Manuscript Number: TE/2006/023538

MODELING THE PROBABILITY OF FREEWAY REAR-END CRASH OCCURRENCE

by

Joon-Ki Kim[1], Yinhai Wang[2], and Gudmundur F. Ulfarsson[3*]

Initial submission: February 8, 2006. Revised submission: June 12, 2006.

Submitted to the *Journal of Transportation Engineering*

[1] Graduate Research Assistant, Department of Civil Engineering, Campus Box 1130, Washington University in St. Louis, One Brookings Drive, St. Louis MO 63130-4899; jk9@cec.wustl.edu; telephone (314) 935-4126; fax (314) 935-4338

[2] Assistant Professor, Department of Civil and Environmental Engineering, Box 352700, University of Washington, Seattle WA 98195-2700; yinhai@u.washington.edu; telephone (206) 616-2696, fax (206) 543-5965

[3] Assistant Professor, Department of Civil Engineering, Campus Box 1130, Washington University in St. Louis, One Brookings Drive, St. Louis MO 63130-4899; gfu@wustl.edu; telephone (314) 935-9354; fax (314) 935-4338

CE Database Keywords: Collisions; Crash; Accident; Analysis; Models; Highway Safety; Driver Response Time; Risk

[*] Corresponding author

ABSTRACT

A microscopic model of freeway rear-end crash risk is developed based on a modified negative binomial regression and estimated using Washington State data. Compared with most existing models, this model has two major advantages: 1) it directly considers a driver's response time distribution; and 2) it applies a new dual-impact structure accounting for the probability of both, a vehicle becoming an obstacle (P_o) and the following vehicle's reaction failure (P_f).

The results show for example that truck percentage-mile-per-lane has a dual impact, it increases P_o and decreases P_f, yielding a net decrease in rear-end crash probabilities. Urban area, curvature, off-ramp and merge, shoulder width, and merge section are factors found to increase rear-end crash probabilities. Daily VMT per lane has a dual impact, it decreases P_o and increases P_f, yielding a net increase, indicating for example that focusing VMT related safety improvement efforts on reducing drivers' failure to avoid crashes, such as crash-avoidance systems, is of key importance. Understanding such dual impacts is important for selecting and evaluating safety improvement plans for freeways.

INTRODUCTION

Approximately 60% of freeway traffic congestion is caused by incidents (Lindley, 1987). Incidents can be classified as either predictable events such as work zones, or unexpected events such as accidents. Rear-end crashes are the most common type of crash in Washington State: rear-end crashes (35.9%), fixed object crashes (17.0%), and sideswipes (10.7%) (WSDOT, 1996). When rear-end crashes occur, they temporarily reduce roadway capacity and cause congestion. According to the 2003 Urban Mobility Report (Schrank and Lomax, 2003), the annual average delay per person in the 75 surveyed urban areas was 26 hours in 2001, a 371% increase compared to 1982. Congestion costs an average of $520 per traveler in the surveyed urban areas in 2001. Therefore, through finding the factors which influence rear-end crashes, we can identify controllable factors which can improve highway design, leading to a decrease in the frequency of rear-end crashes. If successful, this will help reduce number of injuries and reduce overall congestion, thus also saving time and money. This paper describes a numerical approach that can be used to evaluate freeway rear-end crash risk based on known traffic and roadway factors.

LITERATURE REVIEW

In recent years, a significant amount of research has been performed to understand crashes on freeways using modeling methods such as linear regression, Poisson regression, and negative binomial regression. Jovanis and Chang (1986) found some undesirable problems with the use of linear regression in their study. Miaou et al. (1992) used a Poisson model and found that the Poisson constraint (the mean and variance of the crash frequency have to be equal) was violated. The performances of the Poisson regression and negative binomial regression were compared and the overdispersion of crash data was addressed (Miaou, 1994 and Shankar et al., 1995). Poch and Mannering (1996) found that the negative binomial model was the appropriate model for determining crash frequency at intersections due to overdispersion in the data. Shankar et al. (1997) applied zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB) to handle data which violate the Poisson and negative binomial model assumptions due to numerous observations of sections with no crashes in the observed period. Wang (1998) modeled the mean rates of rear-end crashes at four-legged signalized intersections through multiplying traffic volume by rear-end crash probability.

A common criticism of many previous studies is that they do not usually consider human factors. Massie et al. (1993), however, pointed out that the classical human factors approach ignored the problem associated with classifying collisions and their related causes, be it human or otherwise, and failed to address the issue of helping drivers avoid collisions. By identifying geometric conditions that lend themselves to producing crashes, these conditions could be corrected.

Milton and Mannering (1998) estimated annual crash frequency on sections of principal arterials with negative binomial regression models and found numerous traffic and geometric characteristics to be important. Carson and Mannering (2001) identified significant spatial, temporal, traffic, roadway, and crash characteristics that influenced ice involved-crash frequency and severity. Lee and Mannering (2002) used a nested logit model for run-off-roadway crash modeling. Golob and Recker (2003) applied linear and nonlinear multivariate statistical analyses to determine how the types of crashes occurring on heavily used freeways in Southern California are related both to the flow of traffic and to weather and ambient lighting conditions. Ulfarsson and Shankar (2003) explored the negative multinomial model to predict median crossover crash frequencies. Golob and Regan (2004) studied, by applying a multinomial logit model, how various types of truck crashes are related to traffic flow conditions and roadway characteristics on urban freeways.

Although a significant amount of research has attempted to study crashes based on crash type, location, and severity, very few studies have been conducted to model rear-end crashes on freeways. Shankar et al. (1995) noted that separate regression models focused on specific crash types have greater explanatory power than an overall frequency model. Therefore, there is need for studying rear-end crashes separately from other types of crashes on freeways.

METHODOLOGY

For modeling rear-end crashes on freeways, we employ a microscopic modeling approach introduced by Wang (1998). This modeling approach has been successfully applied to intersection safety studies (Wang et al., 2002 and Wang and Nihan, 2003). The occurrence of rear-end crashes on freeways is a combined result of a lead vehicle's time-headway reduction action and a following vehicle's inadequate action or the following vehicle's ineffective response. In this study, the occurrence of crashes is considered to be based on two premises: one is that a lead vehicle becomes an obstacle vehicle to a following vehicle and the other is that a following vehicle fails to avoid a collision given the obstacle vehicle.

