PB2000-106678

ESTIMATING BENEFITS FROM SPECIFIC HIGHWAY SAFETY IMPROVEMENTS

John N. Ivan
Paul J. Ossenbruggen
Chunyan Wang
Nelson R. Bernardo

UNITED STATES DEPARTMENT OF TRANSPORTATION

REGION I UNIVERSITY TRANSPORTATION CENTER

PROJECT UCNR 10-7

FINAL REPORT

April 11, 2000

Performed by

University of Connecticut
Connecticut Transportation Institute
Storrs CT 06269-2037

and

University of New Hampshire
Department of Civil Engineering
Durham NH 03824

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle ESTIMATING BENEFITS FROM SPECIFIC HIGHWAY SAFETY IMPROVEMENTS		5. Report Date March, 2000
		6. Performing Organization Code
7. Author(s) John N. Ivan, Paul J. Ossenbruggen, Chunyan Wang, Nelson R. Bernardo		8. Performing Organization Report No. NEUTC UCNR10-7
9. Performing Organization Name and Address University of Connecticut Connecticut Transportation Institute Civil & Environmental Engineering, U-37 Storrs, CT 06269-2037		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. DTRS95-G-0001
12. Sponsoring Agency Name and Address New England (Region One) UTC Massachusetts Institute of Technology 77 Massachusetts Avenue, Room 1-235 Cambridge, MA 02139		13. Type of Report and Period Covered Final 1997/09/01-1999/08/31
		14. Sponsoring Agency Code

15. Supplementary Notes

Supported by a grant from the US Department of Transportation, University Transportation Centers Program

16. Abstract

In the past thirty years, highway fatality rates have declined steadily because, most notably, of dramatic changes in motor vehicle design, passage of laws making seat belt use mandatory and driving while intoxicated a criminal offense, and educating the public through focused advertising campaigns. However, the practice of highway design has changed little.

Standards, guidelines and warrants are based largely on the opinion of experts and the principles of geometric design. A key element in geometric design is providing the driver with an adequate sighting distance. For a given design speed, a highway is constructed with horizontal and vertical curvatures such that a driver has a sight-line and sufficient time to recognize danger and to stop in a timely fashion. While the principle of adequate sight distance is embraced in practice, it has not been subject to in-depth scientific scrutiny.

The objective of this research is to build upon the scientific framework for identifying hazardous highway locations already begun by these researchers by forming models for predicting the effects of traffic density and land use on highway safety. This will be accomplished by studying a small number of highway locations with varying background conditions. To focus the analysis and improve the quality of the results, study locations will be restricted to rural, two-lane highways. Effects attributable to these factors will be identified by comparing accident histories at sites where other background conditions are similar.

At the present stage of development, models developed by these researchers prove helpful in identifying hazardous highway locations and in identifying contributing factors as to why the highways are deemed hazardous. With a predictive model, proposed measures to reduce both traffic demand and speed with changes in land use policy and highway design changes could be evaluated. The aim to develop these kind of predictive tool with our scientific framework of risk analysis is justifiable.

17. Key Words	18. Distribution Statement
highway safety, risk, Poisson regression, public health, highway capacity, level of service, highway design, traffic exposure	No restrictions, This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified		

Form DOT F 1700.7 Reproduction of form and completed page is authorized

DISCLAIMER

This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers or University Research Institutes Program, in the interest of information exchange. The U. S. Government assumes no liability for the contents

TABLE OF CONTENTS

PART I: Explaining Two-Lane Highway Crash Rates Using Land Use and Hourly Exposure

ABSTRACT

This paper describes the estimation of Poisson regression models for predicting both single and multi-vehicle highway crash rates as a function of traffic density and land use, as well as ambient light conditions and time of day. The study focuses on seventeen rural, two-lane highway segments, each one-half mile in length with varying land use patterns and where actual hourly exposure values are available in the form of observed traffic counts. Land use effects are represented by the number of driveways of various types on each segment. Hourly exposure is represented for single-vehicle crashes as the total vehicle miles traveled and volume/capacity ratio; for multi-vehicle crashes it is the product of the hourly volumes on the main highway and the roads intersecting it along the study segment.

