Arab Journal of
Administrative Sciences

2005, Vol. 12, No. 3, 473 - 499

TRAFFIC COLLISION ANALYSIS MODELS: REVIEW AND EMPIRICAL EVALUATION

*Said M. Easa,**
*Mohamad K. Hasan,***
*Mohammad M. Hamed****

* Ryerson University, Canada
** Kuwait Universuty, Kuwait
*** Jordan University of Science and Technology, Jordan

Key Words

Traffic Collision Frequency; Exponential Models; Poisson Model; Negative Binomial Model; Multiple Linear Models; Survival Models Logit Model; Qualitative Data Analysis Models; Time Series Models.

Abstract

Collision analysis models were classified in this paper into two categories: aggregate and disaggregate. A theoretical review of these models was presented, followed by an empirical (comparative) evaluation of multiple linear models, Poisson models, negative binomial models, and exponential models using three functional forms (additive, multiplicative, and quadratic). The evaluation results showed that the functional form of the model is very important for prediction accuracy. The Poisson model with a quadratic form provided the best accuracy, while multiple linear and expential models provided the least accuracy.

Introduction

Predicting highway traffic-collision frequency is one of the most critical issues for city planning decision makers and auto insurance companies. The primary function of any highway traffic department is to ensure safety of highway users and minimize the collision likelihoods through highway geometric improvements and traffic regulation enhancement. Therefore, developing an adequate collision frequency model that can relate collision frequency to highway geometrics and traffic factors will be very useful in determining the type and magnitude of improvements that should be conducted. These models can help decision makers in any city planning department to evaluate different alternatives (scenarios) by measuring the level of safety that can be

Submitted April 2003, accepted April 2004.

maintained with respect to the cost of improvements.

Most auto insurance companies consider a specific risk classification system in their risk assessment and pricing policy. This classification system can include many factors, like driver characteristics (e.g., age, gender, marital status, and previous driving experience) and driving area characteristics (e.g., home location, work location, trip frequency, and routes) (Hamdy et al. 1998). In this case, an adequate collision frequency model can be very helpful in the auto insurance risk assessment. This paper attempts to provide guidance, through theoretical and empirical evaluations, regarding the selection of an adequate model among the many models that have been presented in the literature.

Highway traffic collisions are complex events that involve the interaction among many factors related to roadway geometric design (e.g. horizontal and vertical alignments), the driver (e.g., age and sex), the vehicle, the traffic (e.g. vehicle speed and congestion level), and the environment (e.g. weather and lighting condition). In this paper we are concerned with collision models that address the relationship between collision frequency and roadway geometric design and traffic factors.

Traffic-collision models attempt to address the following safety questions for a given highway section with specified traffic and geometric characteristics (Miaou and Lum 1993): what vehicle collision involvement rate and collision probability can be expected? which traffic or geometric design variables are more critical to road safety performance? and what percentage of collision reduction can be expected from various improvements? The authors examined four types of models: a linear model with additive and multiplicative forms and two Poisson models. The models were examined with regard to distribution assumptions, choice of geometric design and confounding variables, choice of functional form, delineation of road sections, treatment of vehicle exposure and traffic conditions, and sampling and non-sampling errors. It was found that linear models lacked the distribution property to adequately describe vehicle collisions and did not closely fit the observed data. As such, test statistics from these models would be questionable. On the other hand, Poisson models were found to possess the desirable statistical properties in describing vehicle collision data, but one limitation of these models is that the variance of collision data is constrained to be equal to the mean. It was suggested that more general probability distributions, like negative binomial, should be considered in collision analysis.

The objective of this paper is two-fold: (a) to conduct a literature review of various collision models and (b) to evaluate empirically a comprehensive set of models (multiple linear, negative binomial, Poisson, and exponential) with three functional forms (additive, multiplicative, and quadratic). The review not only presents a classification of existing collision models that would be of interest to the readers, it also serves as a useful background for the empirical evaluation. Existing collision models are classified into two categories: aggregate and disaggregate. Aggregate models include multiple linear models, exponential family, time series, qualitative data analysis models, and generalized linear models. Disaggregate models include survival and logit models.

Aggregate Models

Multiple Linear Models

Consider a highway that is divided into n homogeneous segments. For each segment i, the number of collisions in a year Y_i (dependent variable) is a function of traffic and geometric characteristics (explanatory variables) for segment i, $\mathbf{x}_i$, where $\mathbf{x'}_i = (x_{i1}, x_{i2}, ..., x_{ik})$ with $x_{i1} = 1$ and k = number of explanatory variables. The most common multiple linear models are the additive, multiplicative, and quadratic models (Sergio and Sergio 1986).

The additive linear model is given by

$$Y_i = \mathbf{x'}_i\beta + \varepsilon_i = \sum_{j=1}^{k} x_{ij}\beta_j + \varepsilon_i, \qquad (1)$$

$$\varepsilon_i \sim \text{ind N}(0, \sigma^2), \quad i = 1,2, ...,n \text{ (Additive)}$$

where β is a $k \times 1$ vector of unknown regression coefficients, $\varepsilon_i, i = 1, 2, ..., n$, are error terms that are independently and normally distributed with zero mean and constant variance $\sigma^2 > 0$. The normality assumption is used primarily to obtain tests and confidence statements about the estimated regression coefficients. It was seldom to make probabilistic statements about Y_i. The model can also be written as

$$Y_i \sim \text{ind N } (\mu_i, \sigma^2) \text{ where}$$

$$\mu_i = E(Y_i) = \mathbf{x'}_i\beta \text{ , } i = 1, 2, ..., n \qquad (2)$$

The least squares estimates of the coefficients $\beta_1, \beta_2, ..., \beta_k$ in (1) (they are also the maximum likelihood estimates (MLE) under the normal assumption) can be obtained by any standard linear regression computer program.

The multiplicative linear model is given by

$$Y_i + \delta = \beta_1 (\prod_{j=2}^{k} (1 + x_{ij})^{\beta_j}) e^{\varepsilon_i}$$

(Multiplicative) (3)

where ε_i is as defined in (1). Taking the logarithm of both sides of (3), then

$$\ln(Y_i + \delta) = \beta_1^* + \sum_{j=2}^{k} \beta_j \ln(1 + x_{ij}) + \varepsilon_i \quad (4)$$

where δ is a selected small constant, that is added to Y_i to avoid the ln(0) problem, and $\beta_1^* = \ln(\beta_1)$.This model requires that $x_{ij} \geq 0$ for all i and j. One has been added to each independent variable x_{ij} (exceptx_{i1}) to avoid the ln(0) problem. The model can also be expressed as

$$\ln(Y_i + \delta) \sim \text{ind N}(\nu_i, \sigma^2)$$

where $\nu_i = E(\ln(Y_i + \delta)) = \beta_1^* + \sum_{j=2}^{k} \beta_j \ln(1 + x_{ij})$,
$i = 1, 2, ..., n$ (5)

As in the additive model of (1), the lognormal distributional assumption is used primarily to obtain tests and confidence statement about the estimated regression coefficients. It is not intended to make probabilistic statements about Y_i. The expected value of Y_i under this model is

$$\mu_i = E(Y_i) = -\delta + \beta_1 (\prod_{j=2}^{k} (1 + x_{ij})^{\beta_j}) e^{\frac{1}{2}\sigma^2},$$
$i = 1, 2, ..., n$ (6)

The quadratic linear model has been proposed by Sergio and Sergio (1986) who studied the effect of traffic flow and geometric characteristics on collision frequency for Chilean roads. The study compared this model with two linear specification functions explaining the total collisions, collisions involving pedestrians (CIP), and the total collisions minus CIP. The results showed that the second order effects, especially for geometric variables, are highly significant and their omission caused a relevant downward bias on the marginal effect of traffic flow on collisions.