When a lead vehicle reduces the time-headway with respect to the following vehicle (such as by stopping, decelerating, or performing a cut-in movement), it becomes an obstacle vehicle to the following vehicle. The following vehicle driver may need to react to avoid a collision with the obstacle vehicle. Depending on the Maneuvering Time (Perception-Response Time (PRT) plus vehicle response time) available to the driver, the driver's reaction may or may not be successful. If unsuccessful, a rear-end crash occurs. Thus, the probability of having a rear-end crash is determined by: 1) the probability of a leading vehicle becoming an obstacle, denoted by P_o; and 2) the probability of the following vehicle driver's failure to avoid the collision given an obstacle vehicle, denoted by P_f. Noting the conditional nature of P_f avoids problems related to the dependence of the two drivers' decision making when both see a joint event that leads both to brake. Since the following vehicle's failure to avoid a crash is conditional on there being an obstacle vehicle, the total probability of a rear end crash is the multiplication of the probability of an obstacle vehicle and the conditional failure to avoid crashing. Then the probability of having a rear-end crash can be expressed as the product of P_o and P_f:

$$P = P_o P_f . \tag{1}$$

Note that different rear-end crashes are assumed to be independent events because chain-reaction crashes are excluded and only two-vehicle rear-end crashes are used in this study. There were a total of 8,452 rear-end accidents in the data and two-vehicle rear-end accidents accounted for about 64% (5,868). Chain-reaction rear-end crashes are more likely under high volume conditions. Therefore, excluding the chain-reaction rear-end crashes may affect the traffic volume variable.

The Probability of a Leading Vehicle becoming an Obstacle (P_o)

An event that causes a lead vehicle to become an obstacle vehicle is called a disturbance. Disturbances are rare events, non-negative, and discrete. Also, the occurrences of disturbances are independent during non-overlapping time intervals and different disturbances are independent of each other. Therefore, the occurrence of disturbances is assumed to follow a

Poisson process. The intervals between Poisson-distributed disturbances follow an exponential distribution. The probability density function (PDF) of the exponential distribution is

$$f(t,\eta_j)=\eta_j\, e^{-\eta_j t}, \quad for\ t>0,\ \eta_j>0\ , \tag{2}$$

where j is a disturbance, η_j is the occurrence rate of disturbance j, and t is the time interval. This leads to the probability of a disturbance j occurring at least once in t

$$P_j=\int_0^t \eta_j e^{-\eta_j t} dt \ =\ 1-e^{-\eta_j t}\ \ . \tag{3}$$

Since any of the disturbances can cause the lead vehicle to become an obstacle vehicle, the probability of that, P_o, is the same as the probability that at least one disturbance occurs in t expressed as

$$P_o=1-\prod_{j=1}^{J}(1-P_j)\ , \tag{4}$$

where $1-P_j$ is the probability that disturbance j does not occur, J is a theoretical maximum number of disturbances that can occur in time interval t (since we cannot have an infinite number of disturbances occur in a finite time), and $\prod_j(1-P_j)$ is the probability that no disturbance occurred during time interval t. Substitute P_j in (4) by (3) and P_o can be written

$$P_o=1-e^{-\Sigma_j \eta_j t}\,. \tag{5}$$

To let (5) depend on a set of explanatory variables such as geometric features and traffic flow we parameterize $\sum_j \eta_j t$, noting it must be positive since probability cannot be greater than 1. A loglinear parameterization satisfies this condition:

$$\ln\left(\sum_j \eta_j t\right) = \boldsymbol{\beta}_o \mathbf{x}_o\,, \tag{6}$$

$$\sum_j \eta_j t = \exp(\boldsymbol{\beta}_o \mathbf{x}_o)\,, \tag{7}$$

where $\boldsymbol{\beta}_o$ is a vector of estimable parameters and $\mathbf{x}_o$ is a vector of explanatory variables. By combining (5) and (7), the probability of a leading vehicle becoming an obstacle (P_o) in a given period of time is written

$$P_o = 1 - e^{-e^{\boldsymbol{\beta}_o \mathbf{x}_o}}\,. \tag{8}$$

The Probability of Failure to Avoid a Collision given an Obstacle Vehicle (P_f)

One of the most important factors to avoid crashes on freeways is a driver's Perception-Response Time (PRT). PRT is defined as the PIEV time in the Manual on Uniform Traffic Control Devices (MUTCD, 2003), which can be summarized as "the total time needed to perceive and complete a reaction". PRT is not a constant value of all driving situations but depends on the complexity of the problem and the driver's expectation of a hazard (Bates, 1995).

To model the probability of a driver's failure to avoid a collision, two concepts are considered: Available Maneuvering Time (AMT) and Needed Maneuvering Time (NMT). AMT refers to the actual time available for a driver to avoid a collision with an obstacle vehicle. NMT refers to the minimum time that a driver needs to avoid a collision (PRT plus vehicle response time). If NMT is greater than AMT, a driver cannot avoid a collision.

To model NMT and AMT with appropriate distributions, we need to know the characteristics of PRT. Summala (2000) addressed the following points: 1) not all drivers perform the expected response in on-road studies, and the obtained PRT estimates may be biased due to the drivers who brake the slowest; 2) drivers' attentions differ between locations so that in certain places they are more attentive to their task than in others; 3) although brake reaction latencies appear to increase with available time, steering response latencies do not, at least within a certain range of time; 4) the total PRT distributions do not differ at all for the two groups (18-40 years and 50-84 years). This result was also noted by other studies. For example, Olson and Sivak (1986) showed that both age groups have the 95th percentile PRT time of about 1.6 s. They indicated that while older drivers' perception time is slower than younger drivers, the brake reaction (including foot movement and decision processes) that follows is faster in older drivers. Lerner (1993) also pointed out that although most of the fastest observed PRTs were from the young group, there were no differences in central tendency (mean = 1.5 s) or upper percentile values (85th percentile = 1.9 s) among the age groups. While AASHTO (2001) suggests a conservative PRT of 2.5 seconds for highways, Mannering et al. (2005) mentioned that a driver's PRT is a function of a number of factors including the driver's physical condition, emotional state, and complexity of the situation.

