Examining Methods for Estimating Crash Counts
According to Their Collision Type

Srinivas Reddy Geedipally¹
Engineering Research Associate
Texas Transportation Institute
Texas A&M University
3136 TAMU
College Station, TX 77843-3136
Tel. (979) 845-9892
Fax. (979) 845-6481
Email: srinivas8@tamu.edu

Sunil Patil
PhD Candidate
Texas A&M University
3136 TAMU
College Station, TX 77843-3136
Tel. (979) 845-9892
Fax. (979) 845-6481
E-mail: sunilpatil@tamu.edu

Dominique Lord
Assistant Professor
Zachry Department of Civil Engineering
Texas A&M University
3136 TAMU
College Station, TX 77843-3136
Tel. (979) 458-3949
Fax. (979) 845-6481
Email: d-lord@tamu.edu

Word Count: 4,977 + 2,500 (4 tables + 6 figures) = 7,477 words

March 12, 2010

¹Corresponding author
ABSTRACT

Multinomial logit (MNL) models have been applied extensively in the fields of transportation engineering, marketing and recreational demand modeling. So far, this type of model has not been used to estimate the proportion of crashes by collision type. Consequently, the objective of this study consists of investigating the applicability of MNL models for predicting the proportion of crashes by collision type and their use in estimation of crash counts by collision type. This method is compared with two other methods described in recent publications for estimation of crash counts by collision type: 1) estimated using fixed proportions of crash counts for all collision types; 2) estimated using collision type models. To accomplish the study objective, crash data collected from 2002-2006 on rural two-lane undivided highway segments in Minnesota were employed. The results of this study show that the MNL model can be used for predicting the proportion of crashes by collision type, at least for the dataset used in this study. Furthermore, the method based on the MNL model was found useful in estimation of crash counts by collision type and it performed better than the method based on the use of fixed proportions. However, using collision type models was still found to be the best method for estimation of crash counts by specific collision type. In cases where collision type models are affected by the small sample size and low sample mean problem, the method based on MNL model is therefore recommended.
INTRODUCTION

Crash prediction models or Safety Performance Functions (SPFs) are still one of the primary tools for traffic safety analysis. They are needed because of the random nature of the crash process. Previous work related to the analysis of crash data occurring at intersections and on highway segments has mainly focused on developing crash prediction models which predict the total number of crashes for the entire facility, either for all or different severity levels (e.g., 1 - 4). Few studies have documented models that are used for predicting the number of crashes according to their collision type or manner of collision (e.g., see, 5 - 8). Evaluating crashes according to their collision type can provide important characteristics that cannot be captured using an aggregated model that combined all crashes together.

From the literature, there exist two different methods that have been used for predicting number of crashes according to their collision type (9). The first method is based on the assumption that the proportion of crash counts for all types remains fixed over time and for the entire range of the traffic flow; this method is referred to as the ‘fixed proportion method’ hereafter. Hence, with this method, a model for total crash counts (total crash model) is estimated and then the count of a specific crash type is estimated using the assumed proportion which may be obtained from the data. This simplification however comes at the cost of estimation error which can be attributed to the fact that the crash proportions at a site are not fixed and could vary as a function of traffic flow and highway geometric design characteristics.

The second method involves developing models corresponding to each crash type separately; this method is referred to as the ‘crash type model method’ hereafter. According to Kim et al. (10), estimating crash counts using collision type models has three main advantages. The first advantage is related to the fact that a total crash model by itself cannot identify a high risk location for a specific type of crash. The second advantage is that not all countermeasures are aimed at reducing crashes of all types simultaneously. Often, countermeasures are designed to reduce or influence specific crash types (e.g., head-on, cross-median or red-light running crashes). Hence, a more accurate estimation of the crash count by collision type is necessary and can be achieved by estimating a specific crash type model. The third advantage in estimating individual crash type models is that they can help in identifying different roadway, traffic and environmental variables which may affect each collision type differently.