The quadratic specification form represents a second order Taylor expansion around a point of approximation chosen as the mean of the explanatory variables. If $Y_i = f(\mathbf{x}_i)$ then the second order Taylor expansion of Y_i around the point of approximation $\bar{\mathbf{x}}_i = (\bar{x}_{i1}, \bar{x}_{i2}, ..., \bar{x}_{ik})$ is given by,

$$Y_i \cong f(\bar{\mathbf{x}}_i) + \sum_{j=2}^{k} \frac{\partial f(\bar{\mathbf{x}}_i)}{\partial x_{ij}} (x_{ij} - \bar{x}_{ij})$$
$$+ \frac{1}{2} \sum_{j=2}^{k} \sum_{l=2}^{k} \frac{\partial^2 f(\bar{\mathbf{x}}_i)}{\partial x_{ij} \partial x_{il}} (x_{ij} - \bar{x}_{ij})(x_{il} - \bar{x}_{il}) \quad (7)$$

Hence the quadratic model for estimating an unknown function that relates Y_i to $\mathbf{x}_i$is

$$Y_i = \beta_1 + \sum_{j=2}^{k} \beta_j (x_{ij} - \bar{x}_{ij})$$
$$+ \frac{1}{2} \sum_{j=2}^{k} \sum_{l=2}^{k} \beta_{jl} (x_{ij} - \bar{x}_{ij})(x_{il} - \bar{x}_{il}) + \varepsilon_i \quad (8)$$

$\varepsilon_i \sim ind\ N(0, \sigma^2), = 1, 2, ..., n$ (quadratic)

The model can also be written as

$Y_i \sim \text{ind N}(\mu_i, \sigma^2)$

and

$$\mu_i = E(Y_i) = \beta_1 + \sum_{j=2}^{k} \beta_j (x_{ij} - \bar{x}_{ij})$$

$$+\frac{1}{2}\sum_{j=2}^{k}\sum_{l=2}^{k} \beta_{jl}(x_{ij} - \bar{x}_{ij})(x_{il} - \bar{x}_{il}), \quad i = 1, 2, ..., n \quad (9)$$

where β_1, β_j and β_{jl} are coefficients to be estimated that represent the value of the function, the gradient, and the Hessian, respectively. The quadratic around the mean specification is flexible in the sense that it makes no a priori restrictions in terms of first and second derivatives. It provides unrestricted estimates of both the gradient and the Hessian of the unknown underlying function.

Exponential Family Models

Poisson Model

The Poisson model is widely used in the analysis of count data because the model has discrete, nonnegative integer-distribution characteristics (Frome et al. 1973, Holford 1983, Frome 1983, Cameron and Trivedi 1986). It can also be generalized to more flexible model forms. The Poisson model states that the probability that segment i has y_i collisions is given by

$$P(Y_i = y_i / \mathbf{x}_i) = \frac{\exp(-\mu_i)\mu_i^{y_i}}{y_i!}, y_i = 0, 1, 2, ..., n \quad (10)$$

where μ_i is the Poisson parameter of segment i (also the mean of the Poisson distribution)

$$\mu_i = \text{E}(Y_i / \mathbf{x}_i)$$

which will be some estimable function of the independent variables $\mathbf{x}_i$. That is,

$$\mu_i = \exp(\mathbf{x'}_i \beta) \quad (11)$$

where β is a vector of parameters that can be estimated through standard maximum likelihood procedure (Greene 1993).

From equations (10) and (11) the likelihood function is

$$L(\beta) = \prod_i \frac{\exp[-\exp(\mathbf{x}_i\beta)][exp(\mathbf{x'}_i\beta)]^{y_i}}{y_i!}$$

$$i = 1, 2, \ldots, n \quad (12)$$

which gives a log-likelihood function

$$l(\beta) = logL(\beta) = \sum_i [-\log y_i! - \exp(\mathbf{x'}_i\beta) + y_i \mathbf{x'}_i \beta \quad (13)$$

This function can be maximized to estimate β, thereby providing an estimate of the mean number of collisions for segment i. Unlike standard least-square regression, μ_i is a deterministic function of $\mathbf{x}_i$ with randomness coming from the probability specification of y_i.

Negative Binomial Model

The Poisson model constrains the mean and variance to be equal, that is,

$$E(Y_i/\mathbf{x}_i) = \mathrm{Var}(Y_i/\mathbf{x}_i) \tag{14}$$

This equidispersion restriction is not always compatible with the data. Some heterogeneity may not be captured by the regression component $(\mathbf{x}'_i\beta)$. That is, if the data are overdispersed or underdispersed, this will result in biased estimates of β. Previous research showed that collision frequency data tend to be overdispersed, with the variance being significantly greater than the mean.

To overcome this restriction, Gouieroux et al. (1984) suggested extending the Poisson model by adding a random term ε_i in the regression component to account for the omission of unobserved exogenous variables or other random effects. That is,

$$\mu_i = \exp(\mathbf{x}'_i\beta + \varepsilon_i) \tag{15}$$

If we assume that the random variable $\exp(\varepsilon_i)$ is gamma distributed with a mean equal to 1 and variance $\alpha > 0$, then the Poisson model of (10) becomes

$$P(Y_i = y_i/\mathbf{x}_i) = \frac{\Gamma(y_i+\frac{1}{\alpha})[\alpha\exp(\mathbf{x}'_i\beta)]^{y_i}}{\Gamma(\frac{1}{\alpha})y_i![1+\alpha\exp(\mathbf{x}'_i\beta)]^{y_i+\frac{1}{\alpha}}},$$

$$y_i = 0, 1, 2, \ldots, n \tag{16}$$

where $\Gamma(.)$ is the gamma function, which defines a negative binomial model with a variance-mean relationship given by

$$\mathrm{Var}(Y_i/\mathbf{x}_i) = E(Y_i/\mathbf{x}_i)[1 + \alpha\, E(Y_i/\mathbf{x}_i) \tag{17}$$

If α is significantly different from zero, the data are overdispersed or underdispersed.

If $\alpha = 0$, then the negative binomial reduces to the Poisson model. The maximum likelihood procedure can be used to estimate β andα.

Lawless (1987) showed that the log of the maximum likelihood function of the negative binomial can be written as

$$l(\beta,\alpha) = \sum_{i=1}^{n}\{\sum_{\omega=1}^{y_i-1} Log(1+\alpha\omega) + y_i\mathbf{x}'_i\beta$$

$$-(y_i+\frac{1}{\alpha})Log[1+\alpha\exp(\mathbf{x}'_i\beta)]\}. \tag{18}$$

Although the negative binomial model is more general than the Poisson model, the former requires more extensive computation to estimate model coefficients and to generate inferential statistics. Furthermore, the statistical properties of different estimators, such as the MLE and moment estimators, of the negative binomial model under different sample sizes have not yet been fully investigated. Many statistical studies are underway to modify existing distribution functions within the exponential family to allow more flexible variance-mean relationships (Efron 1986, Gelfand and Dalal 1990).