As explained in the above studies, PRT is not a constant value, but a random variable relating to many factors such as drivers' skill, physical condition, traffic condition, and geometric features. In this study, we assumed that both the AMT and NMT are Weibull distributed because the Weibull distribution is a good approximation to the normal distribution (Plait, 1962) and for this model it results in closed-forms, whereas the lognormal and the log-logistic do not.

The Weibull distribution is a generalized form of the exponential distribution. The Weibull distribution has two parameters, scale $\theta > 0$ and shape $\alpha > 0$. The density function for the Weibull distribution is

$$f(t,\theta,\alpha) = \alpha\,\theta^{\alpha}\,t^{\alpha-1}e^{-(\theta t)^{\alpha}},\ t > 0. \tag{9}$$

A hazard function is a conditional probability that an event occurs between t and $t+\Delta t$ given that an event does not occur until t. The hazard function for the Weibull distribution is

$$h(t) = \alpha\theta^{\alpha}t^{\alpha-1},\ t > 0. \tag{10}$$

When the shape $\alpha = 1$, the Weibull distribution becomes the exponential distribution, and the hazard is constant over time (duration independence). When $\alpha > 1$, the hazard is monotonically decreasing over time (negative duration dependence), and when $\alpha < 1$, the hazard is monotonically increasing over time (positive duration dependence). Note that although the Weibull distribution provides a more flexible means of capturing duration dependence than the exponential distribution, it does not allow the hazard to increase and then decrease over time because it requires the hazard to be monotonic over time. The Log-normal and log-logistic distributions have non-monotonic hazard functions but are computationally cumbersome in this model because of non-closed form solutions. We therefore use the Weibull distribution here in spite of its limitation.

The failure probability is expressed as $P_f = P(AMT < NMT)$, as mentioned before. The probability distributions for AMT and NMT are assumed as follows:

$$AMT = f(t_a,\theta,\alpha) = \alpha\,\theta^{\alpha}\,t^{\alpha-1}e^{-(\theta t_a)^{\alpha}}\ \ for\ \alpha>0,\ \theta>0,\ t>0, \tag{11}$$

$$NMT = f(t_n,\lambda,\alpha) = \alpha\,\lambda^{\alpha}\,t^{\alpha-1}e^{-(\lambda t_n)^{\alpha}}\ \ for\ \alpha>0,\ \lambda>0,\ t>0. \tag{12}$$

Here, we employed two assumptions as follows: 1) The shape parameter, α, is the same for both AMT and NMT. This is a limitation but is necessary to achieve a closed form result. The scale parameter is however allowed to vary. 2) AMT and NMT are independent maneuvering times. Then, the drivers' probability of failure to avoid a collision can be calculated as,

$$
\begin{aligned}
P_f &= P(NMT > AMT), \\
&= \int_0^{\infty}\int_{t_a}^{\infty} f(t_n,\lambda,\alpha)\, f(t_a,\theta,\alpha)\,dt_n dt_a, \\
&= \int_0^{\infty} e^{-(\lambda t_a)^{\alpha}} \alpha\theta^{\alpha} t_a^{\alpha-1} e^{-(\theta t_a)^{\alpha}}\,dt_a, \\
&= \int_0^{\infty} \alpha\theta^{\alpha} t_a^{\alpha-1} e^{-(\lambda^{\alpha}+\theta^{\alpha})t_a^{\alpha}}\,dt_a, \\
&= \frac{-\theta^{\alpha}}{(\lambda^{\alpha}+\theta^{\alpha})} e^{-(\lambda^{\alpha}+\theta^{\alpha})t_a^{\alpha}}\Bigg|_0^{\infty}, \\
&= \frac{1}{1+\left(\dfrac{\lambda}{\theta}\right)^{\alpha}}.
\end{aligned}
\tag{13}
$$

Now, P_f is expressed as a function of λ, θ, and α. Since $\lambda > 0$ and $\theta > 0$, $(\lambda/\theta)^{\alpha}$ is greater than 0. Then, $(\lambda/\theta)^{\alpha}$ can be related to a set of explanatory variables by using the exponential link function:

$$
\left(\frac{\lambda}{\theta}\right)^{\alpha} = e^{-\boldsymbol{\beta}_{\mathbf{f}}\mathbf{x}_{\mathbf{f}}}, \tag{14}
$$

where $\boldsymbol{\beta}_f$ is a vector of estimable parameters and $\mathbf{x}_f$ is a vector of explanatory variables. P_f can then be expressed as

$$
P_f = \frac{1}{1+e^{-\boldsymbol{\beta}_{\mathbf{f}}\mathbf{x}_{\mathbf{f}}}}. \tag{15}
$$

Rear-End Crash Risk Model

By replacing Equation (1) with (8) and (15), the probability of a rear-end crash of an individual vehicle, P, can be rewritten as,

$$P = P_o P_f = \frac{1-e^{-e^{\boldsymbol{\beta}_\mathbf{o}\mathbf{x}_\mathbf{o}}}}{1+e^{-\boldsymbol{\beta}_\mathbf{f}\mathbf{x}_\mathbf{f}}}. \tag{16}$$

The number of crashes (N) for a vehicle flow (v_{ij}) on section i in a given period j, follows a binomial distribution,

$$P(N = n_{ij}) = \binom{v_{ij}}{n_{ij}} P^{n_{ij}} (1-P)^{v_{ij}-n_{ij}}. \tag{17}$$

The mean for this binomial distribution is $m_{ij} = v_{ij}P$. When $v_{ij} \to \infty$ and $P \to 0$, while $v_{ij}P$ remains constant, the number of crashes, N, is a Poisson distributed random variable with the parameter m_{ij},

$$P(N = n_{ij} \mid m_{ij}) = \frac{{m_{ij}}^{n_{ij}} e^{-m_{ij}}}{n_{ij}!}. \tag{18}$$

While the probability of crashes is a very small value, the traffic volume (v_{ij}) is a very large value for a given time period. Therefore, the Poisson distribution can be a good approximation to the binomial distribution as proven above.

Given data such as traffic flow and geometric features, the expected rear-end crashes (m_{ij}) on section i in period j can be parameterized as

$$\ln m_{ij} = \boldsymbol{\beta}\,\mathbf{x}_{ij}, \tag{19}$$

where $\mathbf{x}_{ij}$ is a vector of geometric features, traffic flow, and so on for section i in the given period j, and $\boldsymbol{\beta}$ is a vector of estimable coefficients.

