PERGAMON

Accident Analysis and Prevention 31 (1999) 695–704

ACCIDENT ANALYSIS & PREVENTION

www.elsevier.com/locate/aap

Differences in causality factors for single and multi-vehicle crashes on two-lane roads

John N. Ivan [a,*], Raghubhushan K. Pasupathy [b], Paul J. Ossenbruggen [c]

[a] *Department Civil and Environmental Engineering U-37, University of Connecticut, Connecticut Transportation Institute, Storrs, CT 06269-2037, USA*

[b] *Edwards and Kelcey, 529 Main Street, Suite 203, Charlestown, MA 02129, USA*

[c] *Department of Civil Engineering, University of New Hampshire, Durham, NH 03824, USA*

Received 1 October 1998; received in revised form 9 February 1999; accepted 26 March 1999

Abstract

Past research has found a non-linear relationship between traffic intensity or level of service (LOS) and highway crash rates. This paper investigates this relationship further by including the effects of site characteristics and estimating Poisson regression models for predicting single and multi-vehicle crashes separately. Analysis focuses on rural two-lane highways, with hourly LOS, traffic composition, and highway geometric characteristics as independent variables. The resulting models for single and multi-vehicle crashes have different explanatory variables. Single-vehicle crash rates decrease with increasing traffic intensity (lower LOS), shoulder width and sight distance. Multi-vehicle crash rates increase with the number of signals, the daily single-unit truck percentage, and the shoulder width, and decreased on principal arterials compared to other roadway classes. LOS does not significantly explain variation in the number of multi-vehicle crashes. Ongoing research by the authors is aimed at identifying other site factors, such as driveway density and intersection LOS, that can better explain the differing effects reported here and predict crash rates of both types better.

Keywords: Highway crash rates; Level of service; Crash type; Two-lane roads; Daylight

1. Introduction

One of many important aspects of highway safety research is developing crash prediction models to quantify the relationship between site characteristics and the number of crashes observed. These prediction models can help identify the site characteristics tending to cause highway crashes. Crash prediction models, till now, have tended to be macroscopic in nature, meaning that they predict using summary statistics on traffic such as annual average daily traffic (AADT) rather than microscopic measures such as hourly volume counts. These macroscopic models also tended to consider all crashes together rather than separately by single and multi-vehicle crash types. Studies that have considered hourly volumes have generally not considered geometric effects. In this paper, the significance of all of these factors—site location, highway geometry, daylight conditions, and traffic composition—are investigated together.

The focus of this research was to find differences in the effects of both traffic condition and site characteristics as independent variables for the expected single and multi-vehicle crash rates, measured in crashes per year. Eight different rural two-lane state highway locations across Connecticut with hourly volume counts between September 1990 and October 1996 were studied. Poisson regression was used to estimate and evaluate models for predicting the expected number of single and multi-vehicle crashes at these sites.

2. Previous work

A number of researchers have investigated the effect of exposure on the number of crashes per year, or crash

* Corresponding author. Tel.: +1-860-486-0352; fax: +1-860-486-2298.

E-mail address: johnivan@eng2.uconn.edu (J.N. Ivan)

0001-4575/99/$ - see front matter

PII: S0001-4575(99)00030-5

rate. One of the first such studies was by Gwynn (1967) who analyzed crashes and traffic flow on US Route 22 through the city of Newark, New Jersey. Hourly volumes on every day between the years 1959 and 1963 were classified into 100 volume ranges by magnitude. Crash rates were computed and plotted for each of these volume classes, leading to a distinct 'U' relationship, with more crashes observed at the higher and the lower traffic volumes. Zhou and Sisiopiku (1997) performed a similar study on Interstate 94 in Michigan. This study was slightly different from the previous one in that it included volume/capacity (v/c) ratio instead of the absolute traffic volume. As did Gwynn, they found a distinct 'U' relationship between traffic flow and crash rates. The Gwynn and Zhou and Sisiopiku studies considered only one roadway segment. Since the v/c ratio is the same as the absolute traffic volume where the capacity is constant, the results are the same. An interesting extension of these studies would be to consider multiple roadway segments, with segments having different capacities. Frantzeskakis (1983) suggested that congestion measures, such as the v/c ratio or the level of service (LOS), may be better predictor variables.

Ivan and O'Mara (1997) studied crash rates on mostly two-lane highways in Connecticut. Like other studies, they found the frequency of intersections on the segment to be an important variable. Intersections represent opportunities for vehicle conflicts; however, it is important to realize that mid-block crashes may be caused by completely different phenomena than intersection crashes. Furthermore, vehicle interactions at mid-block locations differ substantially from those at intersections. Vogt and Bared (1998) accounted for this by studying mid-block crashes apart from intersection crashes, and used million-vehicle-miles-traveled for mid-block exposure and the product of the major and minor AADT's for intersection exposure. This study reported different predictors for each type of crash.