Exponential Model

The exponential model can be written as

$$Y_i + \delta = \exp(\sum_{j=1}^{k} x_{ij}\beta_j + \varepsilon_i) \quad (19)$$

$$\varepsilon_i \sim indN(0,\sigma^2), i = 1, 2, ..., n$$

$$\text{or } \ln(Y_i + \delta) = \sum_{j=1}^{k} x_{ij}\beta_j + \varepsilon_i \quad (20)$$

The model can also be expressed as $\ln(Y_i + \delta) \sim \text{ind N}(\mu_i, \sigma^2)$where

$$\mu_i = E(Y_i) = \exp(\sum_{j=1}^{k} x_{ij}\beta_j), i = 1, 2, ..., n \quad (21)$$

Time-Series Models

Time-Based Model

In this approach, a time series of monthly (or yearly) total number of collisions in a specific area during a given time period can be related to a number of explanatory variables (Scott 1986)

$$Y_t = f(\mathbf{x}_t) + e_t, t = 1, 2, ..., T \quad (22)$$

where $Y_t =$ number of collisions in a specific area at time t, T =time period of study, $\mathbf{x}_t$ =vector of explanatory variables that contribute to collision frequency at time t, f = function that relates collision frequency Y_t to the explanatory variables $\mathbf{x}_t$, and e_t =residual error term representing the fraction of Y_t that is not explained by the functional form of $\mathbf{x}_t$ at time t.

The explanatory variables may include variables whose values can be changed during the period of study and that may have a direct or indirect effect on collision frequency (acting via some induced changes in characteristics of traffic or road environment). Numerous economic, policy, and weather indicators can also be used. Note that this model does not relate collisions to locations in which the geometric design effects can be measured, and traffic volume is an average for the area of study. Many time series techniques, like Box-Jenkins ARIMA, can be used to model the time series model of (22).

ARIMA models are flexible and widely used in time-series analysis. The ARIMA procedure combines three types of processes: autoregressive (AR), differencing to strip off the integration (I) of the series, and the moving average (MA). All three types are based on the simple concept of random disturbances or shocks. A disturbance between two observations in the series occurs, which somehow affects the level of the series. These disturbances can be mathematically described by the ARIMA model. Each of the three types has its own characteristic way of responding to a random disturbance. One of the basic principles in the ARIMA procedure is that the series should first be made as stationary as possible by means of differencing which reduces the series to a form in which the trend and seasonality are not clearly present by visual inspection. The variability of the residuals should also appear to be fairly consistent (white noise) throughout the series.

Scott (1986) applied ARIMA technique to road collision data for British highways. The generally very "noisy" autocorrelation among the residuals from standard regression is not very strong; time-series models are unlikely to represent collision data appreciably better than the regression based on the assumption of autocorrelated residuals.

Space-Based Model

The space-based model considers the highway as a system of connected segments where each segment represents a space increment instead of a time increment (Hasan et al. 1998). The model takes the following form,

$$Y_s = \beta \mathbf{x}_s + e_s \qquad (23)$$

where Y_s = collision frequency on segment s, $\mathbf{x}_s$ = vector of traffic and geometric variables that contribute to collision frequency, β = vector of parameters to be estimated, and e_s = residual error term representing a fraction of Y_s that is not explained by the linear combination of the variables $\mathbf{x}_s$ [e_s = ARIMA(ε_s), where ε_s is a white noise process]. The space-based model was able to measure the effect of gradual and sudden changes in traffic and geometric characteristics from one segment to another. As such, the model would be useful in assessing highway geometric design consistency. The model was shown to fit the observed collision data much better than the linear ordinary regression models.

Qualitative Data Analysis Models

These models include principal component analysis (PCA) and canonical correlation analysis (CCA) (Oppe 1992). PCA considers the relational structure of a group of variables. It results in a linear combination of the variables, which is the best representation of all variables in the group. This combination is the first component. Subsequently, it finds the second component, being the best representation for the residuals of all variables, given the first component. If there are M variables, then at least M components describe the variables perfectly. If some variables are linearly dependent, less than M components will give a perfect description. PCA can be regarded as a technique that finds the best representation of a matrix of N observation (rows) and M variables (columns), by a matrix of smaller rank, say N rows and P columns, where P is considerably smaller than M. Classical PCA is used to obtain insight into the interrelations of large numbers of metric variables.

For a set of specific events, such as collisions, CCA calculates linear combinations of variables in a particular group of characteristics of these events that maximizes correlation with linear combinations of variables in a second group. Such a correlation is called canonical correlation coefficient. With only one dependent variable in the second set, CCA is similar to multiple linear regression analysis. A two-dimensional solution is given by two planes for the best representations of the predictor variables and the dependent variables. The smallest angle between these planes corresponds to the first correlation coefficient, while perpendicular to this we find the angle corresponding to the second correlation coefficient. The smaller the angle, the better the prediction is. All vectors corresponding to the predictor and dependent variables will be projected on this plane. If there were only two dependent variables, the plane would contain both corresponding vectors. With one dependent variable, there would be only a straight line through the corresponding vector.

Generalized Linear Interactive Models

The generalized linear interactive models (GLIM) have several important aspects in modeling the data. First, a linear model is assumed for the relationship between the predictor variables and the so-called linear model predictor which need not predict the dependent variable directly. Second, one can select a link function (identity, log, and so on) to define the relationship between the expected value of the dependent variable and the linear model predictor. Finally, one can choose the error distribution for the observations on the dependent variable (normal, Poisson, and so on).

The linear and generalized linear models most frequently used in traffic collision analysis are further classified by Oppe (1992) into two groups: object-oriented and design-oriented techniques. Object-oriented techniques have a particular object as the unit of analysis, such as a collision, a collision location, or a road user (normally use regression analysis). Design-oriented techniques do not work with the object themselves, but with the characteristics of objects, such as the class of young road users driving at night or during daytime, or collision classifications according to age and sex. These techniques, which are generally applied in experimental design, include analysis of variance and log-linear analysis of contingency tables.

Disaggregate Models

Survival Models

Survival analysis has been used to model traffic collisions, and several studies have utilized the concept of the survivor function to model traffic collisions. Survival analysis provides researchers with tools to investigate the relationship between traffic collision occurrence (at the driver level) and the corresponding exposure data. This permits consideration of more policy-oriented variables that could not be handled by aggregate data (Chang and Jovanis 1990). Hensher and Mannering (1994) provide a full review of the application of the survival theory in transportation. Basically, the survivor function is defined as,

$$S(t) = 1 - F_T(t) \qquad (24)$$

where $F_T(t)$ = cumulative distribution function of the random variable T (e.g. time between traffic collisions). The probability density function of the random variable T can be written as,

$$f_T(t) = d\, F_T(t)/dt \qquad (25)$$

The hazard (h(t)), survival, and the probability density functions are related by,

$$h(t) = f(t) / S(t) \qquad (26)$$

The hazard function can be thought of as the rate at which traffic collisions occur (Kalbfliesch and Prentice 1980). The slope of the hazard function determines whether the hazard is monotonically increasing or decreasing in duration, constant, or nonomonotonic (Kalbfliesch and Prentice 1980). Mannering and Hamed (1990) developed a set of survival models with the aim of modeling the duration of commuters' delay following their departure from work. The work by Hamed and Mannering (1993) is considered a cornerstone in the development of duration models to model a number of commuters' post-work activities. These included duration at activity (shopping, social, etc) prior to home arrival from work, length of home-stay duration, and time taken to end an activity and pursue another one. Hamed et al. (1997) estimated survival models to quantify the time between traffic collisions for taxicab drivers. The study estimated a number of duration models and the duration to the first, second, third, and fourth traffic collisions. To include the effect of covariates (explanatory variables) on the hazard function, the hazard function can be written as,

$$h(t, X) = h_0(t) \exp(\gamma X) \qquad (27)$$

where h(t, X) = hazard function with covariate vector X, $h_0(t)$ = unknown hazard function with covariate vector X = 0, and γ = vector of estimable parameters. The above model is referred to as the proportional hazard model (Cox 1972, Kalbfliesch and Prentice 1981, Fleming and Harington 1990). In this model, the vector X acts multiplicatively on the hazard function.

