# Examination of Methods to Estimate Crash Counts by Collision Type

Srinivas Reddy Geedipally, Sunil Patil, and Dominique Lord

**Multinomial logit (MNL) models have been applied extensively in transportation engineering, marketing, and recreational demand modeling. Thus far, this type of model has not been used to estimate the proportion of crashes by collision type. This study investigated the applicability of MNL models to predict the proportion of crashes by collision type and to estimate crash counts by collision type. MNL models were compared with two other methods described in recent publications to estimate crash counts by collision type: (*a*) fixed proportions of crash counts for all collision types and (*b*) collision type models. This study employed data collected between 2002 and 2006 on crashes that occurred on rural, two-lane, undivided highway segments in Minnesota. The study results showed that the MNL model could be used to predict the proportion of crashes by collision type, at least for the data set used. Furthermore, the method based on the MNL model was found useful to estimate crash counts by collision type, and it performed better than the method based on the use of fixed proportions. The use of collision type models, however, was still found to be the best way to estimate crash counts by specific collision type. In cases where collision type models are affected by the small sample size and a low sample-mean problem, the method based on the MNL model is recommended.**

Crash prediction models, or safety performance functions, are still among the primary tools used to analyze traffic safety. They are needed because of the random nature of the crash process. Previous data analyses of crashes that occurred at intersections and on highway segments have focused on the development of crash prediction models to predict the total number of crashes for an entire facility, either at all or various severity levels (*1–4*). Few studies have documented models that predict the number of crashes according to collision type or manner (*5–8*). Evaluation of crashes according to collision type can reveal important characteristics, which cannot be captured by an aggregated model that combines all crashes together.

The literature on research to date describes two methods that have been used to predict number of crashes according to 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 traffic flow. Referred to here as the fixed proportion method, it works as follows: A model for total crash counts (total crash model) is estimated. The count of a specific crash type then is estimated through the use of 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 development of models that correspond to each crash type separately; it is referred to here as the crash type model method. According to Kim et al., estimation of crash counts by using collision type models has three main advantages (*10*). First, a total crash model cannot by itself identify a high-risk location for a specific type of crash. Second, not all countermeasures aim to reduce 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 the estimation of individual crash type models is that they can help to identify roadway, traffic, and environmental variables that may affect each collision type differently.

Development of models by collision types also has limitations. Such models can be negatively influenced by the small sample size and low sample-mean problem (*11–12*). Because the data are disaggregated by collision type, the subset of original data will have a smaller sample-mean value, which can negatively affect the estimation of the dispersion parameter of a Poisson–gamma model. Such models may also be less robust than others. Furthermore, the data may 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 (*13*). Because the two proposed methods described above do have limitations, there is a need to determine if an alternative approach could be used to estimate crashes by collision type.

Recently, researchers have started to use multinomial logit (MNL) or similar models to estimate crash severity levels as a function of various covariates, including highway geometric design features (*14–16*). By capitalizing on this body of work, it may be possible to estimate crashes by collision type. This might be done by multiplying the total crash counts (estimated through the use of a total crash model) times the output of an 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 to here as the MNL model method.

The objectives of this research were twofold: (*a*) to examine the applicability of the MNL model to predict the proportion of crashes by collision type; and (*b*) to evaluate whether the output of the MNL model could be used to estimate crash count by collision type. To accomplish these objectives, count data and MNL models were estimated by using data from crashes that occurred on rural, two-lane highways in Minnesota between 2002 and 2006.

S. R. Geedipally, Texas Transportation Institute, D. Lord, Zachry Department of Civil Engineering, and S. Patil, Texas A&M University, 3136 TAMU, College Station, TX 77843-3136. Corresponding author: S. R. Geedipally, srinivas8@tamu.edu.

*Transportation Research Record: Journal of the Transportation Research Board, No. 2165,* Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 12–20.
DOI: 10.3141/2165-02

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

Shankar et al. developed models for six crash types (*6*). They concluded that models that predict crashes for different crash types have a greater explanatory power than a single model that predicts for all crash types combined together. Kockelman and Kweon developed crash type models (e.g., total, single-vehicle, and multivehicle crashes) by using ordered-probit models to examine the risk associated with various driver injury severity levels (*17*). Their study estimated the safety effects on drivers of different types of vehicles.

Qin et al. developed zero-inflated Poisson models for different crash types (*18*). [See Lord et al. for a discussion of the application of such models in highway safety (*13, 19).*] Qin et al. examined crashes that occurred on highway segments in Michigan and concluded that crashes have different associations with traffic flows for different crash types. Abdel-Aty et al. considered seven collision types at intersections (*8*). The results of their study suggested that the influences of contributing factors vary across collision types.

Kim et al. estimated several crash type models, where crashes were divided into seven types (*10*). 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. developed distinct models for four collision types that occurred at intersections on rural, four-lane highways in California (*20*). These authors concluded that the crash type models exhibited dissimilar relationships with traffic flow and other covariates.

Recently, Jonsson et al. 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 (*9*). The crash type models were developed for four types of crashes. They concluded that crash type models are preferred over the estimation of collision types through the use of fixed proportions. The study reported in this paper is a direct continuation of the work done by Jonsson et al. (*9*).