For single-vehicle crashes, the following variables were found to be significant, with a positive or negative effect as noted: daytime (6am-7pm, negative effect), the natural log of the segment volume/capacity ratio (negative), percent of the segment with no passing zones (positive), shoulder width (positive), number of intersections (negative), and driveways (mixed effects by type). Good multi-vehicle crash prediction models had quite different variables: daylight conditions from 10am-3pm and 3-7pm (positive), number of intersections (negative), and driveways (positive for all types). The results show that traffic intensity explains differences in crash rates even when controlling for time of day and light conditions, and that these effects are quite different for single and multi-vehicle crashes. Suggestions for future research are also given.

INTRODUCTION

This study is motivated by research results found by the first author (Ivan *et. al.,* 1999), specifically that single-vehicle and multi-vehicle highway crashes occur under markedly different circumstances with respect to traffic volume, light and roadway conditions. Many other researchers have found similar results. For example, Persaud and Mucsi (1995) found that the effect of daylight conditions is different for single-vehicle and multi-vehicle crashes. For single-vehicle crashes the potential is higher at night, whereas for multi-vehicle crashes the opposite is the case. Persaud and Mucsi also found that while single-vehicle crashes tend to increase with shoulder width, the trend is the opposite with multi-vehicle crashes.

The previous research (mentioned above) by the first author investigated predictive variables for both types of crash using Poisson regression. One variable was hourly Levels of Service (LOS) computed from actual traffic volumes measured at permanent count stations on two-lane highways in Connecticut. Site characteristics were extracted from the Highway Performance Monitoring System (HPMS), and daylight conditions were also considered.

Specifically, single-vehicle crashes were found to occur at better LOS – most at LOS A, next most at LOS B, and the least at LOS C, D or E. These findings are consistent with results reported by Zhou and Sisiopiku (1997). Single-vehicle crashes also seemed to occur where there are narrow shoulders and poor sight distance. Light conditions were not significant. Conversely, LOS did not help predict the number of multi-vehicle crashes at all – the best predictive variables were the number of signalized intersections, a dummy indicating whether or not the facility is a principal arterial, and the percentage of trucks using the road. The number of intersections and truck percentage both increased the number of multi-vehicle crashes; the principal arterial indicator decreased the number of crashes. Again, light conditions were not significant.

These findings raised several issues:

1. Good LOS tends to occur at night when volumes are low, so is the single-vehicle crash rate higher then because there are fewer vehicles on the road (LOS effect), or because more drivers are drowsy or less alert (time of day effect)?
2. LOS was computed for highway segments (segment LOS effect), but multi-vehicle crashes are more related to vehicle conflicts, so would a traffic intensity variable that includes volumes on cross roads be a better predictor (intersection LOS effect)?
3. Is the surrounding land use significant for predicting single-vehicle or multi-vehicle crash rates (conflict or distraction effect)?

In fact, several researchers have recently investigated some of these issues. Mensah and Hauer (1998) studied the relationship between crashes and time of day. They concluded that it is more accurate to estimate separate models for daytime and nighttime conditions, or at least to include a variable to control for the differences between these two time of day conditions. Levinson and Gluck (1997) reviewed studies of the safety effects of driveway access spacings, finding that for many different types of highways, access type and density are good predictors of crash rates. Consequently, the research described in this paper aimed to answer the above questions using these results as a starting point.

STUDY DESIGN

We prepared a study design which permits us to answer these questions. Following are specific features we included in the design:

1. Single-vehicle and multi-vehicle crashes are modeled separately. We included all crashes occurring on each study segment irrespective of severity. Information about crashes came from Connecticut Department of Transportation (ConnDOT) accident experience reports.
2. Hourly exposure data is collected from ConnDOT two-lane automatic traffic recorder (ATR) sites. However, we expanded our data set by redefining our analysis sites exclusive of HPMS locations. Specifically, we defined one-half mile (approximately 0.8 km) highway segments each with homogeneous cross-sectional features (lane and shoulder width), that were also near enough to the ATR sites that the hourly volume could be considered consistent. This permitted us to define a total of seventeen sites with a greater variety of site characteristics than in the previous research. However, we no longer had the HPMS to get all of the site data needed. We will discuss acquisition of site description data sources in the next section.
3. Hourly traffic volumes for the intersecting roads (none of which are state highways) over the entire six year period were simply unavailable, so we used tube counters to observe an average daily traffic (ADT) count for each road over one day, which we then converted to annual average daily traffic (AADT) counts using factors generated by ConnDOT for this purpose.