The proportional hazard model can be estimated without making any assumptions regarding the probability distribution of the variable related to the time between collisions. Hamed et al. (1997) specified and estimated a set of proportional hazard models to determine the factors influencing the time between traffic collisions in the taxi cab sector. Other studies by Chang and Jovanis (1990) estimated a set of proportional hazard models to study truck collisions. In addition to the pooled data, they estimated proportional hazard models for collision severity (property damage only, injury, and fatality), and collision type (single and multiple-vehicle collisions).

Alternatively, one could select an appropriate probability distribution for the time variable. Commonly-used distributions include the Weibull, log-normal, log-logistic, and the exponential. For example, in the case of the Weibull distribution, the hazard function with covariate vector X can be written as,

$$h(t\,, X) = \alpha\, p(\alpha p)^{p-1} \exp(-\gamma X) \quad (28)$$

where α and p = Weibull distribution shape and scale parameters, respectively. Mannering (1993) applied this formulation to investigate traffic collision involvement and duration for both male and female drivers. The study provided much insight into the factors influencing each duration. Other studies also include Jones et al. (1991).To estimate the vector γ, the proportional hazard approach by Cox (1972) is usually used.

Logit Models

The logit model has been used extensively in the transportation field as a discrete choice model. In the context of disaggregate modeling of traffic collisions, Hamed and Easa (1998) and Mannering and Winston (1987) explicitly addressed the drivers' beliefs in the effectiveness of seat belts. The beliefs variable was a binary variable taking the value of one if the driver believed that wearing a seat belt increases the likelihood of surviving a traffic collision and zero otherwise. To estimate such a binary model, the logit model was specified and estimated. Formally, the likelihood function of this binary logit model can be stated as,

$$l = \prod_{i=1}^{M} [\Pr ob_i]^{q_i} [1 - \mathrm{Prob}_i]^{1-q_i} \quad (31)$$

where Prob_i = probability of believing in the effectiveness of the seat belt, q_i = value of the binary belief variable (0 or 1), and M = number of drivers in the random sample. To address the influence of a number of external factors on the belief probabilities, the Prob_i variable was assumed to be a function of the explanatory variables. The likelihood and log-likelihood functions were stated as,

$$l = \prod_{i=1}^{M} [F(X_i'\beta)]^{q_i} [1 - F(X_i'\beta)]^{1-q_i} \quad (32)$$

$$\ln(l) = \sum_{i=1}^{M} \{q_i \ln[F(X_i'\beta)] + (1 - q_i)\ln[1 - F(X_i'\beta)]\} \quad (33)$$

where $Prob_i = F(X_i'\beta)$ = logistic cumulative distribution function, X_i = vector of external factors for driver i, and β = vector of estimable parameters. The maximum likelihood method was used to estimate the parameters of the above binary logit model.

Empirical Evaluation

The relative prediction ability of different types of the aggregate collision models and functional forms were evaluated using empirical data. This section presents details on the data collection and preliminary analysis, evaluated models, and evaluation results.

Data Collection and Preliminary Analysis

Collision data for the TOME Expressway in Japan were used in this analysis. The expressway has two lanes in each direction and begins just outside of Tokyo outer ring road. The study section extends approximately from kilopost 50 to 280, with a total length of 230 km. Most of the study section is located in rural areas. The expressway is heavily used by trucks that account for about 20% of the total traffic volume. The collision, traffic, and geometric characteristics of the study expressway were generated from a comprehensive database for major Japanese expressways. The database includes information about each collision, such as kilopost, collision type, vehicle type, horizontal alignment, vertical alignment, and environmental conditions.

For purposes of this study, the expressway was divided into segments (5 km long), resulting in 46 segments of the study section. A computer program was written to allow the extraction of specific information related to each segment, including number of collisions, passenger and truck volumes, land use, and geometric design (design speed, horizontal alignment, and vertical alignment). Details on the basic fundamentals of geometric design can be found elsewhere (Easa 2002). The following variables describe the traffic and geometric characteristics for segment *i* used in this study:

$Acc(Y_i)$ = total number of collisions/year

x_{i1} = dummy intercept ($x_{i1} = 1$)

$Hor(x_{i2})$ = sum of the deflection angles of horizontal curves per kilometer (in degrees)

$Ver(x_{i3})$ = sum of the absolute values of the algebraic difference in grades per kilometer (in percent)

$Pas(x_{i4})$ = passenger car volume (average daily traffic in 1,000 vehicles)

$Truck(x_{i5})$ = truck volume (average daily traffic in 1,000 vehicles)

$NH(x_{i6})$ = number of changes in horizontal alignment per kilometer

$NV(x_{i7})$ = number of changes in vertical alignment per kilometer

$SP(x_{i8})$ = design speed (km/h)

The data ranges for the preceding variables are shown in Table 1. The number of collisions per segment ranges from 3 to 31 (total observed collisions of 489). The scatter diagram of the observed collision frequency is shown in Figure 1. The diagram shows that the minimum value of collision frequency is 3 collisions in segment number 38 and the maximum value is 31 collisions in segments number 3 and 4. The average and standard deviation of the collision frequency are 10.63 and 6.55, respectively.

The traffic and geometric characteristics of these specific segments are shown in Table 2. The results show

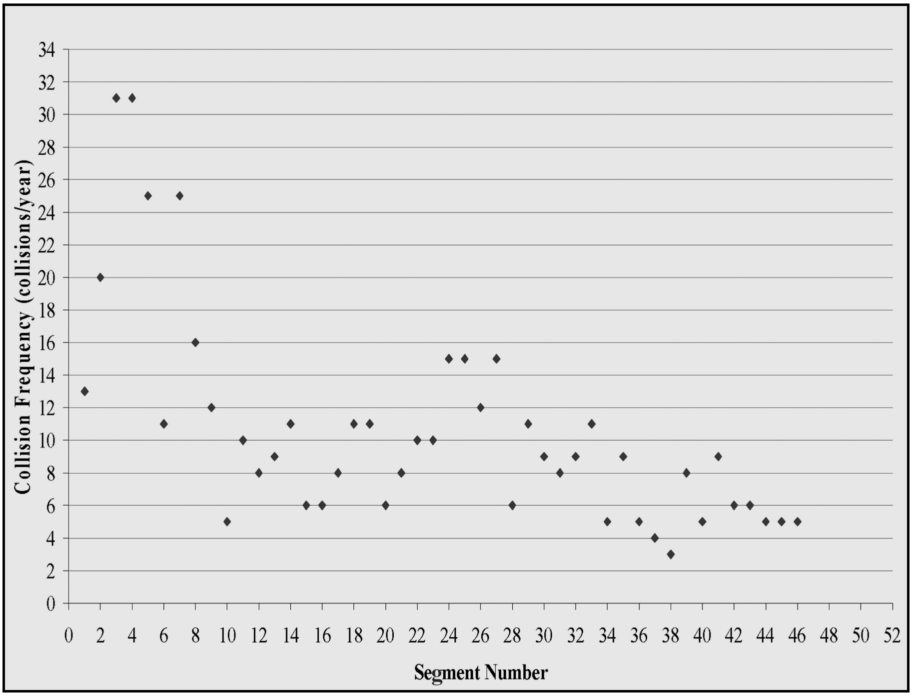
Table 1
Ranges of input data for different variables

Variable	Unit	Range	
		Lower	Upper
Acc	Collisions/year	3	31
Hor	Degrees/km	0.00	56.95
Ver	Percent/km	0.10	3.84
Pas	1,000 vehicles	41	55
Truck	1,000 vehicles	11	16
Nh	Number/km	0	2.2
Nv	Number/km	0.2	1.8
Sp	Km/h	80	100

Table 2
Minimum and maximum observed collision frequencies based on traffic and geometric characteristics of highway segments

Segment Number	Traffic and Geometric Characteristics							
	Acc	Hor	Ver	Pas	Truck	NH	NV	SP
38	3	8.04	0.7	51	11	0.4	0.4	100
3	31	31.22	5.2	55	16	2.2	2	80
4	31	26.73	1.22	55	16	1.6	0.8	80

Figure 1
Scatter diagram of observed collision frequency



that there is a significant difference between traffic and geometric characteristics of segment 38, and between the characteristics of segments 3 and 4. This means that the dataset used in the analysis has no outliers.