MNL models are among the popular econometric models used in transportation engineering and planning. Recently, some researchers have begun to apply them to crash data analysis. Shankar and Mannering used an MNL model specification to estimate motorcycle rider crash severity likelihood, given that a crash had occurred (*14*). Carson and Mannering developed MNL models to identify the effect of warning signs on ice-related crash severities on interstates, principal arterials, and minor arterial state highways (*15*). Finally, Abdel-Aty developed an MNL model for driver's injury severity level and compared it with the ordered-probit model (*16*). Thus far, no one has used the MNL to predict the proportion of crashes by manner of collision.

## METHODOLOGY

This section briefly explains the MNL model, count data models, goodness-of-fit (GOF) statistics, and the steps used to estimate the various models.

### Multinomial Logit Model

In this study, the use of the MNL model was described to predict probabilities for five, discrete types of crashes, given that a crash had occurred. An individual type of crash among the given five crash types was considered to be predicted if the crash type likelihood function was the maximum for that particular type. Each crash type likelihood function, which is a dimensionless measure of the crash likelihood, was considered to be made up of a deterministic component and an error/random component. Although the deterministic part was assumed to contain the variables that could be measured, the random part corresponded to the unaccounted factors that affected the prediction of a type of crash. The deterministic part of the crash type likelihood was specified as a linear function of the operational and segment-specific characteristics, as shown in Equation 1.

$$V_{ij} = \mathrm{ASC}_j + \alpha_{1j}\ln\left(F_i\right) + \alpha_{2j}\mathrm{truckpct}_i + \alpha_{3j}\mathrm{lanewid}_i + \alpha_{4j}\mathrm{shldwid}_i \quad (1)$$

where

$V_{ij}$ = systematic component of crash type likelihood for a segment $i$ and crash type $j$;
$\mathrm{ASC}_j$ = alternative specific constant for crash type $j$;
$\alpha_{kj}$ = coefficient (to be estimated) for crash type $j$ and variable $k$, $k = 1, \ldots, K$;
$F_i$ = annual average daily traffic (AADT) for segment $i$;
$\mathrm{truckpct}_i$ = percentage of trucks for segment $i$;
$\mathrm{lanewid}_i$ = average lane width for segment $i$; and
$\mathrm{shldwid}_i$ = 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*).

$$P_{ij} = \frac{e^{V_{ij}}}{\sum_{j=1}^{J} e^{V_{ij}}} \quad (2)$$

where $P_{ij}$ is the probability of occurrence of crash type $j$ for segment $i$, and $J$ is the total number of crash types to be modeled.

Although 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 a set of alternatives. This restriction implies that the introduction of a new alternative (crash type) to the set will affect all other alternatives proportionately (*22*). The IIA property is a widely acknowledged limitation of the MNL model. Hence, the analysis described in this paper was subject to the same limitations. For a discussion of IIA, see Oh et al. (*23*).

### 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 (e.g., road section, intersections) 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

$$Y_{it} \mid \mu_{it} \sim \text{Poisson}(\mu_{it}) \qquad i = 1, 2, \ldots, I \text{ and } t = 1, 2, \ldots, T \tag{3}$$

The mean of the Poisson is structured as

$$\mu_{it} = f(X; \beta)\exp(e_{it}) \tag{4}$$

where

$f(.)$ = function of the covariates ($X$),
$\beta$ = vector of unknown coefficients, and
$e_{it}$ = 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

$$\mu_{ij} = \beta_{0j} L_i F_i^{\beta_{1j}} e^{\left(\beta_{2j}\text{truckpct} + \beta_{3j}\text{lanewid}_i + \beta_{4j}\text{shldwid}_i\right)} \tag{5}$$

where

$L_i$ = length of segment $i$ (in miles),
$\beta_{0j}$ = intercept (to be estimated) for crash type $j$, and
$\beta_{kj}$ = coefficients (to be estimated) for crash type $j$ and variable $k$, $k = 1, \ldots, K$.

### GOF Statistics

Different methods were used to evaluate the GOF and predictive performance of the models. The methods used included the following.

#### *Mean Absolute Deviance*

The mean absolute deviance (MAD) provides a measure of the average misprediction of the model (*23*). It is computed by using the following equation:

$$\text{MAD} = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right| \tag{6}$$

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

#### *Mean Squared Predictive Error*

Typically, the mean squared predictive error (MSPE) is used to assess the error associated with a validation or external data set as given in Equation 7 (*23*).

$$\text{MSPE} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2 \tag{7}$$

#### *Maximum Cumulative Residual Plot Deviation*

The maximum cumulative residual plot deviation (MCPD) is defined as the maximum absolute value that the cumulative residual (CURE) plot deviates from zero (*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 by using the following five-stage process:

1. An MNL model was estimated by using the LIMDEP software program to predict the probability of a specific crash type, given that a crash had occurred on a roadway segment (*24*). 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 software by using the negative binomial modeling framework (*25*). The number of years and the segment length for each site were used as offsets. Estimates from these models directly yielded 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 the total number of crashes (estimated by using the 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 the specific type of crash count from the total number of crashes. Later, multiplication of 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 GOF statistics were then calculated to identify the best fit among the three approaches. Later, the predicted values that corresponded to each crash type were plotted against AADT for each of these approaches to examine their relationships to the crashes.