Persaud and Mucsi (1995) used hourly volumes derived from aggregate adjustment factors to test the effect of light conditions on these two types of crash. They found that multi-vehicle crashes mostly occurred during the daytime when the light conditions were good, whereas single-vehicle crashes were more likely to occur after sunset. Also, while single-vehicle crashes were associated with narrow lanes and shoulders, multi-vehicle crashes were associated with wide lanes and shoulders. Hence, Persaud and Mucsi recommend that these two types of crashes be modeled separately. It should be noted that these models did not consider actual hourly variation in traffic conditions, so could not account for the actual peaking characteristics of the roadways studied. Furthermore, most analysis to date of factors such as light or surface conditions is often confined to the times when a crash occurred. In this case, one does not have an idea of the number of successful trips made under the same light or surface conditions.

3. Study design

The findings for macroscopic models outlined above demonstrate the importance of highway geometrics, facility type and traffic volume for predicting highway crashes. In addition, there are differences in causality factors for single and multi-vehicle crashes. For this reason, we chose to design a microscopic study with the following specifications:

1. Single and multi-vehicle crashes are modeled separately.
2. Observed hourly volumes are used instead of AADT to capture the actual traffic conditions at the time the number of crashes was observed.
3. The effect of natural light conditions is considered along with other independent variables.
4. LOS is used as a measure of traffic intensity.
5. Different highway geometric variables are studied.
6. Study is limited to two-lane rural roads.

The database used in this study consists of five individual smaller datasets obtained from a number of different sources, which are then merged and aggregated into one database. The following sections describe these datasets.

3.1. ATR database

The Connecticut Department of Transportation (ConnDOT) maintains a total of 37 continuous count stations called automatic traffic recorders (ATR's) to record traffic data through the entire year. Since hourly volume counts are labor intensive and costly, the ATR locations predominantly dictated the sites to be chosen for the study. ATR data were obtained from each of the eight stations on two lane roadways from October 1990 to September 1996. These stations are listed and described on Table 1.

3.2. HPMS database

ConnDOT maintains Highway Performance Monitoring Sites (HPMS) throughout the state. These are randomly selected roadway segments with uniform geometric characteristics that are monitored for roadway condition, environmental and traffic factors. The nearest HPMS site to each of the selected ATR locations provided the geometric characteristics for this dataset. All of the ATR's either coincided with an HPMS site or were close enough to consider the ATR volume to be the volume at the HPMS site as well.

3.3. Crash database

ConnDOT Traffic Accident Summary Reports (TASR) provided the crash data, consisting of detailed information about all crashes that occurred between October 1990 and September 1995 at each of the eight selected HPMS sites. The information from this database included the date, time, location, nature and the type of vehicles involved in each crash, as well as the type of crash.

3.4. Light data

This dataset contains the sunrise and sunset times for each site on every day of the period in consideration. Information about sunrise and sunset times was obtained in order to determine the light conditions under which the trips were made. Data were obtained from the Applied Environmetrics Meteorological Table developed by the National Bushfire Research Unit. The three categories for the light condition variable are:

Light—The time of the day between 1 h after sunrise and 1 h before sunset.

Dark—The time of the day after sunset and before sunrise.

Dusk—The hour after sunrise and the hour before sunset.

3.5. Data aggregation

Merging the crash, ATR and light condition datasets resulted in a large database. Each record contained the number of single and multi-vehicle crashes, observed traffic volumes and prevailing light conditions at all eight highway locations for every hour of every day (with the exception of periods when ATR's were out of service) for the 6-year period between October 1991 and September 1996. A sample of this dataset is shown in Fig. 1; following is an interpretation of the more obscure variable name abbreviations:

st_id = station identification number

dir_tr = direction of travel (1, north; 3, east; 5, south; 7, west)

dat = day of the month (1–31)

day = day of the week (1, Sunday; 2, Monday, etc.)

vol = traffic volume during the indicated hour

time = time of day in 24 h format

l_conds = daylight conditions (1, day; 2, dark; 3, dawn/dusk)

This file was then merged with the much smaller HPMS database (one record for each site for each year), so that each record then also contained the characteristics of the site at its corresponding time.

This database is inappropriate for estimating crash prediction models for several reasons. Each case represents only 1 h of observed crashes under the prevailing traffic, light and site conditions. Since motor vehicle crashes are (thankfully) rather rare events, this results in an enormous number of sites having no observed crashes, and only a handful having one or two crashes. This makes the model estimation task somewhat difficult, as it must overcome the natural bias in the dataset towards the absence of crashes.