4. We again determined light conditions for each hour of data. Light condition is classified as dawn, day, dusk, or dark according to sunrise and sunset times estimated using the Applied Environmetrics Meteorological Table developed by the National Bushfire Research Unit.
5. Finally, we defined a variable called *time of day* with five categories: AM peak (6-10am), midday (10am-3pm), PM peak (3-7pm), evening (7-11pm) and nighttime (11pm-6am).

In addition to the temporal variables just described, we included characteristics to describe each site. Following are characteristics describing the geometric features:

1. shoulder width, in feet
2. percent of segment with no passing zones, and
3. number of intersections on the segment.

Land use effects are represented by the number of driveways observed on each highway segment, classified into the following categories:

1. private residence,
2. apartment building with more than four units,
3. gas station,
4. retail,
5. industrial,
6. office, and
7. other (including churches, campgrounds and other recreational sites).

STUDY DATA

The driveway variables listed above are not available from HPMS. Rather than making time-consuming and costly field visits to each site, we used the ConnDOT photolog archives to find this information. The photolog archives are a collection of driver's eye view images taken at 0.01-mile (16.1-meter) increments along the entire length of every state highway in Connecticut, stored on laser disk. Figure 1 is an example of a photolog image, including the pop-up grid, which is useful for measuring the size of roadway features, such as lane and shoulder width. We have a photolog station in our computer laboratory at UConn. By using the photolog, we were able to collect this information without leaving our building.

Once all of the data were collected, they were compiled into a single file for analysis. Following is the procedure:

1. We started with one case for each hour of the six year period for each site.
2. Next, we merged in the traffic volume and light condition data, so for each hour we knew the two-way segment and intersecting road volumes and whether it was dawn, daylight, dusk or dark at the time.
3. Then we added the crash data, or the number of single-vehicle and multi-vehicle crashes that occurred during each hour. Very few cases had more than one crash, and of course, the vast majority had none.
4. Then we defined the time of day variable and computed the volume/capacity (v/c) ratio for each case. The capacity was estimated using the site characteristics and procedures published in the *Highway Capacity Manual* (TRB 1994) for rural two-lane highway segments (there was no intersection control of any kind on any of the main segment approaches for any of the study sites).

5. Because there were so few cases with crashes (recall that each case represents one hour at each site over the six-year period), we needed to aggregate the dataset. Otherwise, the vast majority of cases would have no crashes, and special modeling techniques would be required. In the aggregated data set, each case represents a unique combination of site, calendar year, light conditions, time of day and v/c range in 0.1 increments. Tables 1 and 2 list for each study segment the number of cases with 0, 1, 2, 3 and 4 crashes as well as the total crash count.
6. Finally, the site characteristics (which vary only by site and sometimes by year) were merged into the database.

METHODOLOGY

We estimated non-linear Poisson regression models for single-vehicle and multi-vehicle crashes using quasi-likelihood estimation techniques. The Poisson distribution assumes that the mean and the variance of the data set are equal. This assumption is often violated for crash data because the variance is greater than the mean, a phenomenon called over-dispersion. When the assumption is violated, the efficiency of the parameter estimates is lost, and the t-statistics are corrupt since they are based on biased standard errors. Quasi-likelihood estimation, as implemented in the S-Plus statistical package, accounts for over or under-dispersion in the count observations by estimating the over or under-dispersion parameter as part of the process (S-Plus 1995).

In the model estimation process we weighed each aggregated case by the number of cases, or hours, it represented in the original, unaggregated database. We did this because many cases in the aggregated data set represented a large number of hours, such as daylight conditions in the middle of the day at moderate v/c range, while other cases represented a very small number of hours, such as daylight conditions in the evening at high v/c range. This way the procedure works harder at fitting the more commonly observed cases, rather than trying to fit the rare and common cases equally well.