## DATA DESCRIPTION

The data set used in this study contained crash data collected on rural, two-lane, undivided highway segments in Minnesota. Crash and network data for the years 2002 through 2006 were obtained from the FHWA Highway Safety Information System website maintained by the University of North Carolina (*26*). The final database included 7,323 segments and 5 years of crash data. To estimate the individual crash type models and the MNL model, crashes were divided into five collision types, namely, head-on, rear-end, passing direction sideswipe, opposite direction sideswipe, and single-vehicle crashes. The data set also contained variables that corresponded 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 Summary Statistics for Data**

| Variable | Minimum | Maximum | Average (SD) | Total |
|---|---|---|---|---|
| Segment length (mi) | 0.016 | 12.915 | 1.036 (1.309) | 7,588.692 |
| Lane width (ft) | 10 | 20.2 | 12.15 (0.62) | — |
| Average shoulder width (ft) | 0 | 15 | 6.4 (3.1) | — |
| AADT | 52.2 | 32,220.8 | 3,349.1 (2852.4) | — |
| Truck percentages | 1.5 | 65.8 | 10.95 | — |
| Total crashes | 0 | 70 | 1.99 (3.43) | 14,586 |
| Head-on crashes | 0 | 7 | 0.23 (0.60) | 1,700 |
| Sideswipe–opposite crashes | 0 | 6 | 0.10 (0.36) | 748 |
| Sideswipe–passing crashes | 0 | 6 | 0.15 (0.47) | 1,102 |
| Rear-end crashes | 0 | 57 | 0.50 (1.62) | 3,668 |
| Single-vehicle crashes | 0 | 48 | 1.01 (1.97) | 7,368 |

NOTE: — = total value cannot be computed.

## 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.

### MNL 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.

Further analyses were carried out to clearly visualize the effect of each variable (e.g., AADT) on the prediction of the proportion for each type of crash as estimated by the MNL model. The proportion of each type of crash was estimated for different values of a particular variable while the other variables were kept constant (Figure 1). The following paragraphs describe findings from the analyses of the applicability of the MNL models to predict the proportions of crash counts as a function of collision types.

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

An increase in the percentage of trucks was found to decrease the proportion of single-vehicle and head-on crashes compared with rear-end crashes (Figure 1). Since trucks often travel at a slower speed than passenger cars, the speed differentials between vehicles increase. Heavy vehicles tend to be longer in length, and the likelihood of passing or overtaking them is anticipated to decrease on rural, two-lane roads. The proportion of rear-end crashes could in turn increase as the truck percentage increased. As a result, the proportion of all other crash types would decrease.

An increase in lane width was found to decrease the proportion of rear-end crashes (Figure 1). Broad lane width is positively correlated with safety, as it allows more room for drivers to correct course if they veer out of lane. It is also possible that wider lanes give drivers

**TABLE 2 Modeling Results for MNL Model**

| Variable | Estimate | *t*-Ratio |
|---|---|---|
| Log(AADT) | | |
| Head-on crashes | −0.874 | −20.85 |
| Sideswipe–opposite crashes | −0.525 | −10.03 |
| Sideswipe–passing crashes | −0.507 | −11.76 |
| Rear-end crashes[a] | | 0 |
| Single-vehicle crashes | −1.203 | −39.27 |
| Shoulder Width (ft) | | |
| Head-on crashes | 0.116 | 11.48 |
| Sideswipe–opposite crashes | 0.094 | 7.14 |
| Sideswipe–passing crashes | | [b] |
| Rear-end crashes[a] | | 0 |
| Single-vehicle crashes | 0.092 | 13.78 |
| Percentage of Trucks | | |
| Head-on crashes | −0.017 | −2.93 |
| Sideswipe–opposite crashes | | [b] |
| Sideswipe–passing crashes | | [b] |
| Rear-end crashes[a] | | 0 |
| Single-vehicle crashes | −0.033 | −8.18 |
| Lane Width (ft) | | |
| Head-on crashes | 0.179 | 3.51 |
| Sideswipe–opposite crashes | | [b] |
| Sideswipe–passing crashes | 0.121 | 2.01 |
| Rear-end crashes[a] | | 0 |
| Single-vehicle crashes | 0.127 | 3.46 |
| Alternative Specific Constant | | |
| Head-on crashes | 3.781 | 5.38 |
| Sideswipe–opposite crashes | 2.246 | 5.23 |
| Sideswipe–passing crashes | 1.588 | 1.95 |
| Rear-end crashes[a] | | 0 |
| Single-vehicle crashes | 8.813 | 17.28 |
| Number of observations | 14,500 | |
| Log likelihood at convergence | −17,645 | |
| Adjusted $\rho^2$ for constants-only model | 0.056 | |

[a]Base alternative.
[b]Insignificant at 5% level of confidence.