Instead, we aggregated the dataset so that each case records the number of crashes occurring in a given 12-month period on a particular highway segment under a given combination of light conditions and LOS. LOS was calculated using the Highway Capacity Manual (HCM) (TRB, 1994) methodology for two-lane highways based on the available sight distance, type of terrain and the prevailing v/c ratio. This aggregation dramatically reduces the number of cases with no crashes. This sets up the dataset to be used for estimating models for predicting single and multi-vehicle crashes under these conditions; a sample of records are shown on Fig. 2; again, a variable list is given for interpreting the abbreviations not already defined:

hpms = HPMS identification number

accs_1 = number of crashes (accidents) occurring under the indicated combination of conditions

veh_mi_1 = number of vehicle miles traveled under the indicated combination of conditions

rate = crash (accident) rate for the aggregated data set in crashes per million vehicle miles traveled

Table 1
ATR locations

Station	Town	Route	Location	Milepost
1	Kent	7	0.5 mi. south of Cornwall TI	55.78
2	East Lyme	1	0.5 mi. east of Stones Ranch Road	89.74
4	Clinton	81	1.0 mi. south of Killingworth TI	2.68
9	New Canaan	124	0.1 mi. south of route N.B. off-ramp	2.76
11	East Windsor	5	0.5 mi. south of route 140	47.23
13	Killingly	12	1.1 mi. north of route 101	40.57
18	Colebrook	8	1.6 mi. north of Winchester Town Line	63.75
20	Hebron	66	1.3 mi. west of route 85	24.60

D:\Research\hwy_risk\WORKING\SANS_PRC\MULTIAND\aap fig 1.sav

	st_id	dir_tr	year	month	dat	day	vol	time	l_conds
1	9016	5	95	10	1	1	210	13:00	1.00
2	9017	5	95	10	1	1	2314	13:00	1.00
3	9022	5	95	10	1	1	237	13:00	1.00
4	9052	5	95	10	1	1	387	13:00	1.00
5	9010	1	95	10	1	1	526	14:00	1.00
6	9016	1	95	10	1	1	221	14:00	1.00
7	9017	1	95	10	1	1	2176	14:00	1.00
8	9022	1	95	10	1	1	208	14:00	1.00
9	9052	1	95	10	1	1	337	14:00	1.00
10	9010	5	95	10	1	1	536	14:00	1.00
11	9016	5	95	10	1	1	196	14:00	1.00
12	9017	5	95	10	1	1	2445	14:00	1.00
13	9022	5	95	10	1	1	218	14:00	1.00
14	9052	5	95	10	1	1	341	14:00	1.00
15	9010	1	95	10	1	1	508	15:00	1.00
16	9016	1	95	10	1	1	236	15:00	1.00
17	9017	1	95	10	1	1	2083	15:00	1.00
18	9022	1	95	10	1	1	260	15:00	1.00
19	9052	1	95	10	1	1	326	15:00	1.00
20	9010	5	95	10	1	1	553	15:00	1.00
21	9016	5	95	10	1	1	212	15:00	1.00
22	9017	5	95	10	1	1	2680	15:00	1.00

Fig. 1. Sample cases from pre-aggregation dataset.

4. Study methodology

4.1. Model form

Statistical estimation procedures require an assumption about the distribution of the dependent variable. The consensus in the research community is that the Poisson distribution is much more suitable than the normal distribution for modeling highway crashes. The Poisson distribution predicts the probability that the number of crashes N is equal to some constant n using the following formula:

$$P(N=n)=\frac{\lambda^n \mathrm{e}^{-\lambda}}{n!} \tag{1}$$

where λ is a constant crash rate. This average crash rate is estimated using Poisson regression analysis with the site and traffic conditions as independent variables.

An important factor in predicting the likelihood of any random event is the number of trials. In highway crash prediction, the number of trials is also called the *exposure*. The exposure is measured by the number of trips made at the site, usually in million-vehicle-miles traveled per year, and typically explains most of the variation in the number of crashes. This being the case, an important step in the model building process is deciding where to incorporate the exposure term into the model. Following are two ways of incorporating exposure:

$$n = Vf(x) \tag{2}$$

$$\lambda = \frac{n}{V} = f(\boldsymbol{x}) \tag{3}$$

where n is the number of crashes, V is the exposure (as defined above), λ is the crash rate in crashes per unit of exposure, and $\boldsymbol{x}$ is a vector of variables representing site and traffic characteristics assumed to be good predictors of highway crashes. These two model forms, while mathematically derivatives of one another, do not con-

vey the same meaning. For instance, the first equation uses the exposure as information for predicting the number of crashes. The second model, by trying to model the crash rate, is indirectly modeling exposure also. Thus, it does not use exposure as information, unlike the first model.

Both models assume that the crash frequency is linearly related to the exposure. However, as discussed earlier, a number of studies have found crash frequency to have a non-linear relationship with exposure. Thus, along with coefficients on the variables $\boldsymbol{x}$ described above, we also estimate an exponent on exposure, to permit the model to include such a non-linear relationship. The choice of a model form is largely arbitrary (Valavanis 1959); the model form chosen here is:

$$n = V^{\alpha}\, e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \varepsilon} \tag{4}$$

where α is the exponent on exposure to be estimated, $x_1, x_2, \ldots$ represent traffic and site characteristics, β_1, $\beta_2, \ldots$ are parameters to be estimated, and ε is an error term.

