PERGAMON

Accident Analysis and Prevention 36 (2004) 183–191

ACCIDENT ANALYSIS & PREVENTION

www.elsevier.com/locate/aap

Selecting exposure measures in crash rate prediction for two-lane highway segments

Xiao Qin[a,*], John N. Ivan[b,1], Nalini Ravishanker[c,2]

[a] Maricopa Association of Governments, 302 North 1st Avenue Ste. 300, Phoenix, AZ 85003, USA
[b] Civil and Environmental Engineering, Connecticut Transportation Institute, University of Connecticut, U-2037, Storrs, CT 06269, USA
[c] Department of Statistics, University of Connecticut, Unit 4120, Storrs, CT 06269, USA

Received 19 July 2002; accepted 20 November 2002

Abstract

A critical part of any risk assessment is identifying how to represent exposure to the risk involved. Recent research shows that the relationship between crash count and traffic volume is non-linear; consequently, a simple crash rate computed as the ratio of crash count to volume is not proper for comparing the safety of sites with different traffic volumes. To solve this problem, we describe a new approach for relating traffic volume and crash incidence. Specifically, we disaggregate crashes into four types: (1) single-vehicle, (2) multi-vehicle same direction, (3) multi-vehicle opposite direction, and (4) multi-vehicle intersecting, and define candidate exposure measures for each that we hypothesize will be linear with respect to each crash type.

This paper describes initial investigation using crash and physical characteristics data for highway segments in Michigan from the Highway Safety Information System (HSIS). We use zero-inflated-Poisson (ZIP) modeling to estimate models for predicting counts for each of the above crash types as a function of the daily volume, segment length, speed limit and roadway width. We found that the relationship between crashes and the daily volume (AADT) is non-linear and varies by crash type, and is significantly different from the relationship between crashes and segment length for all crash types. Our research will provide information to improve accuracy of crash predictions and, thus, facilitate more meaningful comparison of the safety record of seemingly similar highway locations.

Keywords: Exposure; Highway safety; Two-lane rural highway; Zero-inflated-Poisson; Hypothesis test; Crash type

1. Introduction

Estimating the safety performance of an existing or planned roadway is an important task in highway safety management. To do this, highway crash prediction models have been developed to explain uncertainty in the occurrence of crashes, otherwise defined as risk. The purpose of the exposure is to estimate the potential probability for a crash occurring. In practical applications, exposure can vary according to the mode or context being analyzed. For example, the US Coast Guard (USCG) maritime safety program divides the number of incidents by the number of workers, thus measuring occupational safety to enable comparison with other employment-related contexts. On the other hand, commercial aviation safety programs use flight hours or the number of take-offs or landings as the denominator to instead measure an incident rate (Bureau of Transportation Statistics, 2000). Carrying this analogy to highway safety studies, the exposure measure is applied to make a safety risk comparison of the two sites reliable or meaningful. Otherwise, without the exposure measure, the varying spatial-temporal contexts would make the comparison of site safety difficult. In other words, it may mask important risk factors, and mislead people's understanding of the safety relative risk associated with different situations.

The reliability and efficiency of the exposure measured by vehicle-miles traveled (VMT), the number of entering vehicles (NEV) and so on, are doubtful without clear understanding of the relationship between safety and exposure. Since many of the safety evaluations and comparisons are made based on risk, these risk-oriented crash rates are useful only when the relationship between the number of crashes and exposure is linear (Hauer, 1995). In order to explore the relationship, our study focuses on defining crash exposure measures at a more disaggregate level, specifically by crash type using corresponding exposure functions.

* Corresponding author. Tel.: +1-602-254-6300; fax: +1-602-254-6490.
E-mail addresses: xqin@mag.maricopa.gov (X. Qin), johnivan@engr.uconn.edu (J.N. Ivan), nalini@stat.uconn.edu (N. Ravishanker).
[1] Tel.: +1-860-486-0352; fax: +1-860-486-2298.
[2] Tel.: +1-860-486-4760; fax: +1-860-486-4113.

0001-4575/$ – see front matter © 2003 Elsevier Science Ltd. All rights reserved.
doi:10.1016/S0001-4575(02)00148-3

2. Background

A common way to define the safety of a system (usually measured by expected number of crashes) is the product of the probability of having a crash (also called crash risk) given a unit of exposure and the observed level of exposure. Since the number of crashes is the only self-evident quantity in the equation, the resulting crash risk per unit exposure is determined by the selection of exposure measures and vice versa. In other words, these quantities are completely dependent upon one another. Hence, it is vital for the exposure measures to be specified or quantified. Hauer (1982) adopts the definition of "a unit of exposure" and explains it as a trial. The result of such a trial is the occurrence or non-occurrence of an accident (by type, severity and so on.). However, this exposure measure is oriented to the entity (driver or vehicle) involved, e.g. one truck trip or one pedestrian crossing. If it is applied in a site-specified situation, such as a road segment or intersection, the definition is still obscure and the corresponding exposure is difficult to measure. However, on the other hand, other researchers have different definitions about the exposure conveying similar meanings. Stewart (1998) thinks of exposure to risk as a statistical measure providing information on the extent of a road user's exposure to the overall level of travel risk given the road conditions at any point in time and he recommends "kilometers of travel" as a "meaningful, practical and applicable" measure of exposure. Carroll (1971) defines driving exposure as the "frequency of traffic events which create a risk of accident". Chapman (1973) describes the exposure as the number of opportunities for accidents of a certain type in a given time in a given area. These opportunities are occasions when cars cross each others path, when they are following each other, or even when a vehicle is traveling by itself on a winding road. This definition considers that the exposure should include characteristics of drivers and vehicles, characteristics of the roadway, and the environmental condition. These studies indicate that both exposure and risk may depend on some of the same factors, such as traffic volume, time of day, weather and so on. In other words, all of these factors interact with one another. Naturally, some researchers test the same covariates as both exposure and risk factors. Miaou (1994) uses AADT per lane as both exposure and a traffic condition predictor variable in the risk function and finds that both are significant. However, how to use the information consistently and correctly in the safety prediction model is still under investigation.