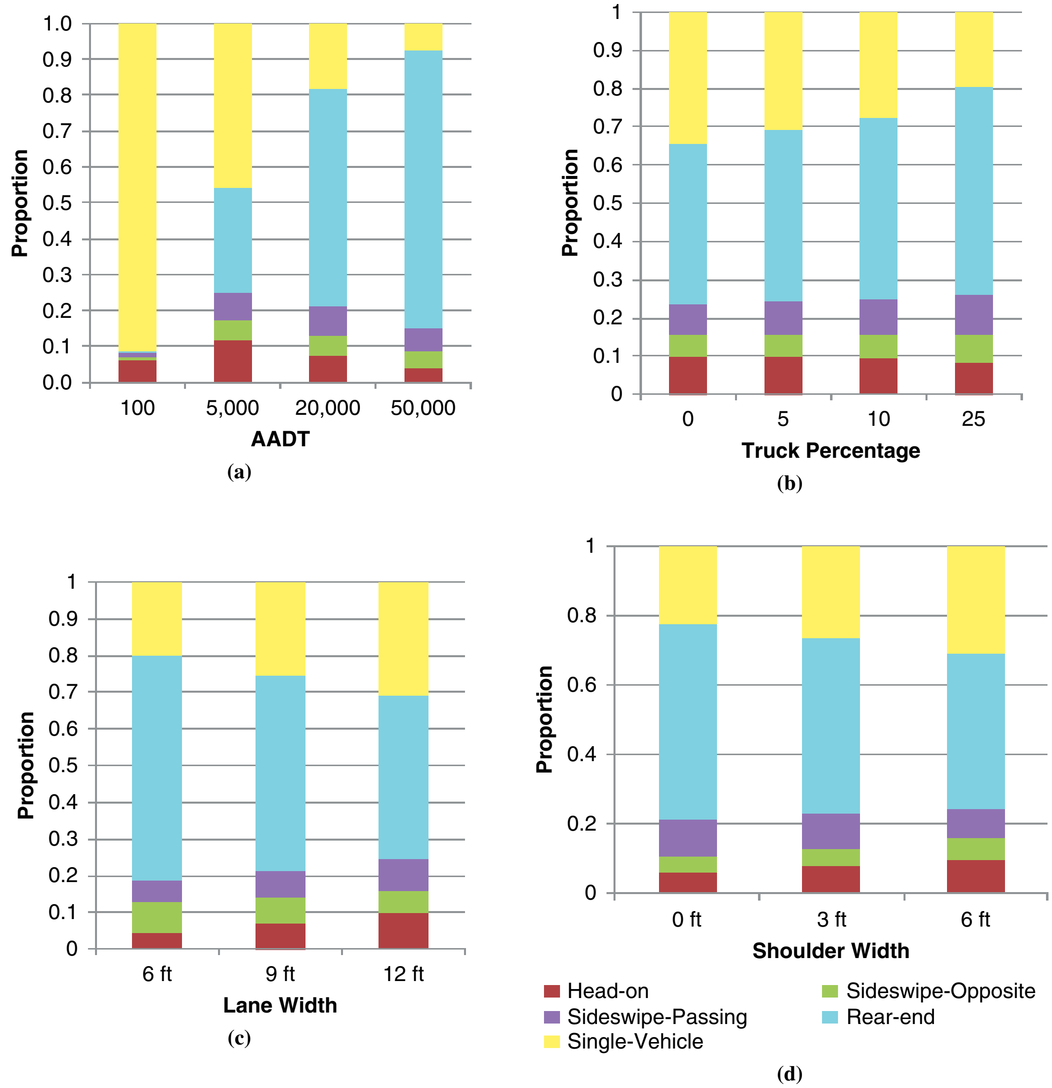


FIGURE 1 Effect of different variables on proportion of crashes by collision pattern: (*a*) AADT, (*b*) truck percentage, (*c*) lane width, and (*d*) shoulder width.

more opportunities in emergencies to leave the traveled way or to otherwise maneuver evasively rather than rear-end the vehicle in front of them. This decrease in the proportion of rear-end crashes will automatically lead to an increase in the proportions of other crash types, as the sum of proportion of all the crash type counts must be equal to 1. The fact that proportions are involved here does not mean that more single-vehicle crashes occur when the lane width increases. Basically it means that a crash on a rural, two-lane highway with a wider lane width is less likely to be classified as a rear-end collision than one that occurs in a narrower lane.

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.

## Crash Count Models

Six Poisson–gamma models were estimated to predict the total number of crashes and the number of crashes that corresponded to the five collision types. The parameter estimates for these six models are summarized in Table 3. As can be observed from the individual crash type model estimates, an increase in AADT increased crash counts at a rising rate for almost all collision types, whereas an increase in truck percentage, lane width, and shoulder width decreased all types of crashes.

## GOF 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 outperformed the other two methods to predict counts of all crash types except for rear-end crashes. Although the MNL model method did not perform as well as the crash type model method, it outperformed the fixed proportion method for all crash patterns.