Developing models by collision types also have some limitations. For instance, these models can be negatively influenced by the small sample size and low sample-mean problem (11 - 12). Since the data are disaggregated by collision type, the subset of the original data will have a smaller sample-mean value, which can negatively affect the estimation of the dispersion parameter of Poisson-gamma model. The models may also be less robust. Furthermore, the data may also contain many zeros for some subsets; this will influence or limit the selection of the appropriate modeling methodologies. In a few cases, some transportation safety analysts may erroneously believe that zero-inflated models are appropriate to analyze such data (see 13). Since the two proposed methods described above have limitations, there is a need to determine whether an alternative approach could be used for estimating crashes by collision type.

Recently, some researchers have started using Multinomial Logit (MNL) or other similar models to estimate crash severity levels as a function of various covariates, including highway geometric design features (14 - 16). Capitalizing on this body work, it may be possible to estimate crashes by collision type by multiplying the total crash counts (estimated using a total crash model) by the output of a MNL model. The MNL model is used to predict the probability
of a specific crash type given that a crash has occurred, as a function of factors that may influence the type of collision. This method is referred as the ‘MNL model method’ subsequently in the text.

The objectives of this research are two-fold. The first objective consists of examining the applicability of the MNL model for predicting the proportion of crashes by collision type. The second objective is to evaluate whether the output of the MNL model can be used to estimate crash count by collision type. To accomplish the study objectives, count data and MNL models were estimated using crashes that occurred on rural two-lane highways in Minnesota for the years 2002-2006.

This paper is organized as follows. The first section provides background information about relevant work done in crash data modeling. The second section describes the methodology used for estimating the count data and MNL models. The third section describes important characteristics of the Minnesota data. The fourth section presents the results of the analysis and associated discussion. The final section provides a summary of the research and outlines avenues for future work.

BACKGROUND

Over the last 20 years, a few researchers have developed crash prediction models by collision type. Hauer et al. (5) were the first to develop such models. They developed models for 15 different crash patterns at urban and suburban signalized intersections in Toronto, Canada.

Shankar et al. (6) developed models for six different crash types. They concluded that models predicting crashes for different crash types had a greater explanatory power than a single model that combined all crash types together. Kockelman and Kweon (17) developed crash type models (such as total, single-vehicle and multi-vehicle crashes) using ordered probit models to examine the risk of different driver injury severity levels. Their study estimated the safety effects on different drivers with different type of vehicles.

Qin et al. (18) developed zero-inflated Poisson models for different crash types (see 13, 19 for a discussion about the application of such models in highway safety). These authors examined crashes occurring on highway segments in Michigan and concluded that crashes are differently associated with traffic flows for different crash types. Abdel-Aty et al. (8) analyzed different crash types by considering seven different collision types at the intersections. The result of this study suggested that the influences of crash contributing factors change between different collision types.

Kim et al. (10) estimated several crash type models, where crashes were divided into seven different crash types. They concluded that a number of variables are related to crash types in different ways and suggested that crash types are associated with different pre-crash conditions. Jonsson et al. (20) developed distinct models for four different collision types occurring at intersections on rural four-lane highways in California. These authors concluded that the different crash type models exhibited dissimilar relationships with traffic flow and other covariates.

Recently, Jonsson et al. (9) developed crash type models to investigate the difference in estimation of crash counts by individual crash type models and total crashes with a fixed proportion. The crash type models were developed for four different types of crashes. They concluded that crash type models are preferred over estimating collision types using fixed proportions. Our study is a direct continuation of the work done by Jonsson et al. (9).
MNL model is one of the very popular econometric models used in the area of transportation engineering. Although these models are often used in transportation planning, some researchers have recently applied such models for crash data analysis. For example, Shankar and Mannering (14) used a MNL model specification for estimating motorcycle rider crash severity likelihood given that a crash has occurred. Carson and Mannering (15) developed MNL models to identify the effect of warning signs on ice-related crash severities on interstates, principal arterials, and minor arterial state highways. Finally, Abdel-Aty (16) also developed a MNL model for driver’s injury severity level and compared it with the ordered probit model. So far, nobody has used the MNL to predict the proportion of crashes by manner of collision.

**METHODOLOGY**

This section briefly explains the MLN model, count data models, goodness-of-fit statistics used in this study, and the steps that were used for estimating the various models.