So far, many of the crash prediction models are macroscopic form, considering all crashes together rather than separately by some stratification variable or variables such as crash type. In these models, the relationship between the total number of crashes and exposure may be ambiguous because, in fact, the opportunity for the occurrence of different crash types is different under the same exposure situation. Hauer et al. (1996) found it best to relate crashes to the actual volumes to which the two colliding vehicles belong, similar to the idea described in the previous section for intersection crashes. He estimated separate crash prediction models for fifteen different possible crash patterns at signalized intersections using the traffic flow relevant to the crash pattern to investigate how crashes depend on the contributory traffic flows. These crash patterns are defined by the maneuvers of the two vehicles before the collision. Similarly, Brown (1981) performed a study, simplifying the number of crash patterns into nine types for predicting crash potential at a four-leg intersection with two-way flow on each leg and a traffic signal.

These studies indicate there is indeed a relationship between the number of crashes and the traffic volume though its exact form is still unknown, and probably depends on the crash type. Consequently, in our study, we still regard exposure as the number of potential opportunities for crashes to occur and consider exposure using the following crash types on rural two-lane highways: (1) single-vehicle crashes (SV), (2) multi-vehicle same direction crashes (SD), (3) multi-vehicle opposite direction crashes (OD), and (4) multi-vehicle intersecting direction crashes (ID). We estimate crash prediction models using a variety of different crash types, and evaluate them to identify the best set of exposure measures for crash prediction.

3. Methodology approach

3.1. Objective

According to previous studies, the safety performance function (the relationship between number of crashes and exposure) is non-linear when AADT is applied as exposure; that is, crashes increase with the traffic volume in a non-linear fashion. Consequently, the crash rate (ratio of crashes to AADT) is not constant with respect to traffic volume even at the same location, so this rate should not be regarded as a measure of the site safety. Fig. 1 depicts this with a hypothetical safety performance function of a site at different volumes with no other treatments. For any point on the safety performance function curve, a relationship between the number of crashes and the exposure during a specified period of time, the crash rate (N/AADT) is defined as the slope of the line joining the origin to that point.

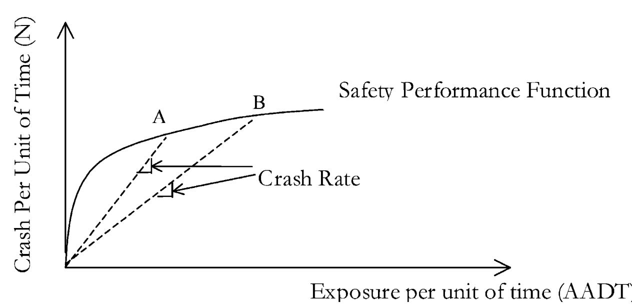

Fig. 1. Relationship between crashes and exposure.

Therefore, if the safety performance function is not a straight line, the crash rate varies with the amount of the exposure. For example, the number of crashes at point B is greater than at point A, but the crash rate at point B, conversely, is smaller than at point A, because the slope of the line joining it to the origin is less steep. From the point of view of highway safety engineers, this crash rate change due only to a change in exposure should not be regarded as an improvement in the site safety because there is no change in the physical characteristics of the site. It seems that a crash rate that ignores the shape of the safety performance function is not appropriate to compare the safety of the physical characteristics of one site to those of another.

Instead it is helpful to define a new exposure function to transform hourly traffic volume into an exposure measure *f*(*V*), forming a linear safety performance function. This exposure function could vary by time of day, land use and regional driver population effect. After implementing the new coordinate, the safety performance function becomes linear and each point on the line has the same slope, representing a normalized crash rate that is constant for all levels of exposure at the same location (Fig. 2). Therefore, this newly defined crash propensity is more meaningful to make comparisons among different entities with different exposures and safety performance functions.

3.2. Crash prediction model

In keeping with the non-linear relationship for the safety performance function, we suggest an exponential form for estimating disaggregate safety indexes for each crash type ρ_{ikt} as follows:

$$\mu_{ikt} = \eta_{ikt}\rho_{ikt} \tag{1}$$

where μ_{ikt} is the expected number of crashes for crash type *k* in site *i* at time *t*, η_{ikt} the computed exposure function of hourly volumes at site *i* for potential crash conflict type *k* at time of day *t*, and ρ_{ikt} is the normalized crash rate of crash type *k* for location *i* at time *t*, also defined as the safety index.

Further, we define functions for η_{ikt} and ρ_{ikt} as follows:

$$\eta_{ikt} = \eta_{kt}(\boldsymbol{V}_{it}, L_i) \tag{2}$$

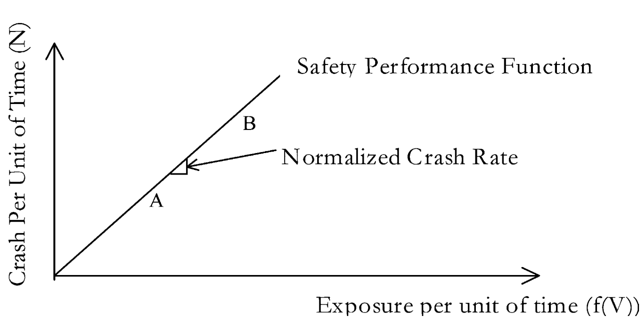

Fig. 2. Relationship between crashes and new exposure function.