The MNL model with only the alternate specific constants (constants-only MNL) predicted proportions equal to those obtained by the fixed proportion method. A comparison of the MNL model shown in Table 2 with the constants-only MNL model may be

**TABLE 3 Estimation Results for Different Crash Type Models**

| | Estimates (standard error) | | | | | |
|---|---|---|---|---|---|---|
| Parameter | Total Crashes | Head-on | Sideswipe–Opposite | Sideswipe–Passing | Rear-end | Single-Vehicle |
| Intercept [$\ln(\beta_0)$] | –6.4628 (0.3060)[1] | –10.0665 (0.3181) | –11.6021 (0.9490) | –13.0764 (0.3589) | –12.9604 (0.6418) | –4.3921 (0.3539) |
| ln(AADT) ($\beta_1$) | 1.0634 (0.0175) | 0.9563 (0.0378) | 1.3143 (0.0508) | 1.3438 (0.0444) | 1.7897 (0.0346) | 0.6166 (0.0214) |
| Truck percentage ($\beta_2$) | –0.0158 (0.0030) | –0.0163 (0.0064) | [a] | [a] | [a] | –0.0298 (0.0035) |
| Lane width ($\beta_3$) | –0.1552 (0.0228) | [a] | –0.1883 (0.0704) | [a] | –0.2177 (0.0474) | –0.1044 (0.0266) |
| Shoulder width ($\beta_4$) | –0.1038 (0.0045) | –0.0420 (0.0101) | –0.0557 (0.0137) | –0.1561 (0.0112) | –0.1429 (0.0082) | –0.0633 (0.0059) |
| Dispersion ($\alpha$) | 0.4965 (0.0204) | 0.5060 (0.0799) | 0.4333 (0.1464) | 1.0283 (0.1502) | 1.5232 (0.0862) | 0.4590 (0.0281) |

[a]Insignificant at 5% level of confidence.

indicative of any advantages to be had from a more complex model 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. For LR test details, see Train (*27*).

## Relationships Between Flow and Collision Types

Further analysis was done to see how these methods would predict the number of crashes by manner of collision as a function of traffic volume (AADT) when all other variables were held constant. Figure 2 shows the head-on crash counts predicted by the three approaches with an increasing AADT (and hence an increasing opposing flow). The crash type and MNL models predicted head-on crashes with a decreasing rate as traffic flow increased. As AADT increased, fewer head-on collisions occurred per unit of exposure. The fixed proportion method showed that the head-on crashes increased linearly with the increase in AADT. On the basis of the results shown in Table 4, it can be assumed that the crash type model provided more realistic trends.

**TABLE 4 Goodness-of-Fit Statistics**

| Crash Type | GOF Test | MNL Model Method | Fixed Proportion Method | Collision Type Model |
|---|---|---|---|---|
| Head-on | MAD | 0.317 | 0.318 | **0.311** |
| | MSPE | 0.289 | 0.296 | **0.286** |
| | MCPD | 143.10 | 132.03 | **48.75** |
| Sideswipe–opposite | MAD | 0.168 | 0.169 | **0.163** |
| | MSPE | 0.114 | 0.114 | **0.113** |
| | MCPD | 70.08 | 108.18 | **24.84** |
| Sideswipe–passing | MAD | 0.233 | 0.236 | **0.231** |
| | MSPE | 0.198 | **0.197** | 0.198 |
| | MCPD | 94.51 | 153.19 | **75.44** |
| Rear-end | MAD | **0.595** | 0.647 | 0.619 |
| | MSPE | **1.976** | 2.212 | 2.008 |
| | MCPD | **404.40** | 1,060.46 | 659.01 |
| Single-vehicle | MAD | 0.864 | 0.902 | **0.846** |
| | MSPE | **2.075** | 2.429 | 2.128 |
| | MCPD | 586.26 | 707.5113 | **201.72** |

NOTE: Bold numbers represent best fit.

Figure 3 shows the prediction of sideswipe–opposite crashes for the three modeling approaches. The crash type model method for the sideswipe crashes predicted that these crashes would increase with an increasing rate with AADT. The fixed proportion method still predicted a linear increase in the sideswipe–opposite type of crash counts with the increase in AADT. The MNL model method predicted a trend similar to that of head-on crashes where the crashes increased with a decreasing rate, although it was almost linear. On the basis of the GOFs shown in Table 4, it can be assumed that the collision type model produced a realistic trend in this case.

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 predicted the sideswipe–passing crashes with an increasing rate, whereas the MNL model method predicted crashes with a decreasing rate. The fixed proportion method showed that the sideswipe–passing crashes increased linearly with the increase in AADT. As Table 4 shows, the crash type model nearly provided realistic trends.

For an increasing AADT, the crash type and MNL models showed that the rear-end crashes increased at an increasing rate (Figure 5). The rate of increase was higher with the crash type model than with the MNL model method. The fixed proportion method showed that the rear-end crashes increased almost linearly with an increase in vehicular traffic. Table 4 makes it clear that the MNL model method fit the data better, and thus it provided more realistic trends than other approaches.

For single-vehicle crashes, the crash type and MNL models showed that the number of crashes increased at a decreasing rate as traffic flow increased (Figure 6). Since the fixed proportion method applies a rigid proportion irrespective of AADT, the trend shown by this approach was linear. As indicated in Table 4, although the model fit was quite different between the MNL model method and the crash type model, they both provided realistic trends.