Poisson models are not suitable for over-dispersed data. However, crash data tends to be over-dispersed. To overcome this limitation, the Poisson model is generalized by introducing an unobserved effect, ε_{ij}, into the expected rear-end crashes parameterization (Greene, 2003),

$$\begin{aligned} \ln m'_{ij} &= \boldsymbol{\beta}\mathbf{X}_{ij} + \varepsilon_{ij} = \ln m_{ij} + \ln u_{ij}, \\ &= \ln m_{ij} u_{ij}, \end{aligned} \tag{20}$$

where for mathematical simplicity we define $\ln u_{ij} = \varepsilon_{ij}$ and use the logarithmic rule $\ln ab = \ln a + \ln b$ to simplify further. Then, the distribution of n_{ij} conditioned on u_{ij} (i.e. ε_{ij}) is

$$P(N = n_{ij} \mid u_{ij}) = \frac{(m_{ij} u_{ij})^{n_{ij}} e^{-m_{ij} u_{ij}}}{n_{ij}!}. \tag{21}$$

The unconditional distribution $P(N = n_{ij})$ is the expected value of $P(N = n_{ij} \mid u_{ij})$,

$$P(n_{ij}) = \int_0^\infty \frac{(m_{ij} u_{ij})^{n_{ij}} e^{-m_{ij} u_{ij}}}{n_{ij}!} g(u_{ij}) du_{ij}. \tag{22}$$

For mathematical convenience, a gamma distribution is assumed for u_{ij}, i.e. exp(ε_{ij}). When $E[u_{ij}]$ is 1 and $V[u_{ij}]$ is δ, the density $g(u_{ij})$ can be expressed as,

$$g(u_{ij}) = \frac{\kappa^\kappa}{\Gamma(\kappa)} e^{-\kappa u_{ij}} u_{ij}^{\kappa-1}, \tag{23}$$

where $\kappa = 1/\delta$ and $\Gamma(\cdot)$ is the gamma function. Then, the probability of the number of crashes is written,

$$
\begin{aligned}
P(N=n_{ij}) &= \int_0^\infty \frac{(m_{ij}u_{ij})^{n_{ij}} e^{-m_{ij}u_{ij}}}{n_{ij}!} \frac{\kappa^\kappa}{\Gamma(\kappa)} e^{-\kappa u_{ij}} u_{ij}^{\kappa-1} du_{ij}, \\
&= \frac{\kappa^\kappa m_{ij}^{n_{ij}}}{\Gamma(n_{ij}+1)\Gamma(\kappa)} \int_0^\infty e^{-(m_{ij}+\kappa)u_{ij}} u_{ij}^{\kappa+n_{ij}-1} du_{ij}, \\
&= \frac{\kappa^\kappa m_{ij}^{n_{ij}} \Gamma(\kappa+n_{ij})}{\Gamma(n_{ij}+1)\Gamma(\kappa)(m_{ij}+\kappa)^{\kappa+n_{ij}}}, \\
&= \frac{\Gamma(\kappa+n_{ij})}{\Gamma(n_{ij}+1)\Gamma(\kappa)} r_{ij}^{n_{ij}} (1-r_{ij})^\kappa,
\end{aligned}
\tag{24}
$$

where $r_{ij} = \dfrac{m_{ij}}{m_{ij}+\kappa}$, $m_{ij} = v_{ij}P$.

This distribution has conditional mean m_{ij} and conditional variance $E[n_{ij}][1+\delta E[n_{ij}]]$ (where $\kappa = 1/\delta$). The negative binomial model can be estimated by maximum likelihood. Using (24), the log-likelihood function for the negative binomial model is,

$$
L(m_{ij}) = \prod_{i=1}^{I} \prod_{j=1}^{T} \ln \left[\frac{\Gamma(\kappa+n_{ij})}{\Gamma(n_{ij}+1)\Gamma(\kappa)} \left[\frac{\kappa}{\kappa+m_{ij}} \right]^\kappa \left[\frac{m_{ij}}{\kappa+m_{ij}} \right]^{n_{ij}} \right], \tag{25}
$$

where $m_{ij} = v_{ij}P = v_{ij} \cdot \dfrac{1-e^{-e^{\boldsymbol{\beta}_o \mathbf{x}_o}}}{1+e^{-\boldsymbol{\beta}_f \mathbf{x}_f}}$, I is the total number of freeway sections, and T is the number of years of crash data. This function is maximized to obtain coefficient estimates for $\boldsymbol{\beta}$ ($\boldsymbol{\beta}_o$ and $\boldsymbol{\beta}_f$) and κ. If the estimated κ is statistically significant (significantly different from zero), the negative binomial regression model is more appropriate than the Poisson model.

DATA DESCRIPTION

Data from the Highway Safety Information System (HSIS) were employed for developing the relationships between rear-end crashes and explanatory variables. The crash data used for this study are two-vehicle rear-end crashes that occurred on I-5 in Washington State from 2001 to 2002. The HSIS classified roadway sections were used as crash observation units. Each roadway section represents a homogenous link in terms of curvature and cross-sectional characteristics, such as number of lanes, lane width, median type and width, and shoulder width. Traffic factors such as traffic volume and truck percentage play an important role in crashes. Unfortunately, traffic data when crashes occur was not available. Therefore, Annual Average Daily Traffic (AADT) and percent trucks were used for calibration in our study and roadway sections without AADT and truck percentage data were excluded from the quantitative analysis.

Traffic variables: AADT and Truck percent data. We generalized the AADT variable by considering section length and number of lanes. Since the number of lanes varies from section to section and the chance for a section to have a crash increases with section length, the AADT variable must be generalized to satisfy the requirement of this microscopic approach. The generalized AADT is called "Daily VMT per lane". It was calculated from AADT, section length, and number of lanes as follows:

$$Daily\ VMT\ per\ lane\ = \frac{AADT \times Section\,length}{The\,number\,of\ lanes \times 1000}\,. \tag{26}$$

The divisor of 1000 was used to expedite the calculation speed of model calibration.