Figure 1 also shows that the collision frequency starts with a moderate value at segment 1, 13 collisions, then increases in the following segments to reach its maximum value, 31 collisions, at segments 3 and 4. After that maximum point, the pattern takes the shape of the exponentially decreasing function. This gives the intuitition that the relationship between the collision frequency and the traffic and geometric characteristics of a highway segment is some kind of a nonlinear relationship.

Evaluated Models

Four types of models were evaluated in this study: multiple linear models, Poisson models, negative binomial models, and exponential models. These four types vary in the functional form of their means (expected values of collision frequency). That is, the functional form that relates collision frequency on a given segment to traffic and geometric characteristics of that segment may be additive, multiplicative, or quadratic. The evaluated models are summarized in Table 3 and are described next.

1. Multiple Linear Models:

(I) Additive linear model (AR) of equation (1)

(II) Multiplicative linear model (MR) of equation (3)

(III) Quadratic linear model (QR) of equation (8)

2. Poisson Models of equation (10):

(I) Additive Poisson Models:

AP1: Additive Poisson model with $\mu_i = E(Y_i) = \sum_{j=1}^{k} x_{ij}\beta_j$

AP2 : Additive Poisson model with $\mu_i = E(Y_i) = \exp(\sum_{j=1}^{k} x_{ij}\beta_j)$

(II) Multiplicative Poisson Models:

MP1:Multiplicative Poisson model with $\mu_i = E(Y_i) = \beta_1(\prod_{j=2}^{k} (1 + x_{ij})^{\beta_j})$

MP2:Multiplicative Poisson model with $\mu_i = E(Y_i) = \exp(\beta_1(\prod_{j=2}^{k} (1 + x_{ij})^{\beta_j}))$

(III) Quadratic Poisson Models:

QP1: Quadratic Poisson model with

$$\mu_i = E(Y_i) = \beta_1 + \sum_{j=2}^{k} \beta_j(x_{ij} - \bar{x}_{ij}) + \frac{1}{2}\sum_{j=2}^{k}\sum_{l=2}^{k} \beta_{jl}(x_{ij} - \bar{x}_{ij})(x_{il} - \bar{x}_{il})$$

QP2: Quadratic Poisson model with

$$\mu_i = E(Y_i) = \exp[\beta_1 + \sum_{j=2}^{k} \beta_j(x_{ij} - \bar{x}_{ij}) + \frac{1}{2}\sum_{j=2}^{k}\sum_{l=2}^{k} \beta_{jl}(x_{ij} - \bar{x}_{ij})(x_{il} - \bar{x}_{il})]$$

3. Negative Binomial Models of equation (16):

(I) Additive Negative Binomial Models:

AN1: Additive negative binomial model with $\mu_i = E(Y_i) = \sum_{j=1}^{k} x_{ij}\beta_j$

AN2 : Additive negative binomial model with $\mu_i = E(Y_i) = \exp(\sum_{j=1}^{k} x_{ij}\beta_j)$

(II) Multiplicative Negative Binomial Models:

MN1: Multiplicative negative binomial model with $\mu_i = E(Y_i) = \beta_1(\prod_{j=2}^{k} (1 + x_{ij})^{\beta_j})$

Table 3
Summary of empirically evaluated models and functional forms

Functional Form	Model Type			
	Multiple Linear	Poisson	Negative Binomial	Exponential
Additive	AR	AP1, AP2	AN1, AN2	AE
Multiplicative	MR	MP1, MP2	MN1, MN2	ME
Quadratic	QR	QP1, QP2	QN1, QN2	QE

MN2: Multiplicative negative binomial model with $\mu_i = E(Y_i) = \exp(\beta_1 (\prod_{j=2}^{k} (1 + x_{ij})^{\beta_j}))$

(III) Quadratic Negative Binomial Models:

QN1: Quadratic negative binomial model with

$$\mu_i = E(Y_i) = \beta_1 + \sum_{j=2}^{k} \beta_j (x_{ij} - \bar{x}_{ij}) + \frac{1}{2} \sum_{j=2}^{k} \sum_{l=2}^{k} \beta_{jl} (x_{ij} - \bar{x}_{ij})(x_{il} - \bar{x}_{il})$$

QN2: Quadratic negative binomial model with

$$\mu_i = E(Y_i) = \exp[\beta_1 + \sum_{j=2}^{k} \beta_j (x_{ij} - \bar{x}_{ij}) + \frac{1}{2} \sum_{j=2}^{k} \sum_{l=2}^{k} \beta_{jl} (x_{ij} - \bar{x}_{ij})(x_{il} - \bar{x}_{il})]$$

4. Exponential Models of (19):

AE: Additive exponential models with $\mu_i = E(Y_i) = \exp(\sum_{j=1}^{k} x_{ij} \beta_j)$

ME: Multiplicative exponential model with $\mu_i = E(Y_i) = \beta_1 (\prod_{j=2}^{k} (1 + x_{ij})^{\beta_j})$

QE: Quadratic exponential model with

$$\mu_i = E(Y_i) = \exp[\beta_1 + \sum_{j=2}^{k} \beta_j (x_{ij} - \bar{x}_{ij}) + \frac{1}{2} \sum_{j=2}^{k} \sum_{l=2}^{k} \beta_{jl} (x_{ij} - \bar{x}_{ij})(x_{il} - \bar{x}_{il})]$$

Evaluation Results

To evaluate the models shown in Table 3 empirically, the models were first applied using all independent variables that represent traffic and geometric characteristics of the highway segments (i.e., *Hor*, *Ver*, *Pas*, *Truck*, *NH*, *NV*, and *SP*) and the dependent variable *Acc*. This application showed that some independent variables were not statistically significant and some have an unexpected coefficient signs. Therefore, the next step was to go through a process of selecting the variables that are statistically significant and have practically logical signs of their coefficients. The results of this selection process showed that only *Truck*, *NV*, and *SP* were candidate variables in all models. This is consistent with other studies that have found truck traffic volume to be a significant factor in traffic collisions in rural freeways (Jovanis and Chang 1986) and urban networks (Jadaan and Nicholson 1992). Using these three independent variables, the evaluation results were as described next.

Tables 4-6 show the application results for the six models presented in Table 3, using the maximum-likelihood estimation. These results consist of:

1. The estimated coefficients and the associated t-statistics (in parentheses) for the independent variables (*Truck*,*NV* and*SP*), and the

value of the parameter α for the negative binomial models.