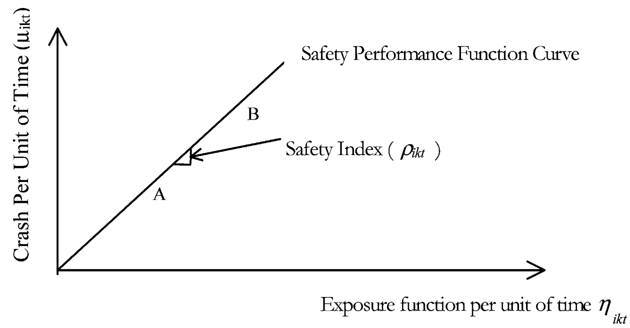

Fig. 3. Relationship between crashes, exposure and safety index.

and

$$\rho_{ikt} = \exp(\boldsymbol{X}_i\boldsymbol{\beta}_{kt}) \tag{3}$$

where $\boldsymbol{V}_{it}$ is the traffic volume by direction on site *i* at time of day *t*, L_i the length of road segment associated with site *i*, $\boldsymbol{X}_i$ the set of the road characteristics for site *i*, and $\boldsymbol{\beta}_{kt}$ is the parameters to be estimated for road characteristic $\boldsymbol{X}$ for crash type *k* and time *t*.

After all the necessary parameters are identified, Fig. 3 shows that the expected safety performance function of crash type *k* for site *i* is a straight line with the slope equal to the safety index ρ_{ikt} at time *t*.

3.3. Exposure function by crash type

In order to simplify the formula, we use η_{ikt} to represent $\eta_{kt}(\boldsymbol{V}_{it})$ and hypothesize potential functions for each type of crash. In fact, we also will permit the function parameters to vary by time, however this dimension is left out here for brevity. The following variables are used in these equations.

Here, η_k is the exposure function for potential crash conflict type *k*, $k \in K$ (SV, SD, OD, ID), v_1 the hourly volume in one direction of the two-lane rural highway, v_2 the hourly volume in the other direction of the two-lane rural highway, α_k the exponent on traffic exposure to be estimated for crash type *k*, $k \in K$ (SV, SD, OD, ID).

(1) Single-vehicle crashes

$$\eta_{\mathrm{SV}} = (v_1 + v_2)^{\alpha_{V_{\mathrm{SV}}}} L^{\alpha_{L_{\mathrm{SV}}}} \tag{4}$$

or

$$\eta_{\mathrm{SV}} = (v_1^{\alpha_{V_{\mathrm{SV}}}} + v_2^{\alpha_{V_{\mathrm{SV}}}}) L^{\alpha_{L_{\mathrm{SV}}}} \tag{5}$$

The first one assumes that each entity (vehicle or driver) on the road segment has a potential opportunity to be in a crash and the crashes on the two directions of the road are not independent from each other. The second will require evaluating crashes in each direction separately, an additional disaggregation step. Since it is not clear whether or not opposing volumes should be added together, both approaches will be investigated.

(2) Multi-vehicle same direction crashes

$$\eta_{\mathrm{SD}} = (v_1 + v_2)^{\alpha_{V_{\mathrm{SD}}}} L^{\alpha_{L_{\mathrm{SD}}}} \tag{6}$$

or

$$\eta_{\mathrm{SD}} = (v_1^{\alpha_{V_{\mathrm{SD}}}} + v_2^{\alpha_{V_{\mathrm{SD}}}}) L^{\alpha_{L_{\mathrm{SD}}}} \tag{7}$$

The assumption is the same as for single-vehicle crashes.

(3) Multi-vehicle opposite direction crashes

$$\eta_{\mathrm{OD}} = (v_1 v_2)^{\alpha_{V_{\mathrm{OD}}}} L^{\alpha_{L_{\mathrm{OD}}}} \tag{8}$$

The assumption is that each vehicle on its own path has the opportunity to collide with a vehicle in the opposite direction, so the number of meetings is proportional to the product of the flows.

(4) Multi-vehicle intersecting crashes

$$\eta_{\mathrm{ID}} = [(v_1 + v_2) v_{\mathrm{c}}]^{\alpha_{V_{\mathrm{ID}}}} L^{\alpha_{L_{\mathrm{ID}}}} \tag{9}$$

The assumption is similar to multi-vehicle opposite direction crashes, except that the conflict flows are mainline traffic and minor road or driveway traffic, where v_{c} is the cumulative traffic volume of the minor intersections or driveways in the road segment because there are no major intersections in our dataset.