Similarly, truck data were also generalized by considering section length and the number of lanes. "Truck percentage-mile per lane" is a function of truck percentage, section length, and the number of lanes. It was calculated as follows:

$$Truck\ percentage\text{-}mile\ per\ lane = \frac{Truck\,\% \times Section\,length}{The\,number\,of\ lanes}\,. \tag{27}$$

Here, the variable is divided by lane to explain the effect of the number of lanes. For example, although a one lane road and a four lane road have the same truck percentage, the effect will be different. Note that although this variable is standardized per lane, there is still potential inaccuracy, since trucks typically concentrate in the right and middle lanes. To account for exposure to truck traffic the percent trucks is multiplied by the section length.

Freeway geometric variables: Shoulder width, horizontal curvature, the number of ramps, and the number of lanes. To reflect the effects of geometric features on rear-end crashes, the variables mentioned above were combined or transformed. Total shoulder width (the sum of left and right shoulder widths) was considered due to the high correlation between left and right shoulder widths. We also created a new variable called "deviation of shoulder width". It is defined as:

$$\textit{Deviation of shoulder width} = \max\{0, 18 - \textit{Total shoulder width}\}\,. \tag{28}$$

Here, we considered 5.5 m (18 feet) as an ideal total shoulder width. This variable explains the effect of the deviation of ideal shoulder width for a particular road section.

Each road section has one horizontal curvature. Curvature-per-length was assumed to have different effects on the occurrence of rear-end crashes. The variable called "curvature-per-length" is defined as:

$$\textit{Curvature-per-length} = \frac{\textit{Degree of curvature}}{\textit{Section length} \times 10}\,. \tag{29}$$

The divisor of 10 was used to expedite the calculation.

To explain the effects of merging on freeway rear-end crashes, a variable called "merge section" is introduced. It is a binary variable with value "1" if a section is within 0.8 km (0.5 mile) upstream of a merge point and with value "0" otherwise. We assume that vehicles have a tendency to change lanes within 0.8 km (0.5 mile) before a merge point.

In sections with off-ramps or on-ramps, vehicles are likely to change lanes to exit or enter into the mainline of traffic. Also, this phenomenon is more likely to occur in sections containing both merging lanes and ramps. To reflect this fact, a variable called "off-ramp and merge" was devised (on-ramps were excluded here because they did not turn out to have statistical significance in the model). This variable is defined as:

$$\textit{Off-ramp and merge} = \textit{the number of off-ramps in a section} \times \textit{Merge ratio}\,, \tag{30}$$

where "merge ratio" is defined as,

$$\textit{Merge ratio} = \frac{\textit{The number of lanes in an upstream section}}{\textit{The number of lanes in a downstream section}}\,. \tag{31}$$

Land use variable: an indicator of land use, split here simply into rural or urban. Freeway sections have different characteristics depending on whether they are in an urban area or rural area. To include this effect in the model, the variable called "urban area" was created. It is a binary variable: "urban area" = 1 when the section is in an urban area; "urban area" = 0 if the section is in a rural area.

Traffic control variable: speed limit. Another important variable in explaining crashes on freeways is the posted speed limit. Posted speed limits on freeways can be expected to be correlated with travel speed, but this correlation breaks down during congestion as travel speeds drop and speeds are governed more by the congestion, not roadway geometrics. During congestion, chain-reaction accidents are more likely to occur whereas in this paper we exclude these and focus on rear-end crashes between only two vehicles. Such accidents are not as closely tied to congestion as chain-reaction rear-end crashes. The posted speed limits are therefore likely to be correlated with travel speed in our study and they are correlated with important unobserved roadway geometrics, such as sight-distance and interchange density, which can influence the likelihood of rear-end crashes. The posted speed limits therefore capture unobserved roadway geometrics and speed effects. The model includes the variable "speed limit" that takes the actual value of the posted speed limit for the section.

In summary, the model includes six continuous explanatory variables: daily VMT per lane, truck percentage-mile-per-lane, deviation of shoulder width, curvature-per-length, off-ramp and merge, and speed limit; and two binary variables: merge section and urban area.

RESULTS

The rear-end crash risk model was estimated by maximum likelihood. In total, twelve coefficients (including intercepts) on eight explanatory variables were found statistically significant at the 90% level in the model (five explanatory variables for P_o, five explanatory variables for P_f, and the reciprocal of the negative binomial dispersion parameter, θ). Model estimation results are shown in Table 1. The sign of an estimated coefficient indicates the direction of the impact of the variable, i.e. a variable with a positive coefficient increases the probability and a variable with a negative coefficient has a decreasing effect. The ρ^2 in this paper compares the log-likelihood at ($\boldsymbol{\beta} = 0$, $\kappa = 1$) to log-likelihood at convergence.

The Probability of a Leading Vehicle becoming an Obstacle (P_o)

Two variables were found to decrease the probability of a leading vehicle becoming an obstacle and four variables were found to increase the probability. The "daily VMT per lane" tends to decrease the probability of a leading vehicle becoming an obstacle. This is somewhat counterintuitive as higher volumes suggest greater opportunities for crashes. However, with increasing flow, traffic is compacted and more vehicles enter into car-following mode, resulting in increasingly similar speeds on the freeway. Importantly, our study focuses on rear-end crashes between two vehicles and omits chain-reaction rear-end crashes. Chain-reaction crashes become more likely with higher volume and the relative number of two-vehicle crashes will drop and cause a negative relationship with increasing volume. There may also be non-linear effects in this variable which are not captured by the model. These effects may contribute to the reduced probability of a vehicle becoming an obstacle with higher volumes. It should be noted, that the net effect from the model does indicate that there is a higher probability of rear-end crashes with higher volumes as expected. That happens because the probability of failing to avoid a crash goes up with increasing volume.

Truck percentage-mile-per-lane was found to increase the probability. Examples that could explain this are as follows: (a) when a leading vehicle is a truck, the following driver may be more likely to switch lanes and overtake the truck due to the relatively slow speed of the truck, and (b) a passenger car sometimes cuts in front of a truck without allowing sufficient headway for a following truck, ignoring the fact that a truck needs a longer headway than a passenger car. Therefore, a higher truck percentage results in more frequent lane changes and such disturbances could contribute to an increase in P_o. Golob and Reagan's study (2004) indicates this tendency. As the number of vehicles increases, lane changes may be difficult and drivers may stay in their current lanes.