## SUMMARY AND CONCLUSIONS

The objectives of this study were to examine the applicability of the MNL model to predict the proportion of crashes by collision type and to evaluate whether the output of the MNL model could be used

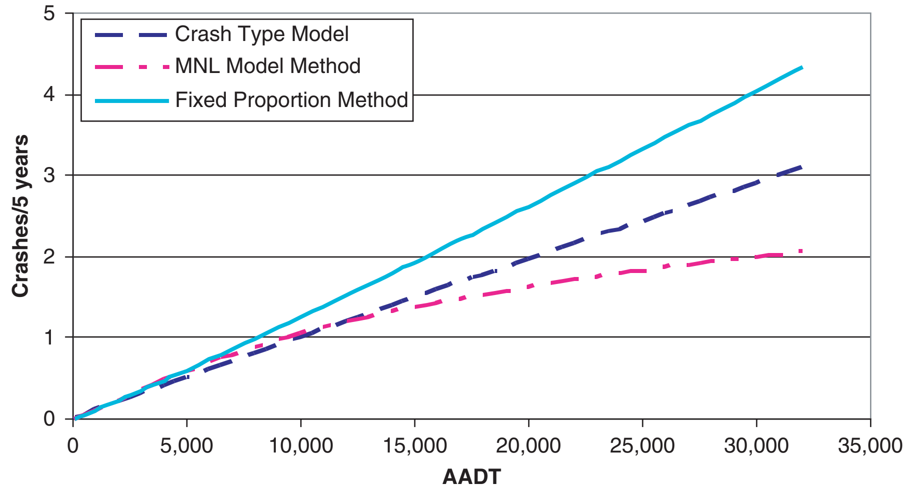


FIGURE 2 Predicted number of head-on crashes as a function of AADT.

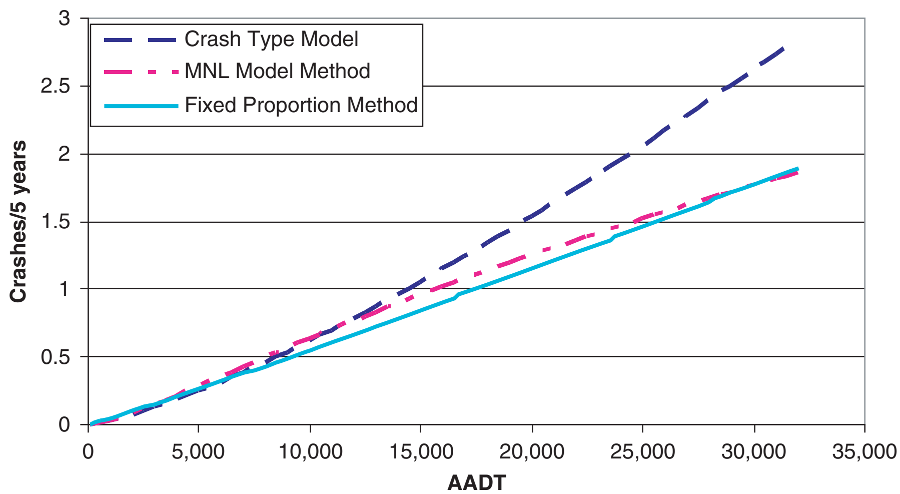


FIGURE 3 Predicted number of sideswipe–opposite direction crashes as function of AADT.

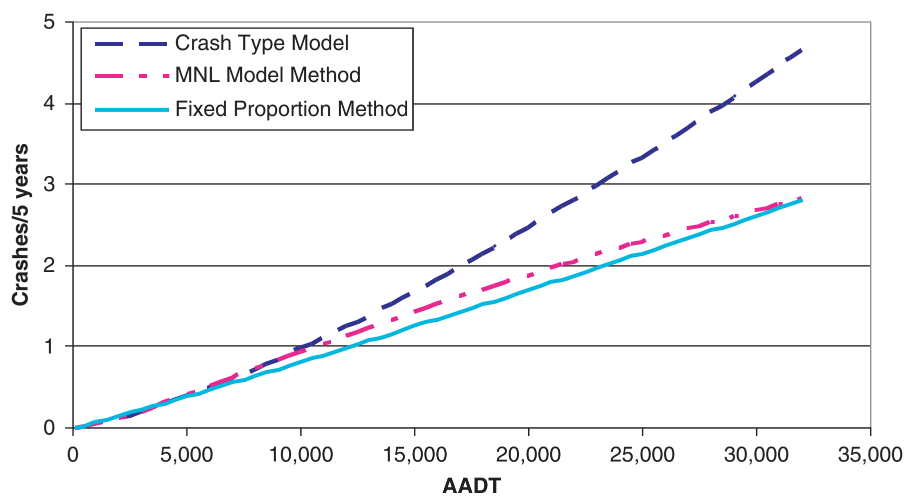


FIGURE 4 Predicted number of sideswipe–passing crashes as function of AADT.