Following is the general form for our prediction model:

$$N = Ve^{\beta x} \tag{1}$$

where N is the number of crashes, V is the exposure to crashes, x is a vector of independent (predictor) variables, and β is a vector of estimated coefficients. For single-vehicle crashes, the standard exposure measure was used – million vehicle miles traveled at the site. Note that in this study, because all segments have the same length, this measure is simply defined as million vehicles. However, for multi-vehicle crashes, we used a different measure based on one suggested by Vogt and Bared (1998), and defined as follows:

$$V_t = \left[\sum_{h \in H_t} q_h \frac{q_h}{AADT_s} \sum_{m=1}^{n} AADT_m \right] \times 10^{-12} \tag{2}$$

where

V_t is the exposure observed under conditions t, defined by site, year, light conditions, time of day and v/c range,

H_t is the set of hours under which conditions t are observed,

q_h is the traffic volume observed during hour h,

$AADT_s$ is the Annual Average Daily Traffic (AADT) observed on the main highway segment,

n is the number of side roads intersecting the main highway along the half-mile segment, and

$AADT_m$ is the AADT on side road segment m.

Essentially, this formula estimates the hourly volume on the intersecting roads by assuming that the ratio of the hourly volume to the AADT is the same as on the main highway segment. Then, the main highway segment hourly volume is multiplied by the sum of the intersecting road hourly volumes and divided by one million squared and used as the exposure measure for multi-vehicle crashes.

In model estimation for both types of crash, we first estimated models using only the temporal variables: v/c (defined in ranges), light condition, time of day and exposure. These effects were all entered as categorical treatments, as were the site effects. Exposure was defined for single-vehicle crashes as the total of all aggregated hourly volumes. For multi-vehicle crashes it was defined as the value defined in Equation (2). The revised model form for this estimation phase is:

$$N_{yijkl} = V_{yijkl}^{1+\alpha} \exp(\beta + S_i + L_j + T_k + X_l) \qquad \textbf{(3)}$$

where

N_{yijkl} is the number of crashes in year y at site i observed under light conditions j at time of day k and with a v/c range l;

V_{yijkl} is the exposure associated with N_{yijkl};

α is an estimated exponent parameter;

β is an estimated intercept parameter;

S_i is the estimated effect of site i;

L_j is the estimated effect of light conditions j;

T_k is the estimated effect of time of day k; and

X_l is the estimated effect of v/c range l.

Note that the α parameter permits the exposure to contribute to the crash *rate* as well as to scale the number of crashes (linear multiplier by the crash rate). After finding the best temporal variables for predicting each type of crash, we estimated models using those temporal variables and the more detailed site characteristics using the following mixed form, with all symbols as defined previously:

$$N_{yijkl} = V_{yijkl}^{1+\alpha} \exp(\beta x + L_j + T_k + X_l) \qquad \textbf{(4)}$$

RESULTS

Temporal Factors

Table 3 summarizes results of single-vehicle crash model estimation using only the categorical temporal and site factors, including the temporal factor coefficients and t-statistics (all t-statistics reported in this paper are adjusted for over-dispersion). The site factors were included in the models simply to control for variation among the sites; since the next phase of the analysis focused on actual site characteristics, the categorical site factor coefficients provide little useful information (other than whether or not they are significant) and are therefore omitted here for brevity.

As with the previous research cited earlier, Model 1 shows that light conditions are not significant at 95 percent confidence for single-vehicle crashes, although the new time of day variable *is* significant. Consequently, Model 2 was estimated without light conditions (to eliminate insignificant variables); here more time of day categories become significant. However, further investigation (paired t-tests) revealed that these time of day groups can be combined into two groups that are significantly different from one another: daytime (6 am - 7pm) and nighttime (7pm - 6 am). Model 3 was then estimated with only these two categories of time of day. The positive coefficient on nighttime shows that more crashes occur at night, even though v/c is also considered.