Freeway sections in an urban area are associated with a higher probability of a vehicle encountering an obstacle vehicle. This may be due to the higher density of entrances and exits that create more frequent lane changes (weaving). This reasoning can be supported by Golob et al. (2004) who found that rear-end crashes have the highest likelihood of occurring in a weaving section.

The degree of curvature is directly related to the radius (R) of the horizontal curve. Therefore, as R decreases, curvature increases. Carson and Mannering (2001) found that crash frequency decreases as horizontal curve radius increases. As shown in Table 1, "curvature-per-length" is identified to have an increasing impact on the probability of a lead vehicle becoming an obstacle.

The "off-ramp and merge" variable increases the probability of the lead vehicle becoming an obstacle. When vehicles' lane change frequency increases, the likelihood of having a rear-end crash grows higher. Jason et al. (1998) found that rear-end crashes involving trucks are more likely to occur in sections with off-ramps than in sections with on-ramps. An "on-ramp" variable was originally included in the model, but removed from the final form because it was not significant. This may indicate that different types of crashes, such as sideswipe, are more frequent than rear-end crashes near on-ramps.

The Probability of Failure to Avoid a Collision given an Obstacle Vehicle (P_f)

Three variables were found to decrease the rear-end crash probability and three variables were found to increase its probability. In the P_f model, "daily VMT per lane" has an increasing impact and "truck percentage-mile-per-lane" has a decreasing impact. Obviously, the impacts of these two variables are opposite to their effects in the P_o model. As "daily VMT per lane" increases, the traffic density increases and the increase of traffic density means the decrease of headway distance if all other conditions are the same. As headway distance decreases, AMT decreases, and as a result, a driver's probability of failure to avoid a collision increases. When following a truck, drivers tend to keep longer gaps. Also, truck drivers are professional drivers. Therefore, the increase of the percent truck means an increase in the number of professional drivers in the traffic stream, and they may be better able to respond to avoid crashes than regular drivers. This tendency results in longer AMT and lowers P_f. This reasoning can be supported by Golob and Reagan (2004). In their study, 45% of crashes not involving trucks were rear-end crashes, whereas only 18% of truck-involved crashes were rear-end crashes.

The P_f model also found that road sections with a higher posted speed limit have lower driver failure rate. This is may be due to the correlations between posted speed limits and unobserved factors such as travel speed, design speed, and roadway geometrics. For example, roadway sections with high posted speed limit have greater stopping sight-distance; other factors, such as reduced frequency of interchanges on sections with a higher speed limit, would reduce weaving maneuvers which can reduce rear-end crash frequencies. Golob and Recker (2003) drew a similar conclusion from their study: rear-end crashes are more likely to occur at lower speeds and during higher variations of speed.

Another variable which increases P_f is "deviation of shoulder width." It has been found that a narrow shoulder width (total shoulder width is smaller than 5.5 m (18 feet) increases the probability of a driver's failure to avoid a collision. On sections with narrow shoulders, drivers have less room to avoid rear-end crashes or take corrective actions, which may explain this result. Milton and Mannering (1998) also concluded that narrow shoulders (including both the right and left shoulders) tend to increase crash frequency.

The "merge section" variable also increases P_f. This indicates that a driver has a greater P_f when driving in a merge section. This can be explained by two reasons: 1) cut-in vehicles significantly reduce AMT; and 2) other vehicles' movements can distract a driver's attention which delays the perception of an obstacle vehicle.

Finally, the t-statistic of the coefficient estimate for the reciprocal of the negative binomial dispersion parameter (κ) was 13.918, which means that this coefficient was statistically very significant, and that it was correct to reject the Poisson model.

The average of the probability of encountering an obstacle vehicle (P_o) was 32.88% and the average of the probability of the following vehicle driver's failure to avoid a collision (P_f) was 0.001158%. This result is consistent with Wang et al. (2002). They reasoned that while traffic flow is frequently interrupted by disturbances, the drivers' AMT is generally greater than NMT and hence allow the appropriate perception and reaction time to accomplish an avoidance maneuver.

Elasticity

Two variables in this model have dual impacts with opposite directions on the rear-end crash risk: daily VMT per lane and truck percentage-mile-per-lane. To know the overall effects on probability of rear-end crashes, elasticity was calculated for those variables. Note that elasticity was calculated for Equation (16), the probability of a rear-end crash occurring, but not Equation (24) the probability of a certain number of rear-end crashes occurring in a section. The elasticities of daily VMT per lane and truck percentage-mile-per-lane were about 0.102 and -0.542 , respectively. That is, as daily VMT per lane increases, the probability of rear-end crashes increases (0.102); as truck percentage-mile-per-lane increases, the likelihood of rear-end crashes decreases (-0.542).

Statistical Tests of Temporal Transferability and Coefficient Stability

We statistically tested the model for temporal transferability and coefficient stability. Table 2 shows the results of the temporal transferability and coefficient stability tests. For the temporal transferability test, the null hypothesis is that the coefficients are transferable between years. We first estimate the model for the two years (2001 and 2002) together, effectively constraining the coefficients to be equal for both years. Then, we estimate the model for the years individually using the same model structure and apply a likelihood ratio test to compare the constrained model to the two unconstrained models. The likelihood ratio test results indicate a χ^2 value of 13.36 with 13 degrees of freedom, which is smaller than the table χ^2 value, 22.3621, at the 95% confidence level. We therefore do not find statistical evidence to reject the null hypothesis of transferability, since allowing the coefficients to be different did not result in a significant change compared to the constrained model.

In the second test, for coefficient stability, a similar likelihood ratio test is performed to test the null hypothesis of coefficient stability. The model was estimated for the entire two-year period. Then, with a uniform distribution, the total data were randomly divided into two sub data sets having the same number of observations. The likelihood ratio test results indicate a χ^2 value of 12.31 with 13 degrees of freedom, which is smaller than the χ^2 table value, 22.3621, at the 95% confidence level. There is therefore no statistical evidence to reject the null hypothesis of stability and the model specification can be used for sub data sets.

CONCLUSION

Following the microscopic modeling approach developed by Wang (1998), a rear-end crash risk model was developed and estimated using freeway rear-end crash data observed in Washington State. Unlike most existing crash models, the model developed in this research considered the occurrence mechanism of rear-end crashes on freeways and was capable of capturing the dual impacts of explanatory variables in the occurrence of rear-end crashes.