2. The value of the log likelihood function evaluated at the estimated coefficients $l(\beta)$(the full model)
3. The value of the log likelihood function evaluated at zero $l(0)$ (the restricted model)
4. The rho-square statistic $\rho^2 = 1 - \frac{l(\beta)}{l(0)}$
   if $\rho^2 = 0$, the model has no explanatory capability.
   if $\rho^2 = 1$, the model has perfect capability to reproduce the observed collision frequency.
5. The rho-bar-square statistic $\overline{\rho}^2 = 1 - \frac{(l(\beta) - N_\beta)}{l(0)}$, where N_β is the number of model parameters. This measure attempts to eliminate the effect of the number of parameters which will be useful in models comparison.
6. The Akaike Information Criterion AIC = $-2L(\beta) + 2N_\beta$
7. The Likelihood Ratio statistic $LR(0) = 2(l(\beta) - l(0))$, which can be used to test the null hypothesis $H_0 : all\beta = 0$ against the alternative hypothesis $H_1 : \beta \neq 0$ (i.e., the independent variables *Truck*, *NV*, and *SP* have contributions to collision frequency). Under the null hypothesis the statistic $LR(0)$ is distributed as a chi-square variable with a number of degrees of freedom equal to N_β.
8. The *t*-Distribution is the observed (tabulated) *t*-Distribution value with (46 -N_β) degrees of freedom and 0.05 significant level.
9. The χ^2-Distribution is the observed (tabulated) χ^2-Distribution value with N_β degrees of freedom and 0.05 significant level.
10. The Total Prediction is the total value of the predicted collision frequencies for the 46 highway segments where the total number of observed collisions is 489.

Before comparing different models, any model should satisfy the following conditions:

a. The absolute value of the t-statistic of the estimated parameter associated with *Truck*,*NV*, or *SP* should be greater than the *t*-Distribution value.
b. The value of $LR(0)$ should be greater than χ^2-Distribution value.
c. The sign of the parameter value of *Truck*, *NV*, and *SP* should be significant and consistent throughout the models.

The above requirements are important for the validity of any of the proposed models in the prediction process of before-and-after studies to evaluate different policy alternatives that are proposed for reducing collision frequency.

The results of Tables 4-6 show that all models satisfy the above conditions. The estimated parameter α of the Negative Binomial models is not significantly different from zero, which indicates that the data are neither over-dispersed nor under-dispersed and that all the Negative Binomial models can be reduced to their corresponding Poisson models. In addition, the solution procedure of the Negative Binomial models that employs the Broyden-Fletcher-Goldfarb-Shanno algorithm (of the SHAZAM Software) is very sensitive to the initial parameter values, and takes many trials to find suitable initial values for the algorithm to converge.

The signs of the coefficients of *Truck* and *NV* are positive for all models, as expected. The positive sign of the coefficient of *Truck* indicates that collision frequency will increase as the average daily truck traffic increases. The positive sign of the coefficient of *NV* indicates that collision frequency will increase as the number of changes in vertical alignment per kilometer increases. The magnitude of the coefficient of *NV* is much greater than that of *Truck* or *SP*. The negative sign of the coefficient of *SP* indicates that increasing the design speed for the highway segment improves the geometric features of that segment and, in turn, reduces collision frequency.

Now the estimated models can be compared based on $\overline{\rho}^2$, AIC, and total number of predicted collisions. The models with high $\overline{\rho}^2$, low AIC, and total number of predicted collisions that is close to the total observed collisions (489) are preferred. Table 7 shows the comparison among different models and suggests that the quadratic Poisson (exponential) model (QP2) is the most appropriate model. This is consistent with the study by Miaou (1994), who evaluated Poisson and negative binomial models. The table also shows that the functional form has an effect on model accuracy. That is, including the cross effect of Truck x NV and Truck x SP in the quadratic functional form improved the accuracy of all models. Figure 2 shows a comparison of the observed and predicted collision frequencies of the quadratic Poisson (exponential) model.

Concluding Remarks

This paper presented a theoretical review of collision analysis models and empirically evaluated some of these models. The models were classified into two broad categories: aggregate and disaggregate. The aggregate models included linear, exponential-family, time-series, qualitative data analysis, and generalized linear models. The disaggregate models included survival and logit-based models.

Table 4
Results of additive models

Independent Variable	Model Type[a]					
	AR	AP1	AP2	AN1	AN2	AE
Constant	9.447 (0.699)	15.460 (1.382)	0.953 (1.007)	16.998 (1.129)	1.059 (0.905)	0.333 (0.253)
Truck	1.672 (4.239)	1.395 (5.808)	0.186 (5.592)	1.332 (4.588)	0.180 (4.615)	0.219 (3.779)
NV	6.145 (3.062)	4.583 (2.573)	0.410 (3.712)	3.942 (1.685)	0.399 (2.609)	0.421 (4.287)
SP	-0.273 (-2.77)	-0.284 (-3.139)	-0.016 (-2.690)	-0.288 (-2.303)	-0.016 (-2.118)	-0.015 (-2.306)
α				0.054 (1.663)	0.037 (1.431)	
Statistics						
$l(\beta)$	-132.800	-128.256	-124.073	-124.705	-122.288	-128.049
$l(0)$	-181.260	-836.740	-882.742	-836.740	-882.742	-178.037
ρ^2	0.267	0.847	0.859	0.851	0.861	0.281
N_β	4	4	4	5	5	4
$\bar{\rho}^2$	0.245	0.842	0.855	0.845	0.856	0.258
AIC	269.600	260.512	252.146	254.410	249.576	260.097
$LR(0)$	96.921	1416.968	1517.338	1424.070	1520.908	99.976
t-Distribution	1.682	1.682	1.682	1.683	1.683	1.682
χ^2-Distribution	9.488	9.488	9.488	11.070	11.070	9.488
Total Prediction	489.00	489.00	489.00	486.36	488.09	485.50

[a] Entries in each cell are the coefficient and the t-statistic (in parentheses) for model variables.

Table 5
Results of multiplicative models

Independent Variable	Model Type[a]					
	MR	MP1	MP2	MN1	MN2	ME
Constant	0.719 (0.552)	3.966 (0.297)	0.552 (0.702)	5.549 (0.242)	0.609 (0.592)	0.352 (1.737)
Truck	3.135 (6.447)	2.674 (5.644)	1.300 (5.132)	2.588 (4.837)	1.271 (4.491)	1.455 (7.533)
NV	0.914 (5.806)	0.856 (4.172)	0.315 (5.012)	0.820 (2.765)	0.316 (3.052)	0.301 (5.193)
SP	-1.387 (-4.322)	-1.475 (-2.708)	-0.497 (-2.533)	-1.494 (-2.184)	-0.501 (-1.960)	-0.492 (-4.568)
α				.036 (1.457)	0.031 (1.269)	
Statistics						
$l(\beta)$	-128.774	-122.680	-121.402	-121.037	-120.101	-128.044
$l(0)$	-181.260	-836.740	-882.742	-836.740	-882.742	-178.037
ρ^2	0.290	0.853	0.862	0.855	0.864	0.281
N_β	4	4	4	5	5	4
$\bar{\rho}^2$	0.267	0.848	0.858	0.848	0.858	0.258
AIC	261.548	249.360	246.804	247.075	245.201	260.087
$LR(0)$	104.973	1428.120	1522.680	1431.405	1525.283	99.986
t-Distribution	1.682	1.682	1.682	1.683	1.683	1.682
χ^2-Distribution	9.488	9.488	9.488	11.070	11.070	9.488
Total Prediction	485.19	489.46	489.02	488.44	488.76	485.51

[a] Entries in each cell are the coefficient and the t-statistic (in parentheses) for model variables.