**Multinomial Logit Model**

In this study, we describe the use of the MNL model for predicting probabilities for five discrete types of crashes given that a crash has occurred. An individual type of crash among the given five crash types is considered to be predicted if the crash type likelihood function is maximum for that particular type. Each crash type likelihood function, which is a dimensionless measure of the crash likelihood, is considered to be made up of a deterministic component and an error/random component. While the deterministic part is assumed to contain the variables which can be measured; the random part corresponds to the unaccounted factors that affect the prediction of a type of crash. We specify the deterministic part of the crash type likelihood as a linear function of the operational and segment specific characteristics as shown in Equation 1.

\[
V_{ij} = ASC_j + \alpha_{ij} \ln(F_i) + \alpha_{ij} \text{Truckpct}_i + \alpha_{ij} \text{Lanewid}_i + \alpha_{ij} \text{Shldwid}_i
\]  

(1)

where,

- \( V_{ij} \) = Systematic component of crash type likelihood for a segment \( i \) and crash type \( j \),
- \( ASC_j \) = Alternative specific constant for crash type \( j \),
- \( \alpha_{ij} \) = Coefficient (to be estimated) for crash type \( j \) and variable \( k, k = 1, \ldots, K \),
- \( F_i \) = Annual Average Daily Traffic (AADT) for segment \( i \),
- \( \text{Truckpct}_i \) = Percentage of trucks for segment \( i \),
- \( \text{Lanewid}_i \) = Average lane width for segment \( i \),
- \( \text{Shldwid}_i \) = Average shoulder width for segment \( i \).

The logit model assumes that the error components are extreme value (or gumbel) distributed and the probability of a discrete event (type of crash) is given by Equation 2 (21).
\[ P_{ij} = \frac{e^{V_i}}{\sum_{j=1}^{J} e^{V_j}} \]  

(2)

where,

- \( P_{ij} \) is the probability of the occurrence of crash type \( j \) for segment \( i \) and
- \( J \) is the total number of crash types to be modeled.

Though this assumption simplifies the probability equation it also adds the Independence from Irrelevant Alternatives (IIA) property in the MNL model. The IIA property of the MNL restricts the ratio of probabilities for any pair of alternatives to be independent of the existence and characteristics of other alternatives in the set of alternatives. This restriction implies that the introduction of a new alternative (crash type) in the set will affect all other alternatives proportionately. (22). The IIA property is a widely acknowledged limitation of MNL model. Hence, analysis in this paper will also be subjected to those limitations (see, (23) for discussion of IIA).

**Negative Binomial Regression**

Poisson-gamma (or negative binomial) models developed for this work have been shown to have the following probabilistic structure: the number of crashes \( Y_{it} \) at the \( i^{th} \) site (road section, intersections, etc.) and \( t^{th} \) time period, conditional on its mean \( \mu_{it} \), is assumed to be Poisson distributed and independent over all entities and time periods as:

\[ Y_{it} | \mu_{it} \sim \text{Poisson}(\mu_{it}), \quad i = 1, 2, \ldots, I \text{ and } t = 1, 2, \ldots, T \]  

(3)

The mean of the Poisson is structured as:

\[ \mu_{it} = f(X; \beta)\exp(e_{it}) \]  

(4)

where,

- \( f(.) \) is a function of the covariates (\( X \));
- \( \beta \) is a vector of unknown coefficients; and,
- \( e_{it} \) is a the model error independent of all the covariates.

It is usually assumed that \( \exp(e_{it}) \) is independent and gamma distributed with a mean equal to 1 and a variance \( 1 / \phi \) for all \( i \) and \( t \) (here \( \phi \) is the inverse of the dispersion parameter and \( \phi > 0 \); note \( \phi = 1/\alpha \)). With this characteristic, it can be shown that \( Y_{it} \), conditional on \( f(.) \) and \( \phi \), is distributed as a Poisson-gamma random variable with a mean \( f(.) \) and a variance \( f(.)(1 + f(.) / \phi) \) respectively.