3.4. Statistical model

While previous studies have provided insight into the factors determining crash frequencies, it is important to realize that traditional application of the Poisson or negative Binomial distribution alone does not address the possibility that more than one underlying process may be influencing crash frequencies (Miaou, 1994; Vogt and Bared, 1998; Lambert, 1992). For instance, if the study segments are collected randomly, a preponderance of zero-crash observations will appear in the data because crashes are rare events. This over-representation of zero-crash observations in the data may erroneously suggest overdispersion in the data even though the Poisson distribution is actually otherwise correct. To account for the large probability "spike" at zero, P_i is used to represent the additional probability of segment i to have no crashes while $1 - P_i$ represents the probability that segment i follows the Poisson distribution. Assuming a Poisson distribution, the probability that a segment will have no crashes (apart from the additional spike) is $\mathrm{e}^{-\mu_i}$. The total probability of observing zero crashes consists of mixing these two probabilities together. The entire distribution is called the zero-inflated-Poisson distribution with the following probability density function (Miaou, 1994; Vogt and Bared, 1998; Lambert, 1992; Shankar et al., 1997).

$$P(N_i) = \begin{cases} P_i + (1 - P_i)\mathrm{e}^{-\mu_i} & (N_i = 0) \\ (1 - P_i)\dfrac{\mathrm{e}^{-\mu_i}\mu_i^{N_i}}{N_i!} & (\text{otherwise}) \end{cases} \tag{10}$$

Now, as defined previously

$$\mu_i = \eta(V_i, L)\mathrm{e}^{X_i\beta} \tag{11}$$

As for the probability of site i to be an inherently safe site P_i, we use a logit function as follows:

$$\mathrm{logit}(P_i) = \log\left(\frac{P_i}{1 - P_i}\right) = \boldsymbol{X}_i\boldsymbol{\gamma} \tag{12}$$

where $\boldsymbol{X}_i$ is a vector of covariance which may influence the value of P_i, and $\boldsymbol{\gamma}$ is a vector of coefficients of the covariates $\boldsymbol{X}_i$.

The maximum likelihood function with which we can estimate all the unknown coefficients is as follows:

$$L(\boldsymbol{\beta}, \boldsymbol{\gamma} | N_{ij}) = \prod_{i=0}^{n} [P_i + (1 - P_i)\mathrm{e}^{-\mu_i}] \left[(1 - P_i)\frac{\mathrm{e}^{\lambda_i}\mu_i^{N_{ij}}}{N_{ij}!}\right] \tag{13}$$

Because P_i represents the perfect state (no crash state), we assume that the same factors have effect on expected number of crashes μ_i and P_i and use the same covariates in their estimation formulas. Recall that each P_i, N_i, V_i and μ_i will be indexed on i, k and t. These indices are omitted here for brevity. Values for the parameters β and the function for η can be obtained using maximum likelihood estimation.

4. Preliminary study design

Gathering data to support estimation of parameters for the above defined models of crash experience and exposure proved to be more challenging than expected. Obviously, we require hourly traffic volume on both the main segment and intersecting roads and associated crash and road segment characteristics. Since these data were not readily available and required some time to assemble, we have in the meantime acquired data from the HSIS, a multi-state safety data base containing crash, roadway inventory, and AADT records for a selected group of states maintained by Federal Highway Administration (FHWA) for a preliminary study to demonstrate our proposed methodology and procedures.

For this preliminary study, we use the Michigan HSIS data, which have complete accident and roadway inventory databases where year-to-year changes on highway geometric features and traffic condition are recorded from 1994 to 1997. There are a total of 29,800 road segments in each year's dataset. The segments are selected randomly from two-lane rural highways in the state without major intersections. The segment length varies from 0.016 to 12.51 km, with an average of 0.992 km. The originally defined crash types have been re-categorized into the four groups defined previously. Descriptive statistics of these road segments are listed in Table 1. As covariates we include highway characteristics available in the dataset shown by previous researchers to be significant for predicting crashes on two-lane highways.

This change in study design also necessitated a modification in the exposure functions. Instead of the hourly volumes needed for these data we have only the two-way AADT

Table 1
variable definition and summary statistics of Michigan state data (1996)

Variable	Definition	Minimum	Maximum	Mean
Number of crashes	SV: single-vehicle crashes	0	61	0.68
	SD: multi-vehicle same direction crashes	0	23	0.15
	OD: multi-vehicle opposite direction crashes	7	0	0.04
	ID: multi-vehicle intersecting direction crashes	0	23	0.08
Segment length	L (km)	0.016	12.51	0.992
AADT (in 1000’s of vehicles)	V	0.24	40	5.45
Speed limit	SL (km/h)	40	88	84
Lane width	LW (m)	3.05	3.66	3.477
Right shoulder width	RTSHD (m)	0	3.66	2.65
Left shoulder width	LTSHD (m)	0	3.66	2.65
Pavement width	PW = RTSHD + LTSHD + 2 × LW (m)	6.1	13.42	12.19

observed on the segment. Consequently, we use the same exposure function for all crash types, without considering variation by time of day, and the t subscript is dropped, as the analysis is preformed without regard for that dimension. This modified exposure function is

$$\eta_{ik} = V_i^{\alpha_{V_k}} L_i^{\alpha_{L_k}} \tag{14}$$

5. Analysis and results

The objective of the preliminary study is to use ZIP modeling for identifying the contribution of the exposure function for the four crash types defined earlier. We choose shoulder width and lane width, speed limit as covariates in the models, AADT and segment length as two components in the exposure functions entered logarithmically so as to estimate exponents for each by crash type as indicated in Eq. (14). The estimated coefficient for each covariate and associated t-value are presented in Table 2 and these coefficients will be discussed later.