Table 6
Results of quadratic models

Independent Variable	Model Type[a]					
	QR	QP1	QP2	QN1	QN2	QE
Constant	11.131 (19.237)	11.041 (21.864)	2.298 (46.159)	10.986 (18.435)	2.305 (33.315)	2.265 (32.631)
Truck	1.4512 (4.011)	1.346 (5.262)	0.169 (4.879)	1.332 (4.416)	0.181 (3.762)	0.202 (3.845)
NV	7.433 (3.875)	6.776 (3.884)	0.475 (2.930)	6.568 (3.099)	0.511 (2.146)	0.477 (1.880)
SP	-0.262 (-2.652)	-0.231 (-2.500)	-0.022 (-3.299)	-0.226 (-1.992)	-0.020 (-1.723)	-0.025 (-4.319)
Truck x NV	3.9 (3.470)	3.174 (3.832)	0.297 (2.878)	3.046 (3.248)	0.323 (2.016)	0.350 (1.992)
Truck x SP	-0.917 (-1.952)	-0.572 (-1.843)	-0.123 (-2.959)	-0.542 (-1.690)	-0.132 (-1.948)	-0.160 (-3.934)
α				0.028 (1.175)	0.018 (0.46)	
Statistics						
$l(\beta)$	-126.814	-121.391	-117.820	-120.332	-117.841	-122.126
$l(0)$	-181.260	-836.740	-882.742	-836.740	-882.742	-178.037
ρ^2	0.300	0.855	0.867	0.856	0.867	0.314
N_β	6	6	6	7	7	6
$\bar{\rho}^2$	0.269	0.849	0.860	0.849	0.859	0.280
AIC	259.628	248.782	241.640	247.664	242.681	250.252
$LR(0)$	108.893	1430.698	1529.844	1432.816	1529.803	111.822
t-Distribution	1.684	1.684	1.684	1.685	1.685	1.684
χ^2-Distribution	12.592	12.592	12.592	14.067	14.067	12.592
Total Prediction	489.00	489.00	489.00	487.22	494.47	483.74

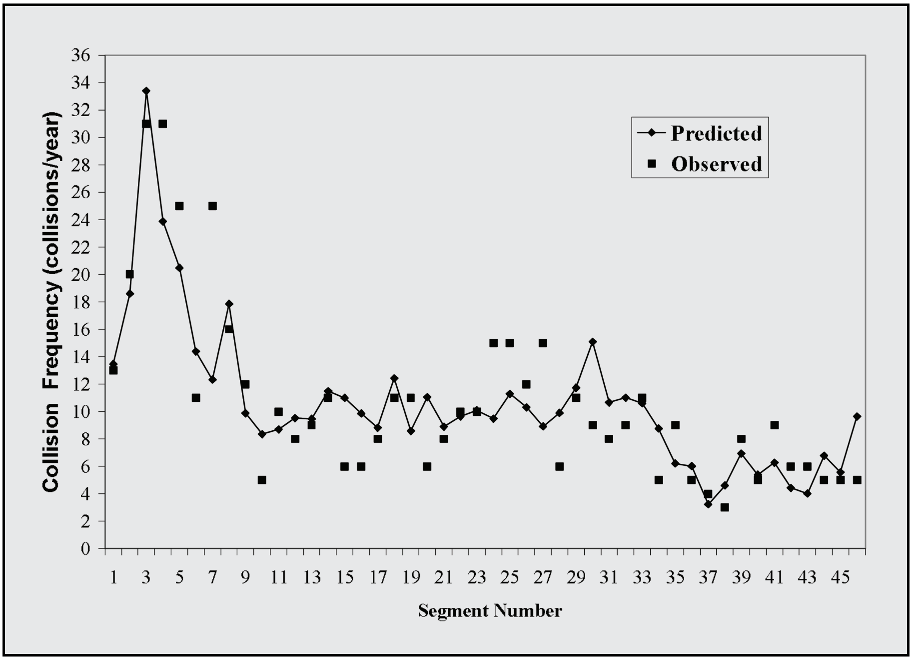
[a] Entries in each cell are the coefficient and the t-statistic (in parentheses) for model variables.

Table 7
$\bar{\rho}^2$, AIC, and total predicted collisions for different functional forms and model types

Functional Form	Model Type[a]					
	Multiple Linear	Poisson	Poisson (exponential)	Negative Binomial	Negative Binomial (exponential)	Exponential
Additive	0.245 (269.6) [489]	0.842 (260.5) [489]	0.855 (252.2) [489]	0.845 (254.4) [486.4]	0.856 (249.6) [488.1]	0.258 (260.1) [485.5]
Multiplicative	0.267 (261.6) [485.2]	0.848 (249.4) [489.5]	0.858 (246.8) [489.0]	0.848 (247.1) [488.4]	0.858 (245.2) [488.8]	0.258 (260.1) [485.5]
Quadratic	0..269 (259.6) [489]	0.849 (248.8) [489]	0.860 (241.6) [489]	0.849 (247.7) [487.2]	0.859 (242.7) [494.5]	0.280 (250.3) [483.7]

[a] Entries in each cell are, $\bar{\rho}^2$, AIC (in parentheses), and total predicted collisions (in brackets).

Figure 2
Observed and predicted collision frequencies of quadratic (exponential) Poisson model



Based on this study, a number of concluding remarks are stated below:

1. The review of collision analysis models presented in this paper included a variety of aggregate and disaggregate models that have sporadically appeared in the literature. Some of the reviewed models represent relatively new concepts, such as the space-based time series model. The characteristics and assumptions of various models are discussed, along with actual applications. It is hoped that this comprehensive review will help provide a better perspective on available collision analysis models and promote their applications to improve road safety.
2. The empirical evaluation of the aggregate models (multiple linear, Poisson, negative binomial, and exponential) with three functional forms (additive, multiplicative, and quadratic) has shown that the quadratic Poisson (exponential) model provided the best results. Generally, the Poisson models were superior to the negative binomial models, while the multiple linear and exponential models were the least favorable. Despite the wide use of Poisson model, it has a restrictive property - the mean should equal the conditional variance. The negative binomial model is likely to provide better results than the Poisson model if the data are over or underdispersed in relation to the mean. The results also showed that the quadratic functional form significantly improved the prediction ability of all models.
3. The data used for evaluating the collision models belong to a multilane expressway in Japan. The expressway is located generally in rural areas and has a gentle (flat) horizontal alignment. The estimated models showed that truck volume, number of vertical alignment changes, and design speed had a significant impact on collision frequency. The estimated models, consistent with previous studies, have confirmed the importance of considering strategies related to truck traffic when attempting to improve road safety.
4. Although highway geometric characteristics are similar worldwide, especially for freeways, the models estimated in this study are not likely to be transferable to other countries in the Middle East or even in North America or Europe. This is mainly due to varying driver and vehicle characteristics. Nonetheless, the authors believe that the finding related to the superiority of the quadratic Poisson (exponential) model will hold regardless of the geographic loca-

tion. The methodologies adopted and reviewed in this paper are likely to open up further research avenues particularly in the field of model transferability.

Acknoeledgements

The data used in this study were prepared while the principal author was a Visiting Scholar at the Institute of Industrial Science, University of Tokyo. The help provided by Professor Masao Kuwahara during data preparation and his generosity are gratefully acknowledged. The authors are grateful to two anonymous reviewers for their thorough and most helpful comments. This research was financially supported by a Discover Grant from the Natural Sciences and Engineering Research Council of Canada.

References

Cameron, A.C. and Trivedi, P.K. 1986. Econometric models based on count data: Comparison and application of some estimations and tests. *Journal of Applied Econometrics*, 1: 29-53.

Chang, H.L. and Jovanis, P.P. 1990. Formulating accident occurrence as a survival process. *Accident Analysis and Prevention,* 22(5): 407-419.