The mean value (the number of crashes per year) for segment \( i \) and crash type \( j \) can be calculated by,
\[ \mu_{ij} = \beta_{0j} L_i^{\beta_{ij}} F_i^{\beta_{ij}} e^{(\beta_{ij} \text{Truckpct} + \beta_{ij} \text{Lanewid} + \beta_{ij} \text{Shldwid})} \]  \hspace{1cm} (5)

where,

- \( L_i \) = Length of segment \( i \) (in miles),
- \( F_i, \text{Truckpct}, \text{Lanewid}, \text{Shldwid} \) are as defined in Equation 1,
- \( \beta_{0j} \) = Intercept (to be estimated) for crash type \( j \),
- \( \beta_{ij} \) = Coefficients (to be estimated) for crash type \( j \) and variable \( k, k=1,..,K \).

**Goodness-of-Fit Statistics**

Different methods were used for evaluating the goodness-of-fit (GOF) and predictive performance of the models. The methods used included the following:

**Mean Absolute Deviance (MAD)**

The MAD provides a measure of the average mis-prediction of the model \( (23) \). It is computed using the following equation:

\[
\text{Mean Absolute Deviance (MAD)} = \frac{1}{n} \sum_{i=1}^{n} \left| \hat{y}_i - y_i \right| \hspace{1cm} (6)
\]

where, \( n \) is the sample size, \( \hat{y}_i \) and \( y_i \) are the predicted and observed crash counts at site \( i \) respectively.

**Mean Squared Predictive Error (MSPE)**

The MSPE is typically used to assess the error associated with a validation or external data set as given in Equation 7 \( (23) \).

\[
\text{Mean Squared Predictive Error (MSPE)} = \frac{1}{n} \sum_{i=1}^{n} \left( \hat{y}_i - y_i \right)^2 \hspace{1cm} (7)
\]

**Maximum Cumulative Residual Plot Deviation (MCPD)**

The MCPD is defined as the maximum absolute value that the Cumulative Residual (CURE) plot deviates from 0 \( (9) \). The residual is the difference between the observed and predicted crash frequencies. A CURE plot presents how the model fits the data with respect to each covariate by plotting the cumulative residuals in the increasing order for each key covariate. A better fit is presented when the cumulative residuals oscillate around the value of zero for that covariate.
Modeling Process

The study was carried out using the following 5-stage process:

1. First, a MNL model was estimated using LIMDEP (24) to predict the probability of a specific crash type given that a crash has occurred on a roadway segment. Various segment specific operational variables and geometric variables were used to predict the probability of a type of crash.
2. A total crash model and five individual crash type models were then developed in SAS (25) using the negative binomial modeling framework. The number of years and the segment length for each site were used as offsets. Estimates from these models directly yield the crash counts per segment for total crashes and for each crash type respectively.
3. The MNL model probabilities for each site were then multiplied by total number of crashes (estimated using total crash model) to estimate the crash counts for each crash type at a particular segment.
4. The fixed proportion for each crash type was directly calculated from the data by dividing the sum of specific type of crash count with total number of crashes. Later, multiplying these proportions with the estimate of total crashes (obtained from the total crash model) gave the crash counts for each crash type at a segment.
5. The goodness-of-fit statistics were then calculated to identify the best fit among the three approaches. Later, the predicted values corresponding to each crash type were plotted against AADT for each of these approaches to examine their relation with crashes.

DATA DESCRIPTION

The dataset used for this study contained crash data collected on rural two-lane undivided highway segments in Minnesota. The crash and network data for the years 2002-2006 were obtained from the Federal Highway Administration’s (FHWA) Highway Safety Information System (HSIS) website maintained by the University of North Carolina (www.hsis.org). The final database included 7,323 segments and five years of crash data. To estimate the individual crash type models and the MNL model, crashes were divided into five different collision types namely, head-on, rear-end, passing direction sideswipe, opposite direction sideswipe, and single-vehicle crashes. The dataset also contained variables corresponding to operational and segment characteristics such as AADT, percentage of trucks, segment length, lane width and average shoulder width. Summary statistics for the model variables are given in Table 1.

Table 1 here
RESULTS

This section describes the modeling results for the MNL and Poisson-gamma models, the GOF comparison analysis, and the relationship between crashes by collision pattern and vehicular traffic.

Multinomial Logit Model

Table 2 summarizes the MNL model to predict the probability of occurrence of a crash type. The rear-end collision was considered the base type scenario.