4.2. Model estimation

A common criticism of the Poisson distribution is its assumption that the variance of the dependent variable is equal to the mean, which in practice is often violated. Usually, the observed variance is greater than the mean (overdispersion) though some studies have also reported the opposite condition (underdispersion). In Poisson regression, the violation of the variance assumption does not change parameter estimates but does render the t-statistics inaccurate, although they can be divided by the square root of the dispersion parameter to correct this (Agresti 1990). The dispersion parameter (τ) is given by

$$\tau = \frac{\chi^2}{N-p} \tag{5}$$

where χ^2 is the computed chi–squared statistic for the model, N is the number of observations, and p is the number of parameters considered in the model.

An alternative method is to use quasi-likelihood rather than maximum likelihood estimation techniques. Both techniques use the same log-likelihood function, but quasi-likelihood estimation does not make any assumptions about the distribution because it allows for separate mean and variance structures by computing the dispersion parameter and assuming that the variance is equal to the product of the mean and the dispersion. This study uses the quasi-likelihood estimation techniques in model estimation as implemented in the statistical package S-PLUS (1995).

Finally, the study uses Akaike's Information Criterion (AIC) for model selection (Bozdogan 1987). AIC uses a theoretical information criterion to identify the optimal model. Unlike other model evaluation methods, AIC identifies the 'best approximating' model among a class of competing models with different numbers of parameters without specifying a significance level. AIC is defined as follows:

D:\Research\hwy_risk\WORKING\SANS_PRC\MULTI_LA\HP_LI_LO\TOTAL.SAV

	hpms	l_conds	los	accs_1	veh_mi_1	rate
1	A001020940	1.00	1.00	.00	192492.2	.00
2	A001020940	1.00	2.00	34.00	3249302	10.46
3	A001020940	1.00	3.00	22.00	3238469	6.79
4	A001020940	1.00	4.00	15.00	4402484	3.41
5	A001020940	1.00	5.00	1.00	306360.3	3.26
6	A001020940	2.00	1.00	5.00	1742991	2.87
7	A001020940	2.00	2.00	7.00	1402349	4.99
8	A001020940	2.00	3.00	2.00	262396.4	7.62
9	A001020940	2.00	4.00	1.00	28606.56	34.96
10	A001020940	3.00	1.00	1.00	213275.5	4.69
11	A001020940	3.00	2.00	8.00	819441.1	9.76
12	A001020940	3.00	3.00	6.00	416462.3	14.41
13	A001020940	3.00	4.00	.00	177984.0	.00
14	A001020940	3.00	5.00	.00	12394.59	.00

Fig. 2. Sample cases from post-aggregation dataset.

Table 2
Preliminary models for single-vehicle crashes

Variable	Model 1	Model 2	Model 3	Model 4
Intercept	−11.00 (−10.03)	−13.57 (−10.88)	−13.09 (−10.75)	−12.78 (−7.74)
Route 1	Base		Base	Base
Route 5	−2.50 (−2.71)		−1.12 (−1.35)	−1.19 (−1.42)
Route 7	−0.38 (−1.11)		−0.66 (−2.14)	−0.65 (−2.10)
Route 8	−0.97 (−1.41)		−1.66 (−2.74)	−1.64 (−2.62)
Route 12	−7.49 (−0.70)		−6.91 (−0.78)	−6.96 (−0.79)
Route 66	−1.42 (−3.70)		−0.68 (−1.89)	−0.70 (−1.94)
Route 81	−0.76 (−1.51)		−0.01 (−0.03)	−0.06 (0.12)
Route 124	−0.72 (−2.29)		0.87 (1.91)	0.79 (1.69)
LOS A		Base	Base	Base
LOS B		−1.32 (−3.75)	−1.67 (−5.18)	−1.60 (−4.52)
LOS C		−2.54 (−5.17)	−2.83 (−6.08)	−2.74 (−5.42)
LOS D		−2.99 (−4.14)	−3.80 (−5.34)	−3.67 (−4.93)
LOS E		−1.91 (−3.85)	−3.49 (−5.30)	−3.31 (−4.55)
Light				Base
Dark				0.02 (−0.08)
Dusk				−0.23 (−0.47)
Exponent on exposure	0.86 (9.52)	1.14 (9.82)	1.16 (10.51)	1.13 (8.21)
Dispersion	0.80	0.90	0.60	0.60
Null deviance	363.50	363.50	363.50	363.50
Residual deviance	218.72	204.73	185.98	185.80
AIC value	236.72	216.73	211.98	215.80

$$\mathrm{AIC} = 2ML + 2K \tag{6}$$

where ML is the maximum log-likelihood and K is the number of free parameters in the model. The best model has the minimum AIC value.