5.1. Comparing exponents of AADT and segment length by crash type

Four years data of Michigan are analyzed separately in order to avoid issues with variation over the years. The regression results show that the exponents for AADT and road segment length vary little by year but greatly from one crash type to another. The exponent on AADT is lowest (about 0.4) for single-vehicle crash prediction. In this type of curve, the single-vehicle marginal crash rate is high at low traffic volumes and low at high traffic volumes, suggesting that as volume increases, this crash type becomes less likely. This may be because the more common presence of other vehicles on the road results in more multi-vehicle crashes. The multi-vehicle same direction marginal crash rate, on the contrary, is low at low traffic volumes and high at high traffic volumes, quite the opposite of the observation for single-vehicle crashes. This may be because shorter time headways at high volumes increase the likelihood of this type of crash. Multi-vehicle opposite direction crashes have a nearly linear relationship with AADT. Note however, that when using AADT, we ignore the directional distribution of the traffic, which in fact is likely to be an important part of the exposure function for this crash type. Interestingly, the multi-vehicle intersecting direction crash exhibits similar behavior to the single-vehicle crashes, that is, the marginal crash rate for single-vehicle is high at low traffic volume and vice versa. In fact, these crashes are mostly associated with driveways, since there are only minor intersecting roads and driveways on the segments in our sample segments and there is no other way to experience this type of crash. Perhaps drivers entering from driveways become more cautious when the time headway between vehicles in the traffic stream becomes shorter, as happens at higher volumes. In any case, it seems clear that this exponent on AADT differs substantially from one crash type to the next, confirming our study design.

The exponent on road segment length, the other important factor in the exposure function, also varies by crash type. As can be seen in Table 2, all the exponents on segment length are positive and less than 1.0. The value is highest for single-vehicle crashes (close to 1.0) and lowest for multi-vehicle intersecting crashes. The relationship between single-vehicle crashes and length is nearly linear, possibly because the driver or the vehicle has the same opportunity to run off the road or collide with the roadside objects for each mile of the segment. On the other hand, multi-vehicle crashes depend on meeting another vehicle, which may not be linearly proportional to segment length. Moreover, with multi-vehicle intersecting crashes, the opportunity for such a crash does not depend so much on the segment length as on the number of the driveways or minor intersecting streets. This quantity was not available in the dataset but they may be important for further study with hourly counts.

Figs. 4(a) and 5(a) plot the simple relationship between crash and VMT from the 1996 dataset with the same site characteristics: speed limit 88 km/h, shoulder width 3.05 m and lane width of 3.66 m. These plots show the safety

Table 2
Estimated ZIP model regression parameters

Crash type	Covariate	1994		1995		1996		1997	
		Coefficient	t-Value	Coefficient	t-Value	Coefficient	t-Value	Coefficient	t-Value
Single-vehicle	Intercept	**−0.086**	**−0.296**	−0.672	−2.405	0.435	2.402	0.375	2.011
	ln(AADT)	0.302	17.096	0.254	14.540	0.305	23.299	0.363	25.781
	ln(segment length)	0.746	62.064	0.725	63.547	0.792	101.952	0.776	96.227
	Shoulder width	0.025	3.111	0.026	3.295	**−0.004**	**−0.864**	**−0.003**	**−0.542**
	Lane width	0.077	3.789	0.136	6.925	0.035	3.215	0.025	2.111
	Speed limit	**−0.002**	**−0.568**	**0.000**	**−0.132**	**0.005**	**1.796**	**0.005**	**1.872**
Multi-vehicle same direction	Intercept	1.298	22.785	2.133	3.226	**−0.141**	**−0.440**	**−0.589**	**−1.752**
	ln(AADT)	1.159	14.417	1.270	22.796	1.063	25.668	1.193	26.293
	ln(segment length)	0.442	5.713	0.429	13.815	0.407	22.289	0.416	21.628
	Shoulder width	−0.058	−2.739	−0.054	−4.702	−0.027	−3.642	−0.036	−4.497
	Lane width	−0.160	−3.807	−0.261	−4.729	**−0.022**	**−0.870**	**−0.007**	**−0.282**
	Speed limit	**−0.015**	**−1.642**	−0.013	−2.805	−0.018	−5.900	−0.018	−5.250
Multi-vehicle opposite direction	Intercept	8.989	4.543	8.882	6.188	**−0.115**	**−0.119**	−2.307	−2.064
	ln(AADT)	0.551	4.283	1.126	12.050	0.987	8.184	0.949	6.828
	ln(segment length)	0.565	5.090	0.748	11.759	0.354	7.196	0.520	9.692
	Shoulder width	−0.148	−5.306	−0.123	−5.393	−0.050	−2.235	−0.066	−2.471
	Lane width	−0.832	−4.964	−0.908	−7.469	**−0.056**	**−0.722**	**0.115**	**1.258**
	Speed limit	**−0.002**	**−0.165**	**−0.012**	**−1.300**	−0.026	−2.966	**−0.017**	**−1.593**
Multi-vehicle intersecting direction	Intercept	**1.081**	**1.113**	**0.883**	**0.857**	**0.666**	**1.381**	**−0.293**	**−0.667**
	ln(AADT)	0.807	11.808	0.720	10.470	0.757	12.613	0.650	11.337
	ln(segment length)	0.106	2.572	0.245	5.658	0.171	6.515	0.191	7.828
	Shoulder width	−0.049	−3.615	−0.064	−4.674	**−0.016**	**−1.587**	−0.044	−4.549
	Lane width	−0.208	−2.555	**−0.148**	**−1.715**	−0.103	−2.717	**−0.015**	**−0.448**
	Speed limit	**0.001**	**0.149**	**0.003**	**0.533**	−0.015	−3.344	**−0.005**	**−1.113**

Coefficients in bold are insignificant at 95%.

performance function, computed using the predicted crash rates for each case. As discussed earlier, for the plot with AADT, the safety performance function is a curve, and the apparent crash rate is not constant. For comparison, Figs. 4(b) and 5(b) plot the adjusted relationship using the exposure function instead of VMT. Due to our model definition, for the plot with the exposure function, the safety performance function becomes a straight line with the slope being the safety index or normalized crash rate that is constant for all levels of exposure for the same site characteristics. Single-vehicle crashes and multi-vehicle same direction crashes are displayed in the figures and the other two crash types: multi-vehicle opposite and multi-vehicle intersecting crashes show the similar patterns and are omitted in the interest of brevity. The plots show that the adjusted exposure function provides a good linear pattern for these data.