Table 2 here

In order to clearly visualize the effect of each of the variable (such as AADT) on the prediction of the proportion for each type of crash as estimated by the MNL model, we carried out further analyses. We estimated the proportion of each type of crash for different values of a particular variable while keeping the other variables constant (Figure 1). The following describes the findings of these analyses representing the applicability of the MNL models in predicting the proportions of crash counts as a function collision types.

An increase in AADT was found to decrease the proportion of all other types of crashes compared to rear-end crashes when everything else is held constant (Figure 1). In other words, the proportion of rear-end crashes increases greatly when compared to other crashes. This could be explained by fact that as traffic flow increases, the gaps between the vehicles decreases and the probability of a rear-end crash increases. Also, it can be observed that as the AADT increases, the proportion of single-vehicle crashes decreases with respect to rear-end crashes. This basically means that, for single-vehicle crashes, the crash risk per vehicle diminishes when traffic flow increases.

An increase in the percentage of trucks was found to decrease the proportion of single-vehicle and head-on crashes in comparison with rear-end crashes (Figure 1). Since trucks often travel at a slower speed than passenger cars, the speed differentials between vehicles increase. Due to the longer length of heavy vehicles, it is anticipated that the likelihood for passing or overtaking such vehicle will decrease on rural two-lane roads. This, in turn, could increase the proportion of rear-end crashes as the truck percentage increases. As a result, the proportion for all other crash types will decrease.

Increasing the lane width was found to decrease the proportion of rear-end crashes (Figure 1). Generally, the lane width is positively correlated with safety, as it allows drivers more room when the driver starts veering off the lane and regain control of the vehicle. With wider lane widths, it is possible that drivers have more opportunities to leave the traveled way rather than rear-ending the vehicle traveling in front in cases of an emergency or evasive maneuver. This decrease in the proportion of rear-end crashes will automatically lead to an increase in the proportions for the other crash types, as the sum of proportion of all the crash type counts must be equal to one. It should be pointed out that since we are dealing with proportions, this does not mean that more single-vehicle crashes occur when the lane width increases. It basically says that, if a crash occurs at on rural two-lane highways with a wider lane width, it is less likely to be classified as a rear-end collision.
An increase in shoulder width was also found to decrease the proportion of rear-end crashes (Figure 1). The same explanation as that of lane width applies here.

**Figure 1 here**

**Crash Count Models**

Six Poisson-gamma models were estimated to predict the total number of crashes and the number of crashes corresponding to the five collision types. The parameter estimates for these six models are summarized in Table 3. It can be observed from the individual crash type models estimates that the increase in AADT increases crash counts at an increasing rate for almost all the collision types whereas the increase in truck percentage, lane width and shoulder width decreases all types of crashes.

**Table 3 here**

**Goodness-of-fit Statistics for Three Modeling Methods**

The GOF statistics for each method with respect to prediction of individual type of crash counts are presented in Table 4. This table shows that the crash type model method outperforms the other two methods in predicting counts of all crash types except for rear-end crashes. Though, the MNL model method does not perform as well as the crash type model method, it outperforms the fixed proportion method for all crash patterns.

**Table 4 here**

It is important to note that the MNL model with only the alternate specific constants (constants only-MNL) predicts proportions which are equal to those obtained by the fixed proportion method. Hence, the comparison of the MNL model shown in Table 2 with the constants only-MNL model can be indicative of any advantage of using a more complex model such as the one in Table 2 over the fixed proportion method. A likelihood ratio (LR) test was carried out for these two models and the MNL model was found to be the best one as it offered a significant improvement over the constants only-MNL model in terms of the log-likelihood value (the reader is referred to 26 for the details of the LR test).

**Relationships between Flow and Collision Types**

We further analyzed how these methods predict the number of crashes by manner of collision as a function of traffic volume (AADT) when all other variables are held constant. The head-on crash counts predicted by the three approaches for increasing AADT are presented in Figure 2. For an increasing AADT (hence increasing opposing flow), the crash type and MNL models predict head-on crashes with a decreasing rate as traffic flow increases. In other words, as AADT increases, there are fewer head-on collisions per unit of exposure. The fixed proportion method shows that the head-on crashes increase linearly with the increase in AADT. Using the results shown in Table 4, we can assume that the crash type model provides more realistic trends.