5. Results

5.1. Categorical analysis

Prior to estimating crash models with the specific site variables, preliminary model estimation was undertaken with the site location, the LOS and the light conditions entered as categorical independent variables. Following is the slightly revised form of this categorical model:

$$n_{ijkt} = V^{\alpha}_{ijkt} \exp(\beta_0 + \beta_{Si} + \beta_{Lj} + \beta_{Dk} + \in_{ijkt}) \tag{7}$$

where the symbols n and V are as defined before, and subscripts i, j, k and t denote values for case t at site i under LOS j and daylight conditions k. The β subscripts 0, S, L and D, respectively indicate the estimated intercept and groups of coefficients describing the effect of each site, LOS level and daylight condition (day, dawn/dusk and dark). Regression coefficients were estimated for each variable level except a base level for each. Hence, the resulting t-statistics test a null hypothesis that the difference in means between a particular level of the variable and the base value is zero, in other words,

$$H_0{:}\lambda_{\mathrm{Rte.7,LOSA,Dark}} - \lambda_{\mathrm{Rte.LOSA.light}} = 0 \tag{8}$$

Such an analysis only directly provides information about comparisons between the base case and each level; other pair-wise comparisons require paired t-tests between the desired variable levels.

Four models were estimated for single-vehicle crashes; their results are displayed in Table 2. The base case for this analysis is route 1 at LOS A under daylight conditions. Based on the AIC and t-statistics, model 3 is best. Models 1 and 2 each have just one explanatory variable, site location and LOS, respectively, and explain less variation than model 3, which includes both. Model 4 adds light conditions, but this actually diminishes model performance. In model 3, all sites are significantly different from the base case, route 1, except for routes 12 and 81. These results suggest that both LOS and the site characteristics are required for explaining the number of single vehicle crashes, but that light conditions do not help.

These results must be used carefully. LOS is highly correlated with the site characteristics because it is the site geometry that determines the capacity of a roadway and LOS is computed from the capacity. Also, note that the single-vehicle crash rate decreases monotonically as LOS becomes poorer. This is probably because the better levels of service are more likely to be observed during very early and very late hours of the day, when drivers can be very tired or sleepy and less alert. This point will be discussed later in more detail.

The same four variable combinations were estimated for multi-vehicle crashes; the results are shown on Table 3. This time, model 1, including only the site

Table 3
Preliminary models for multi-vehicle crashes[a]

Variable	Model 1	Model 2	Model 3	Model 4
Intercept	−10.08 (−11.53)	−9.88 (−12.04)	−9.98 (−10.72)	−11.31 (−7.33)
Route 1	Base		Base	Base
Route 5	−2.03 (−3.72)		−2.28 (−3.86)	−2.30 (−3.78)
Route 7	−1.86 (−4.19)		−1.84 (−4.12)	−1.81 (−3.99)
Route 8	−1.17 (−2.10)		−1.11 (−1.90)	−0.94 (−1.50)
Route 12	0.88 (3.13)		0.86 (3.00)	−1.01 (−3.18)
Route 66	−1.95 (−5.70)		−2.09 (−5.45)	−2.21 (−5.35)
Route 81	−2.59 (−3.15)		−2.67 (−3.22)	−2.63 (−3.05)
Route 124	−0.99 (−4.00)		−1.54 (−3.18)	−1.66 (−3.13)
LOS A	Base	Base	Base	Base
LOS B		0.64 (1.67)	0.15 (0.38)	0.14 (0.32)
LOS C		−0.14 (−0.32)	0.21 (0.49)	0.33 (0.67)
LOS D		−0.18 (−0.39)	0.63 (1.16)	0.76 (1.24)
LOS E		0.46 (1.02)	0.83 (1.28)	0.86 (1.21)
Light				Base
Dark				0.26 (0.79)
Dusk				0.45 (1.27)
Exponent on exposure	0.88 (12.20)	0.76 (10.37)	0.86 (10.73)	0.96 (7.96)
Dispersion	1.37	1.44	1.38	1.40
Null deviance	788.97	788.97	788.97	788.97
Residual deviance	373.28	504.44	370.26	367.97
AIC value	391.28	516.44	396.26	397.97

[a] Values in parentheses are t-statistics after correction for dispersion.

variables, is clearly the best. It has the smallest AIC value, and none of the LOS or light condition coefficients in the other models are significantly different from zero at 95% confidence. It seems that segment LOS and light conditions are not useful for predicting multi-vehicle crashes. This point will also be discussed later.

5.2. Site characteristics

The preliminary analysis clearly showed that each site has a different average crash rate for both crash types. The next step is to replace the site variable with the actual site characteristics in the model estimations. These site characteristics are represented by a linear combination of specific continuous variables, as represented in the initial model form. The LOS and light condition variables were retained as categorical variables. The list of site variables and their abbreviations as used in the models are shown in Table 4. The base model (with respect to LOS and light conditions) was defined as before. All site characteristic, traffic, and light condition variables, and all two-way interactions between the categorical variable levels, were considered as predictors in estimating models for both types of crashes.