When interpreting a model with dual effects, the effects can contradict each other. In that case it is necessary to draw the interpretation from the overall model result which will show which effect is stronger. Most often, the results are in harmony and it is not necessary to look to the overall model results, since the individual models directly yield useful interpretations that can lead to safety improvements.

For example, the "daily VMT per lane" variable has dual impacts with opposite directions: It reduces the probability of a vehicle becoming an obstacle (P_o) as indicated by the negative coefficient, and it increases the probability of a following vehicle's reaction failure (P_f), yielding a net increase in probability of rear-end crash with volume as indicated by the average elasticity for the overall model. The dual process in this model therefore suggests the increased probability of rear-end crashes with volume happens because of the increasing probability of drivers failing to avoid crashes but not because of an increase in probability of a vehicle becoming an obstacle vehicle. This indicates that focusing safety improvement efforts on reducing drivers' failure to avoid crashes is of key importance. Potential applications could be crash avoidance systems that assist drivers to avoid crashes, e.g., headway warning systems and smart cruise controls that reduce the PRT significantly.

The "truck percentage-mile-per-lane" variable was also significant in both P_o and P_f. A higher truck percentage increases P_o but decreases P_f. When considering the overall model, this variable reduces the probability of rear-end crashes.

Understanding such dual impacts of controllable variables is important for selecting safety improvement plans. This finding can for example be used to improve safety on highways through the use of information systems. Although truck percentage-mile-per-lane decreases the probability of rear-end crashes under simultaneous consideration of both P_o and P_f, this variable increases P_o. Therefore, real-time information of truck percentage-mile-per-lane may be used in warning systems that inform drivers when traffic conditions are more likely to lead to rear-end crashes.

This model mainly focused on the mechanism of freeway rear-end crash occurrence. Traditional human factors such as age, experience, health condition, and gender play an important role in the mechanism but individual-specific data cannot be used in a frequency model because crash information must be aggregated over a time period in each section. Therefore, a distribution of drivers' response time was employed as a surrogate variable for reflecting the impacts of human factors.

A modified negative binomial regression approach was employed to calibrate the risk model using observed rear-end crash data and was successfully estimated by maximum likelihood. The estimated negative binomial distribution parameter was found statistically significant, which indicates the data was overdispersed, and that the Poisson model would have been less appropriate.

For future study, we recommend the use of micro-scale traffic data, such as 5 minute volumes at the time of each crash to better explain the effect of traffic flow on rear-end crashes in a microscopic model.

In summary, this study demonstrated that a microscopic modeling approach can be applied to freeway rear-end crashes and it produced reasonable results. This type of microscopic crash frequency modeling adds to the understanding of the relationships between the risks of freeway rear-end crashes and causal factors. It can also help decision-makers select effective countermeasures against freeway rear-end crashes, especially in the realm of design of roadways (e.g., roadways can be designed with certain shoulder widths and less curvature).

ACKNOWLEDGEMENTS

The authors thank the Washington State Department of Transportation and the Highway Safety Information System, Federal Highway Administration for their help in providing the data used in this study. The authors also thank the anonymous reviewers, whose constructive comments significantly improved the paper.

REFERENCES

AASHTO (2001). *A policy on Geometric Design of Highways and Streets*, American Association of State Highway and Transportation Officials, Washington, D.C.

Bates, J. T. (1995). "Perception-Reaction Time." *ITE Journal*, 65(2), 35–36.

Carson, J., and Mannering, F. (2001). "The effect of ice warning sings on ice-accident frequencies and severities." *Accident Analysis and Prevention*, 33, 99–109.

FHWA. (2003). *Manual on Uniform Traffic Control Devices (MUTCD)*, Federal Highway Administration, U.S. Department of Transportation.

Golob, T. F., and Reagan, A. C. (2004). *Traffic conditions and truck accidents on urban freeways*, University of California at Irvine, UCI-ITS-WP-04-3.

Golob, T. F., and Recker, W. W. (2003). "Relationships among urban freeway accidents, traffic flow, weather, and lighting conditions." *Journal of Transportation Engineering*, 129(4), 342–353.

Golob, T. F., Recker, W. W., and Alvarez, V. M. (2004). "Safety aspects of freeway weaving sections." *Transportation Research Part A*, 38, 35–51.

Greene, W. H. (2003). *Econometric Analysis*, 5th Ed., Prentice Hall, New York.

Jason, B., Awed, W., Robles, J., Kononov, J., and Pinkerton, B. (1998). "Truck Accidents at Freeway Ramps: Data Analysis and High-Risk Site Identification." *Transportation and Statistics*, 76–92.

Jovanis, P., and Chang, H. (1986). "Modeling the relationship of accidents to miles traveled." *Transportation Research Record 1068*, 42–51.

Lee, J., and Mannering, F. (2002). "Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis." *Accident Analysis and Prevention, 34,* 149–161.

Lerner, N. D. (1993). "Brake perception-reaction times of older and younger drivers." *Human Factors and Ergonomics Society*, 1, 206–210.

Lindley, J. (1987). "Urban freeway congestion: quantification of the problem and effectiveness of potential solutions." *ITE Journal*, 57, 27–32.

Mannering, F. L., Kilareski, W. P., Washburn, S. S. (2005). *Principles of Highway Engineering and Traffic Analysis*, 3rd Ed., John Wiley & Sons, New York.

Massie, D. L., Campbell, K. L., and Blower, D. F. (1993). "Development of a collision typology for evaluation of collision avoidance strategies." *Accident Analysis and Prevention*, 25(3), 241–257.

Miaou, S. P. (1994). “The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions.” *Accident Analysis and Prevention*, 26(4), 471–482.

Miaou, S. P., Hu, P., Wright, T., Rathi, A., and Davis, S. (1992). “Relationships between Truck Accidents and Highway Geometric Design: A Poisson Regression Approach.” *Transportation Research Record 1376*, 10–18.

Milton, J., and Mannering, F. (1998). “The relationship among highway geometrics, traffic related elements and motor-vehicle accident frequencies.” *Transportation*, 25, 395–413.

Olson, P. L., and Sivak, M. (1986). “Perception-response time to unexpected roadway hazard.” *Human Factors*, 28(1), 91–96.

Plait, A. (1962). “The Weibull distribution.” *Industrial Quality Control*, 17–26.