**Figure 2 here**
Figure 3 shows the prediction of sideswipe-opposite crashes for the three different modeling approaches. The crash type model method for the sideswipe crashes predicts that these crashes increase with an increasing rate with AADT. The fixed proportion method still predicts a linear increase in the sideswipe-opposite type of crash counts with the increase in AADT. The MNL model method predicts similar trend as that of head-on crashes where the crashes increase with a decreasing rate, although it is almost linear. From the GOFs in Table 4, we can assume that the collision type model produces a realistic trend in this case.

Figure 3 here

The sideswipe-passing crash counts predicted by the three modeling approaches for increasing AADT are presented in Figure 4. For an increasing AADT, the crash type model predicts the sideswipe-passing crashes with an increasing rate, whereas and MNL model method predicts crashes with a decreasing rate. The fixed proportion method shows that the sideswipe-passing crashes increase linearly with the increase in AADT. From Table 4, we can see that the crash type model nearly provides realistic trends.

Figure 4 here

For an increasing AADT, the crash type and MNL models show that the rear-end crashes increases at an increasing rate (Figure 5). The rate of increase is larger with the crash type model than with the MNL model method. The fixed proportion method shows that the rear-end crashes increases almost linearly with an increase in vehicular traffic. From Table 4, it is clear that the MNL model method fit the data better and thus it provides more realistic trends than other approaches.

Figure 5 here

For single-vehicle crashes, the crash type and MNL models show that the number of crashes increases at a decreasing rate as traffic flow increases, as seen in Figure 6. Since the fixed proportion method applies a rigid proportion irrespective of AADT, the trend shown by this approach is linear. As indicated in Table 4, though the model fit is much different between MNL model method and crash type model, they both provide realistic trends.

Figure 6 here

SUMMARY AND CONCLUSIONS

The objectives of this study were to examine the applicability of multinomial logit (MNL) model for predicting the proportion of crashes by collision type and to evaluate whether the output of the MNL model can be used to estimate crash counts by collision type. Crash data collected from rural two-lane highway segments in Minnesota for years 2002-2006 were used for comparing this approach with the two previous approaches documented in the literature: crash type models and collision types estimated using fixed proportions. Crashes that occurred on these segments were divided into five different categories: head-on, rear-end, passing direction sideswipe, opposite direction sideswipe, and single-vehicle crashes.
The application of the multinomial logit model for estimating the proportion of crashes by collision type seems to be promising. The effects of different variables on the occurrence of each crash type were found to meet prior expectations. Furthermore, when the output of the MNL model was used to estimate crash counts by collision type, it performed better than the fixed proportion method with respect to three goodness-of-fit criteria used in this study. The fixed proportion method hence failed to generate realistic trends with increase in the traffic flow volumes for all crash counts by collision type. Predicting crash counts by specific crash type models was, nonetheless, found to be the best method, as documented in Jonsson et al. (9).

However, it should be noted that developing models for collision types can be negatively influenced by the small sample size and low sample-mean problem (11). Using a logit model (such as a MNL model) for estimating the crash count by collision type is recommended if count data models are affected by this problem.

Three avenues for further work on this topic are as follows:

1. This study used data collected on rural two-lane highways and it can be extended for multiline high speed highways and intersections to check if the findings are similar for other type of datasets.
2. The study can also be broadened to include more collision types.
3. Since the MNL model suffers from methodological limitations, it is therefore suggested to evaluate the application of mixed logit models for estimating collision patterns. Mixed logit models relax the assumption of independence from irrelevant alternatives (IIA) and also make it possible to allow for heterogeneity from a variety of sources (see, 26).

REFERENCES


LIST OF TABLES AND FIGURES

TABLE 1 Summary Statistics for the Data
TABLE 2 Modeling Results for the MNL Model
TABLE 3 Estimation Results for Different Crash Type Models
TABLE 4 Goodness-of-fit Statistics (GOFs)