Table 5 shows the results from estimating models for single-vehicle crashes. Model 5 includes the five LOS categories, along with three site-related variables: the percentage of the section with a sight distance of at least 1500 feet, the right shoulder width (feet), and the percentage of single-unit trucks in the peak traffic stream. All coefficients are significantly different from zero with at least 95% confidence; however, paired t-tests show that the coefficients on LOS D and LOS E are not significantly different from each other. Consequently, model 6 was run, which combines LOS D and E into a single category. The resulting increase in residual deviance is offset by the reduction in model complexity (one less parameter), so this model has a better (smaller) AIC value.

Table 4
Site characteristic variables

Variable name	Type of variable	Observed values
Functional classification	Categorical	Principal arterial—other[a]; minor arterial; major collector
Right shoulder width (feet)	Continuous	0–8
Sight distance over 1500 feet (% of section)		0, 10, 80
Number of signals	Continuous	0–1
% Single unit trucks (daily)	Continuous	0–4
% Single unit trucks (peak)	Continuous	0–2
Speed limit (mph)	Continuous	35, 40, 45

[a] Non-freeway.

Table 5
Final models for single vehicle crashes[a]

Variable	Model 5	Model 6	Model 7	Model 8	Model 9
Intercept	−14.78 (−11.74)	−14.88 (−12.22)	−13.06 (−10.48)	−13.06 (−11.84)	−8.91 (−4.13)
LOS A	Base	Base	Base	Base	Base
LOS B	−1.65 (−5.22)	−1.66 (−5.30)	−1.66 (−4.89)	−1.65 (−4.91)	−1.59 (−4.90)
LOS C	−2.81 (−6.27)	−2.83 (−6.31)	−2.53 (−5.36)		
LOS D	−3.90 (−5.72)				
LOS E	−3.70 (−6.31)				
LOS (D, E)		−3.77 (−7.01)	−2.59 (−5.90)		
LOS (C, D, E,)				−2.56 (−6.57)	−2.76 (−6.86)
Sight distance over 1500 feet	−0.02 (−2.30)	−0.02 (−2.30)	−0.02 (−2.31)	−0.02 (−2.44)	−0.02 (−2.49)
Right shoulder width (feet)	−0.27 (−4.23)	−0.27 (−4.36)	−0.12 (−2.32)	−0.12 (−2.33)	−0.15 (−2.84)
% Single-unit trucks (peak)	1.47 (3.68)	1.49 (3.80)			
% Single-unit trucks (daily)			−0.01 (−0.06)		
Speed limit (mph)					−0.09 (−2.10)
Exponent on exposure	1.18 (11.71)	1.19 (12.13)	1.16 (11.26)	1.16 (11.50)	1.13 (11.53)
Dispersion	0.64	0.64	0.75	0.75	0.71
Null deviance	363.50	363.50	363.50	363.50	363.50
Residual deviance	185.95	186.01	194.94	194.96	191.44
AIC value	203.95	202.01	210.94	206.96	205.44

[a] Values in parentheses are t-statistics after correction for dispersion.

The peak hour truck percentage variable seemed inappropriate, since each case potentially represents conditions aggregated over periods of time other than the peak. We therefore ran model 7, which replaces the peak truck variable with a corresponding daily variable; this model performs much worse; the coefficient on the daily truck percentage is not significantly different from zero at even 90% confidence. Furthermore, a paired t-test comparing the coefficients on LOS C and the combined LOS D/E category showed them to be statistically identical at 95% confidence ($t^* = 1.96$).

In consideration of these findings, we then combined LOS C, D and E into one category and removed both truck variables: this was run as model 8. Because other research by the senior author (Ivan and O'Mara 1997) has found the posted speed limit to be a significant variable for predicting crash rates, model 9 was then run adding 'speed limit' to model 8. Model 9 performs better than model 8, having a smaller AIC, but not as well as model 6, which included the suspect peak single-unit truck percentage variable.

Similar analysis was performed with the number of multi-vehicle crashes as the dependent variable; these results are shown on Table 6. In the first model (number 5), the best predictor variables were found to be the functional classification of the road, the number of traffic signals, the daily truck percentage, and the right shoulder width. Only the coefficient on the principal regional arterial roadway class was significantly different from the others (minor arterial or major collector) at 95% confidence; this class has a significantly lower average crash rate than the others. The positive coefficients on both the number of signals and the daily single unit truck percentage make intuitive sense; one would expect both of these variables to increase the intensity of conflicts between vehicles.

