end and right angle annual property damage collisions frequency

<table>
<thead>
<tr>
<th>Variable</th>
<th>Symbol</th>
<th>Variable</th>
<th>Symbol</th>
<th>Variable</th>
<th>Symbol</th>
<th>Variable</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.3418</td>
<td>Intercept</td>
<td>0.3388</td>
<td>Intercept</td>
<td>1.4207</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protected exclusive left turn phase</td>
<td>X_{PLT}</td>
<td>0.3207</td>
<td>X_{PLT}</td>
<td>0.3207</td>
<td>X_{PLT}</td>
<td>0.3207</td>
<td></td>
</tr>
<tr>
<td>Median width in approach (in meter)</td>
<td>MW</td>
<td>0.1166</td>
<td>MW</td>
<td>0.1166</td>
<td>MW</td>
<td>0.1166</td>
<td></td>
</tr>
<tr>
<td>Skew angle (in degrees)</td>
<td>SA</td>
<td>-0.0030</td>
<td>SA</td>
<td>-0.0030</td>
<td>SA</td>
<td>-0.0030</td>
<td></td>
</tr>
<tr>
<td>Number of phases in each cycle</td>
<td>NPH</td>
<td>0.1882</td>
<td>NPH</td>
<td>0.1882</td>
<td>NPH</td>
<td>0.1882</td>
<td></td>
</tr>
<tr>
<td>Through traffic volume (in log)</td>
<td>THTV</td>
<td>0.9220</td>
<td>THTV</td>
<td>0.9220</td>
<td>THTV</td>
<td>0.9220</td>
<td></td>
</tr>
<tr>
<td>Right turn traffic volume (in log)</td>
<td>X_{RTV}</td>
<td>0.1126</td>
<td>X_{RTV}</td>
<td>0.1126</td>
<td>X_{RTV}</td>
<td>0.1126</td>
<td></td>
</tr>
<tr>
<td>Left turn traffic volume (in log)</td>
<td>X_{LTV}</td>
<td>0.2222</td>
<td>X_{LTV}</td>
<td>0.2222</td>
<td>X_{LTV}</td>
<td>0.2222</td>
<td></td>
</tr>
<tr>
<td>Negative binomial dispersion parameter, $\alpha$</td>
<td>$\alpha$</td>
<td>0.1454</td>
<td>$\alpha$</td>
<td>0.1454</td>
<td>$\alpha$</td>
<td>0.1454</td>
<td></td>
</tr>
</tbody>
</table>

(-) dash in the column of incidence rate ratio means the value is not defined for intercept.

(-) dash in the column of variance inflation factor means the value is not defined for intercept.

(-) dash in the column of coefficient estimate means the variable is not included in the model.

(-) dash in the column of P-value means the value is not defined for negative binomial dispersion parameter.
Modeling Multiple -Vehicle Property Damage Collisions in Urban ...

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

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

Property damage collisions frequency. Rear-end property damage collisions frequency predicted model as follows:

\[
\ln(y_{\text{re}}) = -11.463 + 0.6302 x_{\text{svr}} + 0.1301 x_{\text{svr}^2} + 0.3463 x_{\text{svr}^3} + 2.1612 x_{\text{svr}^4} + 0.3082 x_{\text{svr}^5} \quad (11)
\]

In table 4, the deviance to degree of freedom ratio is 1.0867 that represents the collision data is over-dispersed. Table 4 also shows the Negative binomial dispersion parameter, \( \alpha \), is 0.3388 that the use of the negative binomial model is justified by the highly significant value of \( \alpha \). 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 are also shown in table 4. As shown in table 4, the variable of skew angle decrease right angel collisions frequency. The increasing median width and left turn traffic volume increase the right angle property damage collisions frequency. Right angel property damage collisions frequency predicted the model as follows:

\[
\ln(y_{\text{ra}}) = 0.1764 x_{\text{svr}} - 0.0038 x_{\text{svr}^2} + 0.3269 x_{\text{svr}^3} \quad (12)
\]

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, \( \alpha \), is 0.3048 that the use of the negative binomial model is justified by the highly significant value of \( \alpha \). 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 by 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 validate the models using the paired samples T-test, several intersections which are not involved in developing the models, are selected and the frequency of collisions in their approaches were predicted using 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 in table 5, the significant value is greater than 0.05 for all of property damage collisions frequency models. If the significance value is greater than 0.05, then the null hypothesis is not rejected and we conclude that there is no difference between the mean of observed and predicted collisions frequency of models.

5. Conclusions
As stated previously, one of the justifications for modelling 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 follows: Through traffic volume, right turn traffic volume, left turn traffic volume, median width in approach, skew angle, number of phases in each cycle, and 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) is increased. 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 about 2.5 per year and 8.68 times, respectively. The increment of right turn traffic volume increases the rear-end annual property damage collisions frequency by %11.9. The increasing left turn traffic volume increases the total, rear-end and right angle 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 angle, 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 angle collisions frequency at %12.4, %13.8 and %19.3, respectively.
- The increment of skew angle decreases the total and right angle collisions frequency at %0.3 and %0.4, respectively.
- The increment of the number of phases in each cycle increases the 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. Recommendations
It 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

<table>
<thead>
<tr>
<th>Model</th>
<th>Number of observed collisions</th>
<th>Number of predicted collisions</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Standard Error Mean</th>
<th>T-statistic</th>
<th>Degree of Freedom</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total property damage collision</td>
<td>Predicted 36</td>
<td>6.28</td>
<td>4.779</td>
<td>0.796</td>
<td>-0.757</td>
<td>35</td>
<td>0.254</td>
<td></td>
</tr>
<tr>
<td></td>
<td>observed 36</td>
<td>5.94</td>
<td>5.534</td>
<td>0.922</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rear-end property damage collision</td>
<td>Predicted 36</td>
<td>1.58</td>
<td>1.811</td>
<td>0.302</td>
<td>0.552</td>
<td>35</td>
<td>0.585</td>
<td></td>
</tr>
<tr>
<td></td>
<td>observed 36</td>
<td>1.42</td>
<td>3.307</td>
<td>0.551</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right angle property damage collision</td>
<td>Predicted 36</td>
<td>3.31</td>
<td>2.109</td>
<td>0.351</td>
<td>-1.276</td>
<td>35</td>
<td>0.210</td>
<td></td>
</tr>
<tr>
<td></td>
<td>observed 36</td>
<td>2.75</td>
<td>3.027</td>
<td>0.505</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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. It should be mentioned that, including other variables such as road pavement and weather conditions may improve the precision of the models, but were not used in 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