FIGURE 1 Effect of different variables on the proportion of crashes by collision pattern.
FIGURE 2 Predicted number of head-on crashes as a function of AADT.
FIGURE 3 Predicted number of sideswipe-opposite direction crashes as a function of AADT.
FIGURE 4 Predicted number of sideswipe-passing crashes as a function of AADT.
FIGURE 5 Predicted Number of Rear-End Crashes as a Function of AADT.
FIGURE 6 Predicted number of single vehicle crashes as a function of AADT.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Min</th>
<th>Max</th>
<th>Average (Std Dev)</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segment Length (mile)</td>
<td>0.016</td>
<td>12.915</td>
<td>1.036 (1.309)</td>
<td>7588.692</td>
</tr>
<tr>
<td>Lane Width (ft)</td>
<td>10</td>
<td>20.2</td>
<td>12.15 (0.62)</td>
<td>--</td>
</tr>
<tr>
<td>Average Shoulder Width (ft)</td>
<td>0</td>
<td>15</td>
<td>6.4 (3.1)</td>
<td>--</td>
</tr>
<tr>
<td>AADT</td>
<td>52.2</td>
<td>32220.8</td>
<td>3349.1 (2852.4)</td>
<td>--</td>
</tr>
<tr>
<td>Truck Percentages</td>
<td>1.5 %</td>
<td>65.8 %</td>
<td>10.95 %</td>
<td>--</td>
</tr>
<tr>
<td>Total Crashes</td>
<td>0</td>
<td>70</td>
<td>1.99 (3.43)</td>
<td>14586</td>
</tr>
<tr>
<td>Head-on Crashes</td>
<td>0</td>
<td>7</td>
<td>0.23 (0.60)</td>
<td>1700</td>
</tr>
<tr>
<td>Sideswipe- Opposite Crashes</td>
<td>0</td>
<td>6</td>
<td>0.10 (0.36)</td>
<td>748</td>
</tr>
<tr>
<td>Sideswipe- Passing Crashes</td>
<td>0</td>
<td>6</td>
<td>0.15 (0.47)</td>
<td>1102</td>
</tr>
<tr>
<td>Rear-end Crashes</td>
<td>0</td>
<td>57</td>
<td>0.50 (1.62)</td>
<td>3668</td>
</tr>
<tr>
<td>Single Vehicle Crashes</td>
<td>0</td>
<td>48</td>
<td>1.01 (1.97)</td>
<td>7368</td>
</tr>
<tr>
<td>Variables</td>
<td>Estimate</td>
<td>t-ratio</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>----------</td>
<td>---------</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Log(AADT)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Head-on Crashes</td>
<td>-0.874</td>
<td>-20.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sideswipe- Opposite Crashes</td>
<td>-0.525</td>
<td>-10.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sideswipe- Passing Crashes</td>
<td>-0.507</td>
<td>-11.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rear-end Crashes**</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single Vehicle Crashes</td>
<td>-1.203</td>
<td>-39.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Shoulder Width (ft)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Head-on Crashes</td>
<td>0.116</td>
<td>11.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sideswipe- Opposite Crashes</td>
<td>0.094</td>
<td>7.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sideswipe- Passing Crashes</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rear-end Crashes**</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single Vehicle Crashes</td>
<td>0.092</td>
<td>13.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Percentage of Trucks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Head-on Crashes</td>
<td>-0.017</td>
<td>-2.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sideswipe- Opposite Crashes</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sideswipe- Passing Crashes</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rear-end Crashes**</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single Vehicle Crashes</td>
<td>-0.033</td>
<td>-8.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Lane Width (ft)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Head-on Crashes</td>
<td>0.179</td>
<td>3.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sideswipe- Opposite Crashes</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sideswipe- Passing Crashes</td>
<td>0.121</td>
<td>2.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rear-end Crashes**</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single Vehicle Crashes</td>
<td>0.127</td>
<td>3.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Alternative Specific Constant</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Head-on Crashes</td>
<td>3.781</td>
<td>5.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sideswipe- Opposite Crashes</td>
<td>2.246</td>
<td>5.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sideswipe- Passing Crashes</td>
<td>1.588</td>
<td>1.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rear-end Crashes**</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single Vehicle Crashes</td>
<td>8.813</td>
<td>17.28</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Number of Observations 14500
Log-likelihood at convergence -17645
Adjusted $\rho^2$ for constants only model 0.056

**Base alternative, * insignificant at 5% level of confidence
### TABLE 3 Estimation Results for Different Crash Type Models