However, the positive coefficient on right shoulder width is troubling; one normally expects a wider shoulder to be a safety feature. Consequently, we examined the data surrounding this variable more carefully. Figure 3 shows how the multi-vehicle crash rate varies with the right shoulder width: note how there is no clear linear relationship between the two quantities. We therefore examined several categorical representations of right shoulder width; those that performed best are shown as models 6, 7 and 8. These three models each divide the right shoulder width into two categories, with the lower category having shoulder widths up to 2, 3 and 4 feet, respectively. None of these models perform as well as model 5. The issue of right shoulder width will be discussed more in the next section.

6. Discussion

Several general observations may be made about these results:

1. The coefficient on shoulder width was *negative* for predicting single vehicle crashes, but *positive* for multi-vehicle crashes.
2. For single-vehicle crashes, the most important predictor variables are the LOS and the percent sight distance over 1500 feet.
3. For multi-vehicle crashes, the most important predictor variables are the class of roadway, number of signals and daily single unit truck percentage.

Single vehicle crashes seem to occur when traffic volume is low and on roads with less forgiving geometry. This makes sense, since a single vehicle crash involves a driver losing control and colliding with a fixed object off the roadway (or an object on the roadway). Under low traffic volumes, there is less likely to be another vehicle to hit, and the less than ideal geometry (sharp horizontal curves and narrow shoulders) reduces the maneuvering room for the driver to recover without leaving the pavement, for example, if the driver had been speeding.

Conversely, multi-vehicle crashes appear to occur in the presence of intense traffic conflicts such as would be observed at a traffic signal, or on roads serving local trips, or when there are many trucks in the traffic stream. Traffic signals tend to be placed in locations where intersecting streets have high traffic volumes. Minor arterials and major collector roads tend to serve local trips and thus have more driveways and local street intersections, increasing the potential for vehicle conflicts.

Clearly, more analysis is required to properly interpret these findings. For one thing, the higher levels of LOS—which correspond to the higher numbers of single-vehicle crashes—are observed more typically at night rather than during the day. Consequently, it is not immediately clear from this analysis whether it is the low volumes, the darkness, or drowsiness of late night drivers that results in greater numbers of single-vehicle crashes. Each of these factors needs to be accounted for explicitly in order to fully understand this phenomenon.

A second issue with these findings is that the LOS was computed for the analysis *segment*, not for the *intersections* on the segment. This might explain why LOS was not significant for multi-vehicle crashes, which occur where there is potential for vehicle conflicts, such as at intersections. If LOS were computed for the intersections on the segment, then it might become significant for predicting multi-vehicle crashes.

Finally, it is difficult to explain the positive coefficient on right shoulder width in the multi-vehicle crash models. One interpretation is that with wider shoulders drivers are more likely to drive too fast and collide with other vehicles exiting or entering driveways. Another is that drivers might be tempted to use a wide shoulder to pass drivers waiting to make left turns, and thus not see a hazardous situation beyond the stopped vehicle. Both of these explanations are related to the presence of intersections on the segment; this issue is addressed in the next section.

7. Conclusions

In response to these findings, ongoing research by this research team is aimed at including new variables which attempt to clarify the effects of the predictive factors used in this study. Specifically, intersection volumes and the potential for conflicts leading to angle and turning collisions are incorporated into the analysis by considering the AADT on roads intersecting each analysis section. Second, the number of driveways of different types (e.g. residential, service station, retail, etc.) is included to account both for the land use intensity and for conflicts associated with driveway intersections. Third, the potential effect of time of day is being investigated by incorporating it as an additional variable along with daylight conditions.

Table 6
Final models for multi-vehicle crashes[a]

Variable	Model 5	Model 6	Model 7	Model 8
Intercept	−15.87 (−14.69)	−14.72 (−14.07)	−15.32 (−11.13)	−18.57 (−12.89)
Minor arterial or major collector	Base	Base	Base	Base
Principal arterial	−5.79 (−9.83)	−3.34 (−8.79)	−4.05 (−7.55)	−7.98 (−7.09)
Number of signals	2.31 (7.78)	1.50 (4.40)	2.53 (6.39)	3.55 (8.53)
Daily single-unit truck percentage	1.26 (6.84)	1.01 (5.39)	1.40 (4.92)	2.24 (8.17)
Right shoulder width (feet)	0.44 (5.15)			
Right shoulder width (≤2 feet)		Base		
Right shoulder width (>2 feet)		1.16 (5.07)		
Right shoulder width (≤3 feet)			Base	
Right shoulder width (>3 feet)			0.71 (1.84)	
Right shoulder width (≤4 feet)				Base
Right shoulder width (>4 feet)				3.79 (4.48)
Exponent on exposure	0.88 (13.83)	0.90 (13.57)	0.87 (11.67)	0.91 (13.22)
Dispersion	1.30	1.32	1.69	1.33
Null deviance	788.97	788.97	788.97	788.97
Residual deviance	370.88	374.07	400.26	384.67
AIC value	382.88	386.07	412.26	396.67

[a] Values in parentheses are t-statistics after correction for dispersion.