Poch, M., and Mannering, F. (1996). “Negative binomial analysis of intersection-accident frequencies.” *Journal of Transportation Engineering*, 122(2), 105–113.

Shankar, V., Mannering, F., and Barfield, W. (1995). “Effect of roadway geometrics and environmental factors on rural freeway accident frequencies.” *Accident Analysis and Prevention*, 27(3), 371–389.

Shankar, V., Milton, V., and Mannering, F. (1997). “Modeling accident frequencies as zero-altered probability process: an empirical inquiry.” *Accident Analysis and Prevention*, 29(6), 829–837.

Schrank, D. and Lomax, T. (2003). *The 2003 Urban Mobility Report*, Texas Transportation Institute, The Texas A&M University System, <http://mobility.tamu.edu/> (Sep. 15, 2004).

Summala, H. (2000). “Brake Reaction Times and Driver Behavior Analysis.” *Transportation Human Factors*, 2(3), 217–226.

Ulfarsson, G. F., and Shankar, V. (2003). “An accident count model based on multi-year cross-sectional roadway data with serial correlation.” *Transportation Research Record 1840*, 193–197.

Wang, Y. (1998) “Modeling Vehicle-to-Vehicle Accident Risks Considering the Occurrence Mechanism at Four-Legged Signalized Intersections.” Ph.D. Dissertation, The University of Tokyo.

Wang, Y., and Nihan, N. (2003). “Quantitative analysis on angle-accident risk at signalized intersections.” *World Transport Research, Selected Proceedings of the 9th World Conference on Transport Research.* Pergamon Press, Oct. 2003, Seoul, Korea.

Wang, Y., Ieda, H., and Mannering, F. (2002). “Estimating Rear-End Accident Probability at Signalized Intersections: An Occurrence-Mechanism Approach.” *Journal of Transportation Engineering*, 129(4), 1–8.

Washington State Department of Transportation. (1996). *1996 Washington State highway Accident Report*, <http://www.wsdot.wa.gov/mapsdata/tdo/> (Sep. 15, 2004).

NOTATION

The following symbols are used in this paper:

$E[\cdot]$ = expected value

e = exponential function

f = function

$g(\cdot)$ = density function of a gamma distribution

h = hazard function

I = total number of freeway sections

J = maximum number of disturbances

$L(\cdot)$ = log-likelihood function

m = Poisson distribution parameter

N = number of crashes

n = number of crashes on a given section in a given time period

P = probability

p = probability value or p-value

T = number of years of crash data

t = time

$V[\cdot]$ = variance

v = vehicle flow

$\mathbf{x}$ = vector of explanatory variables for a single observation

α = shape parameter of the Weibull distribution

$\boldsymbol{\beta}$ = vector of estimable parameters

$\Gamma(\cdot)$ = gamma function

δ = variance of gamma distributed error term

ε = unobserved error term

η = parameter of the exponential distribution

θ = scale parameter of the Weibull distribution for Available Maneuvering Time

κ = dispersion parameter of the negative binomial model

λ = scale parameter of the Weibull distribution for Needed Maneuvering Time

ρ^2 = rho-squared statistic

χ^2 = chi-squared distributed likelihood ratio test statistic

The following subscripts are used in this paper:

a = available maneuvering time

f = following vehicle failure

i = freeway section

j = disturbance type or given period

n = needed maneuvering time

o = obstacle vehicle

TABLE 1 Model Estimation Results

Variable	Estimated Coefficient	t-statistic
Variables affecting the probability of becoming an obstacle vehicle (P_o)		
Constant	-1.158 (0.483)	-2.397
Daily VMT per lane	-0.581 (0.107)*	-5.421
Truck percentage-mile-per-lane	0.771 (0.139)*	5.552
Urban area	0.695 (0.133)*	5.237
Curvature per length	0.019 (0.011)	1.651
Off-ramp and merge	0.190 (0.105)	1.811
Variables affecting the probability of following vehicle's driver failure (P_f)		
Constant	-8.239 (0.749)*	-11.000
Daily VMT per lane	0.552 (0.113)*	4.875
Truck percentage-mile-per-lane	-0.779 (0.146)*	-5.329
Speed limit	-0.103 (0.009)*	-11.069
Deviation of shoulder width	0.040 (0.006)*	6.824
Merge section	0.540 (0.100)*	5.415
Reciprocal of negative binomial dispersion parameter (κ)	0.888 (0.064)*	13.918
Log-likelihood at zero ($\boldsymbol{\beta}$ = zero, κ = 1)		-63,984.570
Log-likelihood at κ only ($\boldsymbol{\beta}$ = 0)		-6,536.126
Log-likelihood at constants and κ (other $\boldsymbol{\beta}$ = 0)		-3,826.912
Log-likelihood at convergence		-3,484.632
ρ^2		0.946

Standard errors are in parentheses.
Level of significance: all greater than 90% and * > 99.9%. Coefficients that weren't significant at the 90% level were restricted to zero and omitted from the table.
ρ^2 was calculated by comparing the log-likelihood at zero ($\boldsymbol{\beta}$ = zero, κ = 1) to log-likelihood at convergence.

TABLE 2 Transferability and Stability Test Results

Temporal transferability Test	
LL ($\boldsymbol{\beta}_t$)	= -3484.63
LL ($\boldsymbol{\beta}_a$)	= -1781.94
LL ($\boldsymbol{\beta}_b$)	= -1696.01
χ^2	= -2(LL ($\boldsymbol{\beta}_t$)- LL ($\boldsymbol{\beta}_a$)- LL ($\boldsymbol{\beta}_b$)) = -2(-3484.63-(-1781.94)-(-1696.01)) = 13.36 (13.36 < 22.3621, with 95% confidence and 13 degrees of freedom) Transferable
Coefficient Stability Test	
LL ($\boldsymbol{\beta}_t$)	= -3484.63
LL ($\boldsymbol{\beta}_a$)	= -1819.27
LL ($\boldsymbol{\beta}_b$)	= -1659.21
χ^2	= -2(LL ($\boldsymbol{\beta}_t$)- LL ($\boldsymbol{\beta}_a$)- LL ($\boldsymbol{\beta}_b$)) = -2(-3484.63-(-1819.27)-(-1659.21)) = 12.31 (12.31< 22.3621, with 95% confidence and 13 degrees of freedom) Transferable