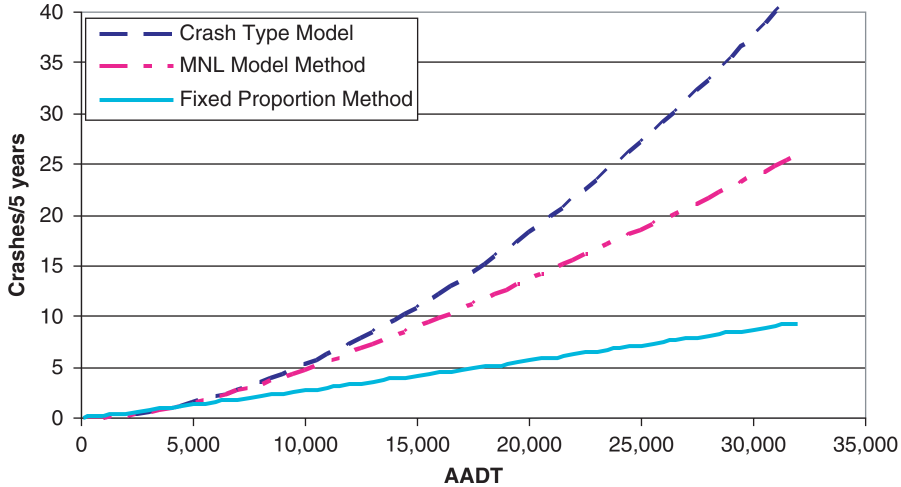


FIGURE 5 Predicted number of rear-end crashes as function of AADT.

to estimate crash counts by collision type. Crash data collected from rural, two-lane highway segments in Minnesota for the years 2002 through 2006 were used to compare this approach with the two previous approaches documented in the literature: crash type models and collision types estimated by using fixed proportions. Crashes that occurred on these segments were divided into five categories: head-on, rear-end, passing-direction sideswipe, opposite-direction sideswipe, and single-vehicle.

The application of the MNL model to estimate the proportion of crashes by collision type seems 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 GOF 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. The prediction of crash counts by specific crash type models was nonetheless found to be the best method, as documented in Jonsson et al. (*9*).

Development of models for collision types can be negatively influenced, however, by the small sample size and low sample-mean problem (*11*). The use of a logit model (such as an MNL model) to estimate 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. The study used crash data collected on rural, two-lane highways. Analysis could be extended to data on multilane, high-speed highways and intersections to see if the findings would be similar.
2. The study could be broadened to include more collision types.
3. Since the MNL model suffers from methodological limitations, the application of mixed logit models might be evaluated to estimate collision patterns. Mixed logit models relax the assumptions of IIAs and allow for heterogeneity from a variety of sources (*27*).

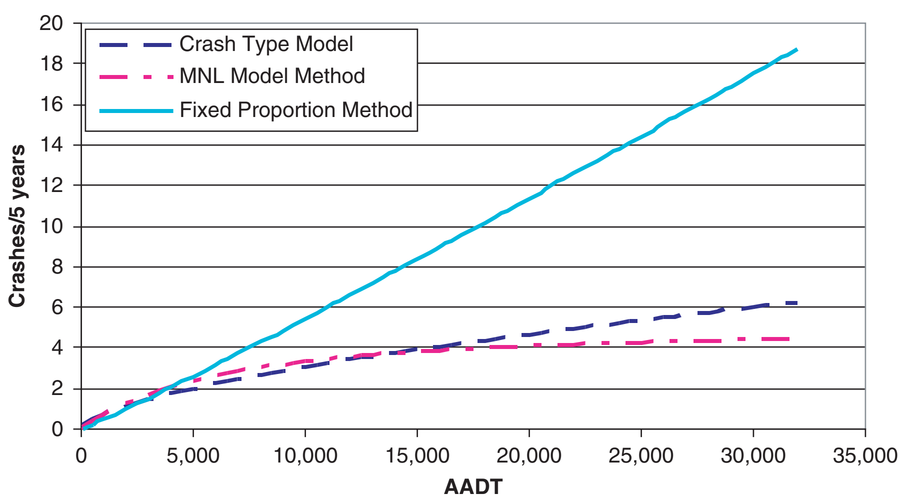


FIGURE 6 Predicted number of single-vehicle crashes as function of AADT.