Table 3
Hypothesis test between AADT and exposure function

Crash type	Covariate	1994			1995			1996			1997		
		α	S_α	t^*	α	S_α	t^*	α	S_α	t^*	α	S_α	t^*
Single-vehicle	AADT	0.302	0.018	−39.522	0.254	0.017	−42.663	0.305	0.013	−53.179	0.363	0.014	−45.169
	Segment length	0.746	0.012	−21.078	0.725	0.011	−24.066	0.792	0.008	−26.822	0.776	0.008	−27.846
Multi-vehicle same direction	AADT	1.159	0.080	1.981	1.270	0.056	4.848	**1.063**	0.041	1.521	1.193	0.045	4.255
	Segment length	0.442	−0.077	7.213	0.429	0.031	−18.395	0.407	0.018	−32.419	0.416	0.019	−30.332
Multi-vehicle opposite direction	AADT	0.551	0.129	−3.485	**1.126**	0.093	1.350	**0.987**	0.121	−0.106	**0.949**	0.139	−0.370
	Segment length	0.565	0.111	−3.924	0.748	0.064	−3.963	0.354	0.049	−13.114	0.520	0.054	−8.963
Multi-vehicle intersecting direction	AADT	0.807	0.068	−2.826	0.720	0.069	−4.080	0.757	0.060	−4.058	0.650	0.057	−6.113
	Segment length	0.106	0.041	−21.654	0.245	0.043	−17.453	0.171	0.026	−31.627	0.191	0.024	−33.103

Values in bold represent hypothesis test 1 or 2 cannot be rejected at 5% level of significance; α: exponents of the AADT or segment length; S_α: standard deviation of the exponents of AADT or segment length; $t^* = (\alpha - 1)/S_\alpha$.

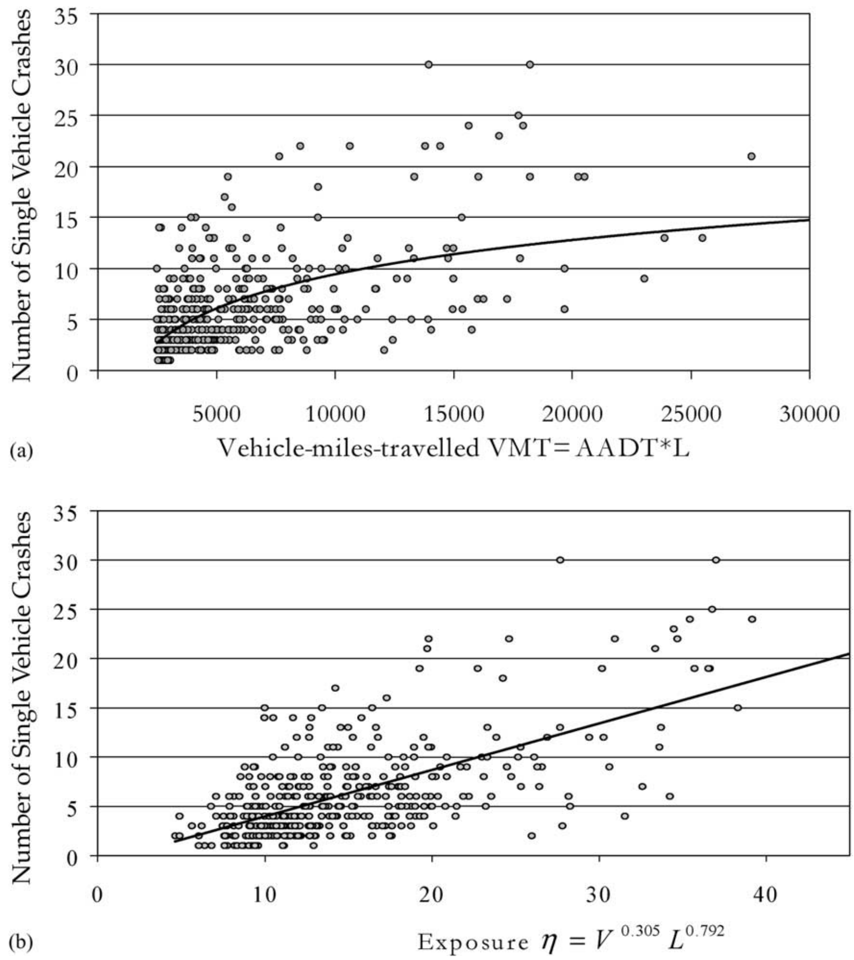

Fig. 4. (a) The observed and predicted crash counts vs. VMT and (b) estimated exposure function for crash type 1.