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Total Crashes</th>
<th>Head-on Opposite</th>
<th>Sideswipe-Passing</th>
<th>Rear-end Single-Vehicle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept ($\ln(\beta_0)$)</td>
<td>-6.4628 (0.3060)</td>
<td>-10.0665 (0.3181)</td>
<td>-11.6021 (0.9490)</td>
<td>-13.0764 (0.3589)</td>
</tr>
<tr>
<td>Ln (AADT) ($\beta_1$)</td>
<td>1.0634 (0.0175)</td>
<td>0.9563 (0.0378)</td>
<td>1.3143 (0.0508)</td>
<td>1.3438 (0.0444)</td>
</tr>
<tr>
<td>Truck fraction ($\beta_2$)</td>
<td>-0.0158 (0.0030)</td>
<td>-0.0163 (0.0064)</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Lane Width ($\beta_3$)</td>
<td>-0.1552 (0.0228)</td>
<td>*</td>
<td>-0.1883 (0.0704)</td>
<td>*</td>
</tr>
<tr>
<td>Shoulder Width ($\beta_4$)</td>
<td>-0.1038 (0.0045)</td>
<td>-0.0420 (0.0101)</td>
<td>-0.0557 (0.0137)</td>
<td>-0.1561 (0.0112)</td>
</tr>
<tr>
<td>Dispersion ($\alpha$)</td>
<td>0.4965 (0.0204)</td>
<td>0.5060 (0.0799)</td>
<td>0.4333 (0.1464)</td>
<td>1.0283 (0.1502)</td>
</tr>
</tbody>
</table>

* insignificant at 5% level of confidence
<table>
<thead>
<tr>
<th>Collision Type</th>
<th>MAD</th>
<th>MSPE</th>
<th>MCPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head On</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MNL Model Method</td>
<td>0.317</td>
<td>0.289</td>
<td>143.10</td>
</tr>
<tr>
<td>Fixed Proportion Method</td>
<td>0.318</td>
<td>0.296</td>
<td>132.03</td>
</tr>
<tr>
<td>Collision Type Model</td>
<td>0.311</td>
<td>0.286</td>
<td>48.75</td>
</tr>
<tr>
<td>Sideswipe-Opposite</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAD</td>
<td>0.168</td>
<td>0.114</td>
<td>70.08</td>
</tr>
<tr>
<td>MSPE</td>
<td>0.169</td>
<td>0.114</td>
<td>108.18</td>
</tr>
<tr>
<td>MCPD</td>
<td>0.163</td>
<td>0.113</td>
<td>24.84</td>
</tr>
<tr>
<td>Sideswipe-Passing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAD</td>
<td>0.233</td>
<td>0.198</td>
<td>94.51</td>
</tr>
<tr>
<td>MSPE</td>
<td>0.236</td>
<td>0.197</td>
<td>153.19</td>
</tr>
<tr>
<td>MCPD</td>
<td>0.231</td>
<td>0.198</td>
<td>75.44</td>
</tr>
<tr>
<td>Rear-end</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAD</td>
<td>0.595</td>
<td>1.976</td>
<td>404.40</td>
</tr>
<tr>
<td>MSPE</td>
<td>0.647</td>
<td>2.212</td>
<td>1060.46</td>
</tr>
<tr>
<td>MCPD</td>
<td>0.619</td>
<td>2.008</td>
<td>659.01</td>
</tr>
<tr>
<td>Single-Vehicle</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAD</td>
<td>0.864</td>
<td>2.075</td>
<td>586.26</td>
</tr>
<tr>
<td>MSPE</td>
<td>0.902</td>
<td>2.429</td>
<td>707.5113</td>
</tr>
<tr>
<td>MCPD</td>
<td>0.846</td>
<td>2.128</td>
<td>201.72</td>
</tr>
</tbody>
</table>
FIGURE 1 Effect of different variables on the proportion of crashes by collision pattern.
FIGURE 2 Predicted number of head-on crashes as a function of AADT.
FIGURE 3 Predicted number of sideswipe-opposite direction crashes as a function of AADT.
FIGURE 4 Predicted number of sideswipe-passing crashes as a function of AADT.
FIGURE 5 Predicted Number of Rear-End Crashes as a Function of AADT.
FIGURE 6 Predicted number of single vehicle crashes as a function of AADT.