## REFERENCES


1. Lord, D., S. R. Geedipally, B. N. Persaud, S. P. Washington, I. van Schalkwyk, J. N. Ivan, C. Lyon, and T. Jonsson. *NCHRP Web-Only Document 126: Methodology to Predict the Safety Performance of Rural Multilane Highways*. Transportation Research Board of the National Academies, Washington, D.C., 2008.
2. Ivan, J. N., C. Wang, and N. R. Bernardo. Explaining Two-Lane Highway Crash Rates Using Land Use and Hourly Exposure. *Accident Analysis and Prevention,* Vol. 32, No. 6, 2000, pp. 787–795.
3. Lyon, C., J. Oh, B. Persaud, S. Washington, and J. Bared. Empirical Investigation of Interactive Highway Safety Design Model Accident Prediction Algorithm: Rural Intersections. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1840,* Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 78–86.
4. Tarko, A. P., M. Inerowicz, J. Ramos, and W. Li. Tool with Road-Level Crash Prediction for Transportation Safety Planning. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2083,* Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 16–25.
5. Hauer, E., J. C. N. Ng, and J. Lovell. Estimation of Safety at Signalized Intersections. In *Transportation Research Record 1185,* TRB, National Research Council, Washington, D.C., 1988, pp. 48–61.
6. Shankar, V., F. Mannering, and W. Barfield. Effect of Roadway Geometric and Environmental Factors on Rural Freeway Accident Frequencies. *Accident Analysis and Prevention,* Vol. 27, No. 3, 1995, pp. 371–389.
7. Geedipally, S. R., and D. Lord. Investigating the Effect of Modeling Single-Vehicle and Multi-Vehicle Crashes Separately on Confidence Intervals of Poisson-Gamma Models. *Accident Analysis and Prevention.* Vol. 42, No. 4, 2010, pp. 1273–1282.
8. Abdel-Aty, M., J. Keller, and P. A. Brady. Analysis of Types of Crashes at Signalized Intersections by Using Complete Crash Data and Tree-Based Regression. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1908,* Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 37–45.
9. Jonsson, T., C. Lyon, J. N. Ivan, S. P. Washington, I. van Schalkwyk, and D. Lord. Differences in the Performance of Safety Performance Functions Estimated for Total Crash Count and for Crash Count by Crash Type. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2102*, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 115–123.
10. Kim, D., J. Oh, and S. Washington. Modeling Crash Outcomes: New Insights into the Effects of Covariates on Crashes at Rural Intersections. *ASCE Journal of Transportation Engineering,* Vol. 132, No. 4, 2006, pp. 282–292.
11. Lord, D. Modeling Motor Vehicle Crashes Using Poisson-Gamma Models: Examining the Effects of Low Sample Mean Values and Small Sample Size on the Estimation of the Fixed Dispersion Parameter. *Accident Analysis and Prevention,* Vol. 38, No. 4, 2006, pp. 751–766.
12. Lord, D., and F. Mannering. The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives. *Transportation Research Part A: Policy and Practice,* Vol. 44. No. 5, pp. 291–305.
13. Lord, D., S. P. Washington, and J. N. Ivan. Poisson, Poisson-Gamma and Zero Inflated Regression Models of Motor Vehicle Crashes: Balancing Statistical Fit and Theory. *Accident Analysis and Prevention,* Vol. 37, No. 1, 2005, pp. 35–46.
14. Shankar, V., and F. Mannering. An Exploratory Multinomial Logit Analysis of Single-Vehicle Motorcycle Accident Severity. *Journal of Safety Research,* Vol. 27, No. 3, 1996, pp. 183–194.
15. Carson, J., and F. Mannering. The Effect of Ice Warning Signs on Accident Frequencies and Severities. *Accident Analysis and Prevention,* Vol. 33, No. 1, 2001, pp. 99–109.
16. Abdel-Aty, M. Analysis of Driver Injury Severity Levels at Multiple Locations Using Ordered Probit Models. *Journal of Safety Research,* Vol. 34, No. 5, 2003, pp. 597–603.
17. Kockelman, K., and Y. J. Kweon. Driver Injury Severity: An Application of Ordered Probit Models. *Accident Analysis and Prevention,* Vol. 34, No. 4, 2002, pp. 313–321.
18. Qin, X., J. N. Ivan, and N. Ravishanker. Selecting Exposure Measures in Crash Rate Prediction for Two-Lane Highway Segments. *Accident Analysis and Prevention,* Vol. 36, No. 2, 2004, pp. 183–191.
19. Lord, D., S. D. Guikema, and S. Geedipally. Application of the Conway-Maxwell-Poisson Generalized Linear Model for Analyzing Motor Vehicle Crashes. *Accident Analysis and Prevention,* Vol. 40, No. 3, 2008, pp. 1123–1134.
20. Jonsson, T., J. N. Ivan, and C. Zhang. Crash Prediction Models for Intersections on Rural Multilane Highways: Differences by Collision Type. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2019,* Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 91–98.
21. McFadden, D. Econometric Models of Probabilistic Choice. In *Structural Analysis of Discrete Data with Econometric Applications* (Manski and D. McFadden, eds.), MIT Press, Cambridge, Mass., 1981.
22. Koppelman, F. S., and C. R. Bhat. A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models. Prepared for the Federal Transit Administration, U.S. Department of Transportation, 2006, pp. 79–80.
23. Oh, J., C. Lyon, S. Washington, B. Persaud, and J. Bared. Validation of FHWA Crash Models for Rural Intersections: Lessons Learned. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1840,* Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 41–49.
24. Greene, W. H. *LIMDEP, Version 9.0: User's Manual.* Econometric Software, Plainview, N.Y., 2007.
25. SAS Institute Inc. *Administrator Guide for the SAS System Version 9 for 64-bit Microsoft Windows.* Cary, N.C., 2002.
26. Highway Safety Information System. FHWA, U.S. Department of Transportation. http://www.hsisinfo.org/.
27. Train, K. E., *Discrete Choice Methods with Simulation.* Cambridge University Press, United Kingdom, 2003.


*The Statistical Methods Committee peer-reviewed this paper.*