Table 4
Hypothesis test between VMT($L \times V$) and exposure function

Crash type	Covariate	1994			1995			1996			1997		
		α	S_α	$S_{V,L}$	α	S_α	$S_{V,L}$	α	S_α	$S_{V,L}$	α	S_α	$S_{V,L}$
Single-vehicle	AADT	0.302	0.018	0.134	0.254	0.017	0.103	0.305	0.013	0.066	0.363	0.014	0.048
	Segment length	0.746	0.012		0.725	0.011		0.792	0.008		0.776	0.008	
	t^*			−22.225			−23.721			−32.998			−25.929
Multi-vehicle same direction	AADT	1.159	0.080	0.077	1.270	0.056	0.071	1.063	0.041	0.060	1.193	0.045	0.063
	Segment length	0.442	−0.077		0.429	0.031		0.407	0.018		0.416	0.019	
	t^*			6.194			13.606			14.814			16.130
Multi-vehicle opposite direction	AADT	**0.551**	0.129	0.095	1.126	0.093	0.079	0.987	0.121	0.060	0.949	0.139	0.063
	Segment length	**0.565**	0.111		0.748	0.064		0.354	0.049		0.520	0.054	
	t^*			−0.883			3.476			4.963			2.944
Multi-vehicle intersecting direction	AADT	0.807	0.068	0.057	0.720	0.069	0.054	0.757	0.060	0.046	0.650	0.057	0.052
	Segment length	0.106	0.041		0.245	0.043		0.171	0.026		0.191	0.024	
	t^*			9.008			5.994			9.013			7.501

Values in bold represent hypothesis 3 cannot be rejected at 5% level of service; α: exponents of the AADT or segment length; S_α: standard deviation of the exponents of AADT or segment length; $S_{V,L}$: the correlation of coefficient of AADT and segment length; $t^* = (\alpha_L - \alpha_V)/\sqrt{S_{\alpha_L}^2 + S_{\alpha_L}^2 - 2S_{\alpha_L} S_{\alpha_V} S_{V,L}}$.

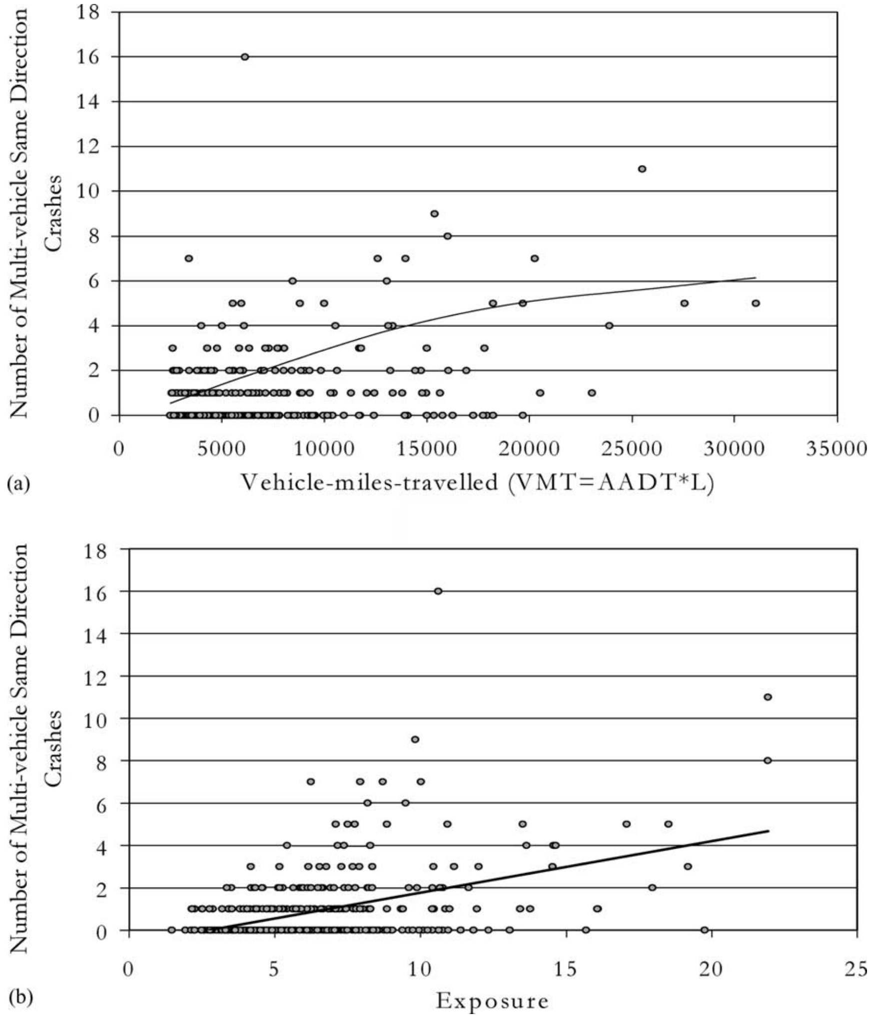

Fig. 5. (a) Observed and predicted crash counts vs. VMT (b) and estimated exposure function for crash type 2.

5.2. Testing hypotheses about exposure

We are interested in comparing the new exposure function with the previous exposure measures such as AADT and VMT. The exposure function can be written as $\eta = V^{\alpha_V} L^{\alpha_L}$, where V is the AADT and L is the segment length. We define three hypotheses to test whether or not the new exposure function, or parts of it, fit the observed data better than the traditional exposure measures. Following are the hypotheses with discussion about their meaning and test results.

The results of the three hypothesis tests are as follows.