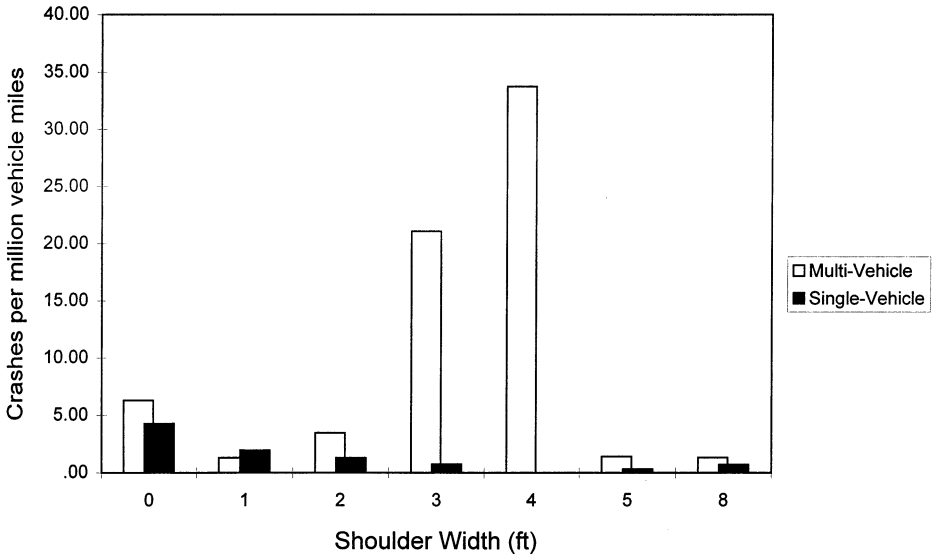

Fig. 3. Single and multi-vehicle crashes per million vehicle miles by right shoulder width.

Intersection traffic volumes and driveway counts are not as readily available as the other data used in this analysis. The permanent count stations only provide data for the main road; consequently, the precise exposure and volumes gathered for each analysis section are not available for the roads intersecting them. Instead, in this new research, these intersecting road volumes are estimated using an AADT observed for each road and the observed hourly volumes. Similarly, the HPMS and other roadway inventory datasets do not provide detailed information about the driveways along the section. This information is being collected using the ConnDOT photolog library, a unique data source comprised of driver's eye view images of the entire length of every state highway in Connecticut, updated once each year. Using the photolog library permits expanding our analysis beyond HPMS sites; consequently, we have identified ten new analysis sites in the vicinity of two-lane ATR stations, each 0.5 mile in length.

This newly expanded dataset permits us to estimate additional models for predicting single and multi-vehicle crashes which consider these new variables. In addition, in this analysis we are including the time of day (divided into five categories), permitting us to isolate time of day, daylight and LOS effects and identify the value of each. We are also computing new variables for measuring exposure to multi-vehicle crashes which consider not only the traffic volume on the section being analyzed, but also the traffic volume on intersecting roads. We expect these new models to explain more variation than the models presented in this paper.

Acknowledgements

The research described in this paper was funded by a grant from the United States Department of Transportation through the New England (Region 1) University Transportation Center and was conducted at the Connecticut Transportation Institute of the University of Connecticut.

References

Agresti, A., 1990. Categorical Data Analysis. Wiley, New York.

Bozdogan, H., 1987. Model selection and Akaike's Information Criterion (AIC): the general theory and its analytical extensions. Psychometrika 52, 345–370.

Frantzeskakis, J., 1983. Accident analysis on two non-controlled access national highways in Greece. ITE Journal 65 (10), 32–34.

Gwynn, D., 1967. Relationship of accident rates and accident involvements with hourly volumes. Traffic Quarterly 21, 407–418.

TRB, 1994. Highway capacity manual. Special Report 209. National Research Council, Washington, DC.

Ivan, J., O'Mara, P., 1997. Prediction of traffic accident rates using Poisson regression. Presented at 76th Annual Meeting of the Transportation Research Board, Washington, DC, paper no. 970861.

Persaud, B., Mucsi, K., 1995. Microscopic accident potential models for two-lane rural roads. Transportation Research Record 1485, TRB, National Research Council, Washington, DC, pp. 134–139.

S-PLUS User's Manual, 1995. Version 3.3 for Windows. Statistical Sciences Division, Mathsoft, Seattle.

Valavanis, S., 1959. Econometrics: An Introduction to Maximum Likelihood Methods. McGraw-Hill, New York.

Vogt. A., Bared, J., 1998. Accident Models for Two-Lane Rural Segments and Intersections. Transportation Research Record 1635, TRB, National Research Council, Washington, DC, 18–29.

Zhou, M., Sisiopiku, V., 1997. Relationship Between Volume to Capacity Ratios and Accident Rates. Transportation Research Record 1581, TRB, National Research Council, Washington, DC, pp. 47–52.