Cox, D. R. 1972. Regression models and life tables. *Journal of Royal Statistical Society,* 34B: 187-220.

Easa, S.M. 2002. *Geometric design*. In Civil Engineering Handbook, W.F. Chen and J.Y.Liew eds., CRC Press, Boca Raton, FL, Chapter 63: 1-39.

Efron, B. 1986. Double exponential families and their use in generalized linear regression. *Journal of American Statistical Association,* 81(395): 709-721.

Fleming, T. R. and Harrington, D. P. 1990. *Counting process and survival analysis*. John Wiley and Sons, New York, N.Y..

Frome, E.L. 1983. The analysis of rates using Poisson regression models. *Biometrics*, 39: 665-674.

Frome, E.L., Kutner, M.H., and Beauchamp, J.J. 1973. Regression analysis of Poisson-distributed data. *Journal of American Statistical Association,* 68: 935-940.

Gelfand, A.E. and Dalal, S.R. 1990. A note on overdispersed Exponential Families. *Biometrika*, 71(1), 55-64.

Gouieroux, C., Monfort, A., and Trognon, A. 1984. Pseudo maximum likelihood methods: Application to Poisson models. *Econometrica*, 52: 701-720.

Greene, W. 1993. *Econometrics analysis*. Macmillan Publishing, New York, N.Y.

Hamdy, H.I., AL-Mahmeed, M.A., and El-Habshi, M. H. 1998. Modeling severity of auto accidents in Kuwait. *Arab Journal of Administrative Sciences*, 5(2): 519-532.

Hamed, M.M. and Easa, S.M. 1998. Ordered probability modeling of seat belt usage. *Journal of Transportation Engineering*, ASCE, 124(3): 271-276.

Hamed, M.M., Hensher, D.A., and Al-Masaied, H.R. 1997. The duration between traffic accidents in the taxi-cab sector: An empirical study. *Road and Transport Research*, 6(1): 64-73.

Hamed, M.M. and Mannering, F.L. 1993. Modeling travelers' post-work activity involvement: Toward a new methodology. *Transportation Science*, 27: 381-394.

Hasan, M.K., Easa, S.M., and Abd El Halim, A.O. 1998. Toward a new methodology for modeling highway accident frequency. *Proceedings of 2nd Transportation Specialty Conference*, CSCE, Halifax, NS, Canada, 1998.

Hensher, D.A. and Mannering, F.L.1994. Hazard-based duration models and their application to transport analysis. *Transport Reviews*. 14(1): 63-82.

Holford, T.R. 1983. The estimation of age, period and cohort effects for vital rates. *Biometrics,* 39: 311-324.

Jadaan, K.S. and Nicholson, A.J. 1992. Relationships between road accidents and traffic flows in an urban netwok. *Traffic Engineering and Control*, 33(9): 507-511.

Jones, B., Janssen, L., and Mannering, F.L. 1991. Analysis of the frequency and duration of freeway accidents in Seattle. *Accident Analysis and Prevention*, 23(2): 239-255.

Jovanis, P.P. and Chang, H.L. 1986. Modeling the relationship of accidents to miles traveled. *Transportation Research record 1068*: 42-51.

Kalbfliesch, J. and Prentice, R. 1981. *The statistical analysis of failure time data*. John Wiley and Sons, New York, N.Y..

Lawless, J.F. 1987. Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics,* 15, 209-225.

Mannering, F.L. 1993. Male/female driver characteristics and accident risk: Some new evidence. *Accident Analysis and Prevention,* 25(1): 77-84.

Mannering, F.L. and Hamed, M.M. 1990. Occurrence, frequency, and duration of commuters' work-to-home departure delay. *Transportation Research*, 24B(1): 99-110.

Mannering, F.L. and Winston, C. 1987. *Recent automobile occupant safety proposal*: *Blind Intersection*. The Brooking Institution, Washington, D.C., 69-88.

Miaou, S. and Lum, H. 1993. Modeling vehicle accidents and highway geometric design relationships. *Accident Analysis and Prevention,* 25(6): 689-709.

Miaou, S.P. 1994. The relationship between truck accidents and geometric design of road section: Poisson versus negative binomial regression. Accident Analysis and Prevention, 26: 471-482.

Oppe, S. 1992. A comparison of some statistical Techniques for road accident analysis. *Accident Analysis and Prevention,* 24(4): 397-423.

Scott, P.P. 1986. Modeling time-series of British road accident data. *Accident Analysis and Prevention,* 18(2): 109-116.

Sergio, R.J. and Sergio, M.G. 1986. Flexible models for accidents on Chilean roads. *Accident Analysis and Prevention,* 18(2): 103-108.

الملخص

نماذج لتحليل حوادث التصادم المروري:
مراجعات أدبية وتقييم تجريبي

سعيد عيسى
جامعة رايرسون

محمد كمال حسن
جامعة الكويت

محمد حامد
جامعة الأردن للعلوم والتكنولوجيا

تم تصنيف نماذج تحليل حوادث التصادم في هذا البحث إلى نوعين: النماذج الإجمالية، والنماذج التفصيلية؛ حيث تم عرض للمراجعات الأدبية من الجانب النظري لهذه النماذج، تلا ذلك تقييم تجريبي مقارن لكل من النماذج الخطية المتعددة: نماذج بواسون، ونماذج ذي الحدين السالب والنماذج الآسية، مستخدمين ثلاثة أنواع من أشكال الدوال الرياضية: (الدوال التجمعية، والدوال المضروبة والدوال من الدرجة الثانية). ولقد وضحت نتائج هذا التقييم أن شكل الدالة الرياضية المستخدم لكل نموذج يعد مهمًا جدًا لمدى دقة التنبؤات باستخدام هذا النموذج. وقد أوضحت النتائج أيضا أن نموذج بواسون مع دالة رياضية من الدرجة الثانية تعطي أفضل دقه للتنبؤ على حين تعطي النماذج الخطية المتعددة والنماذج الآسية نتائج أقل دقه للتنبؤ.

Said M. Easa (Ph.D. Transportation, University of California at Berkeley, California, U.S.A., 1982). Professor and Chair, Department of Civil Engineering, Ryerson University, Toronto, Ontario, Canada. Current research interests include Road safety, Highway geometric design, Human factors in transportation, and Traffic operations and management. He has authored and co-authored more than 300 publications, including 160 refereed-journal articles and book contributions. Dr. Easa's work received numerous national and international best-paper and lifetime achievement awards, including the 2001 Frank M. Masters Transportation Engineering Award from American Society of Civil Engineers in recognition of "outstanding contributions to ASCE and the transportation profession throughout his career."

Mohamad K. Hasan (Ph.D. Transportation Systems, Texas A&M University, Texas, U.S.A., 1991) Associate Professor, Department of Quantitative Methods and Information Systems, College of Business Administration, Kuwait University. His current research interests include: Network Optimization, Transportation Systems Analysis', Transportation Network Modeling, Freight Transportation Modeling, and Travel Demand Modeling, Logistics, Supply Chain Management, Production and Operations Management, Scheduling, Simulation, and Mathematical Programming

Mohammad M. Hamed (Ph.D. Transportation Planning, University of Wshington, Seattle, U.S.A., 1990). Professor of Transportation Planning, Department of Civil Engineering, Jordan University of Science and Technology, Irbid, Jordan and Director General of the Public Transport Regulatory Commission, Amman, Jordan. Current research interests include Travel Demand Modeling, Travel Activity-based analysis, Transportation Systems Analysis, Privatization of transport services, Public sector restructuring and reform, Public transportation demand modeling, Design of travelers information strategies, Monitoring public transportation services, and Public transportation tracking systems.

إعلان