Hypothesis test 1 (H_0: $\alpha_V = 1$; H_1: $\alpha_V \neq 1$). This hypothesis allows us to test whether or not the relationship between AADT and the crash count is linear ($\alpha_V = 1$). Table 3 gives the results of testing this hypothesis. The results show that the null hypothesis is rejected in most cases for all crash types except for multi-vehicle opposite direction crashes. Therefore, we can safely say the relationship between the number of crashes and AADT is non-linear for all but that crash type.

Hypothesis test 2 (H_0: $\alpha_L = 1$; H_1: $\alpha_L \neq 1$). This hypothesis is similar to the previous one, but allows us to test whether or not the relationship between segment length and crash count rather than AADT is linear. The results of testing this hypothesis are also given in Table 3; this hypothesis is rejected for all crash types and all years, meaning the number of crashes increases non-linearly with the road segment length.

Hypothesis test 3 (H_0: $\alpha_V = \alpha_L$; H_1: $\alpha_V \neq \alpha_L$). This hypothesis allows us to test whether or not the non-linear relationships between AADT and crash count and those between segment length and crash count are the same. Table 4 gives the results of this test and shows power inequality of AADT and segment length, meaning the assumption that they have the same exponents in the exposure function is doubtful.

6. Conclusion and future research

Highway safety is an issue of increasingly greater importance. Recent studies show that aggregate crash prediction models and exposure measures ignore significant variation in highway crashes contributing to prevailing hourly volumes and daily human biological cycles. The disaggregate approach used in this research reveals how the relationship between crashes and traffic volume varies from location to location, as well as for different crash types. We found for single-vehicle crashes that the marginal crash rate is high at low traffic volumes and low at high traffic volumes, probably because crashes are more likely to involve multiple vehicles at high traffic volumes. For multi-vehicle same direction crashes, the relationship is opposite: the marginal crash rate is low at low traffic volume and high at high traffic volumes, probably because this type of crash is more likely to occur under short time headways. The number of multi-vehicle opposite direction crashes increases nearly linearly with the traffic volume, while the multi-vehicle intersecting direction crashes exhibit similar behavior to the single-vehicle crashes. In fact, most multi-vehicle intersecting direction crashes on road segments are associated with vehicles entering from driveways, possibly explaining why the relationship is different from the other two multi-vehicle crash types. Besides AADT, road segment is important in exposure function, too. The relationship between number of crashes and segment length is not linear, but less than one. Among them, segment length has the least effect on the multi-vehicle intersecting direction crashes. It supports the notion that this type of crash is related more closely to the driveway density than the vehicle-miles-traveled.

Ongoing study on this project is estimating the exposure functions described in the study design section with newly collected traffic volume by direction. In addition, in this analysis, we are using hourly traffic volume including the time of day (divided into three categories). For intersecting direction crashes, we hope to collect observations of driveway density, or a surrogate for it, such as population density. The results may help us further understand the contribution of vehicle exposure to predicting crashes and provide new insight into developing common denominators for safety measures. Furthermore, the results will provide useful information for highway safety engineers and highway treatment enforcement to evaluate the safety of any location using either the number of crashes or the crash rate. Therefore, highway maintenance expenditure can be more efficiently assigned to more critical situations. Finally, this will facilitate more meaningful comparison of the safety record of seemingly similar highway locations and help to identify truly hazardous locations.

Acknowledgements

The research was funded by the Bureau of Transportation Statistics, and conducted at the Connecticut Transportation Institute. The data were provided by Dr. Yusuf Mohamedshah from Federal Highway Administration (FHWA), Highway Safety Information System (HSIS). The authors also gratefully acknowledge contributions to this project made by Dr. Don Tepas, Jianhe Du and Junfeng Liu. Thanks also go to Dr. Diane Lambert for providing ZIP model computer codes.

References

Brown, R.J., 1981. A method for determining the accident potential of an intersection. Traffic Eng. Contr. 22 (12), 648–651.

Bureau of Transportation Statistics, 2000. Safety in Numbers. Safety Data Action Plan Report. BTS, US.

Carroll, P.S., 1971. Techniques for the use of driving exposure information in highway safety research. HSRI, University of Michigan.

Chapman, R.A., 1973. The concept of exposure. Accid. Anal. Prev. 5 (2), 95–110.

Hauer, E., 1982. Traffic conflicts and exposure. Accid. Anal. Prev. 14 (5), 359–364.

Hauer, E., 1995. On exposure and accident rate. Traffic Eng. Contr. 36 (3), 134–138.

Hauer, E., Ng, J.C.N., Lovell, J., 1996. Estimation of safety at signalized intersection. Transport. Res. Board 1285, 42–51.

Miaou, S.P., 1994. The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. Accid. Anal. Prev. 26 (4), 471–482.

Lambert, D., 1992. Zero-inflated-Poisson regression, with an application to defects in manufacturing. Technometrics 34 (1), 1–14.

Shankar, V., Milton, J., Mannerring, F., 1997. Modeling accident frequencies as zero-altered probability processes: an empirical inquiry. Accid. Anal. Prev. 29 (6), 829–837.

Stewart, D.E., 1998. Methodological approaches for the estimation evaluation, interpretation and accuracy assessment of road travel 'basic risk', 'relative risk', and 'relative risk odds-ratio' performance measure. Indicators: a 'risk analysis and evaluation system model' for measuring, monitoring, comparing and evaluating the level(s) of safety on Canada's roads and highways. Transport Canada Publication No. TP 13238 E.

Vogt, A., Bared, J., 1998. Accident models for two-lane rural segments and intersections. Transport. Res. Board 1635, 18–29.