While most of the v/c categories are significant in these first three models, the coefficients do not follow a logical pattern. Our previous research found the LOS effect to gradually diminish as v/c increased (Ivan *et al.* 1999), but this pattern is not found here. Therefore, the categorical v/c variable was replaced by a continuous variable, the natural log of v/c, in the estimation of Model 4. Model 4 also drops the α exponent on exposure (million vehicle miles traveled), which was not significant in any of the models. Recall that this exponent was added to 1.0, representing an effect of the exposure on the actual crash *rate*, rather than just on scaling the *number* of crashes. Model 4 thus gives the following model with all factors significant at 95 percent confidence:

$$N_{it} = V_{it} X_t^{-0.33} e^{-0.73+1.22D+S_i} \qquad \textbf{(5)}$$

where N_{it} is the number of crashes at site i under conditions t, V_{it} is the exposure at site i under conditions t, X_t is the v/c under conditions t (moved out of the exponential expression to simplify the mathematics), D is a dummy variable equal to 1.0 only if conditions t are observed at nighttime (7pm-6am), and S_i is the effect of site i.

Table 4 shows the results of estimating categorical temporal and site factor models for multi-vehicle crashes. Models including segment v/c are omitted because this variable is not significant in any of them, which is consistent with previous research findings. One difference from the previous multi-vehicle crash estimation findings, though, is that light condition is now significant along with the new time of day variable. However, due to the strong correlation between light conditions and time, we combined these into a single variable which included a category for each observed combination of these two variables; models that contain this combined variable consistently perform better than models that include them separately, as indicated by the smaller residual deviance on the one included in Table 4.

When comparing paired differences among all levels of the combined variable, only two levels are significantly different from the others. These two levels are daylight conditions between 10am and 3pm and between 3pm and 7pm, and both increase the rate of multi-vehicle crashes. The additive exponent on multi-vehicle exposure is also significant, but negative, which means the effect of the product of the volumes on the main and intersecting roads is less than 1.0. This is not alarming, since this quantity is the product of two traffic volumes, and therefore has units of vehicles-squared; it thus might be more proper to use the square root of this exposure measure as the reference. This final temporal model for multi-vehicle crashes has many more coefficients than the one for single-vehicle crashes, so it is not practical to show it in the text.

Site Factors

The categorical site variable was significant for both types of crash (for at least a majority of sites in each model estimated), indicating that even when controlling for temporal factors, site characteristics still explain some of the variation in each type of crash. Therefore, our next step was to estimate models using the site characteristics added to the best model for each type of crash just presented. Table 5 presents results for single-vehicle crash prediction models that consider the driveway variables, along with several key site variables: shoulder width, percent no passing zone and the number of intersections, all of which were found to be significant for predicting highway crashes in previous research (Ivan and O'Mara 1997). The first model considers all driveways in one variable, rather than by categories. The coefficient on driveways is insignificant at 95 percent, and the shoulder width coefficient has the wrong sign – we expect crash rate to <u>decrease</u> as shoulder width increases. The second model considers each type of driveway separately, but only three are significant: gas station driveways decrease the single-vehicle crash rate, but apartment and other driveways increase it. Shoulder width again has an unexpected positive coefficient. These results will be discussed more later.

Table 6 shows the results for similar models estimated for multi-vehicle crashes. Here, total driveways is significant, but this model does not perform very well (note the higher residual deviance). In the second model, the apartment, gas station, retail and office driveway types are combined into one category called commercial, because there were not many sites with these driveway types. What is most interesting here is that other driveways are most dangerous (i.e. greatest coefficient), followed by industrial, commercial and lastly by residential. Other driveways consists of churches, campgrounds and other recreational sites which have inconsistent traffic volume patterns, so drivers on the main road may not expect to see vehicles entering and leaving. Industrial driveways are likely to have slow moving trucks entering and leaving, increasing the opportunities for vehicle conflicts.

DISCUSSION

Contrary to expectations, driveway variables were significant for predicting single-vehicle crashes as well as multi-vehicle crashes. The best single-vehicle crash models tell us that sites with a lot of gas station driveways and street intersections tend to have fewer single-vehicle crashes, and that sites with a lot of apartment driveways tend to have more single-vehicle crashes. This might be explained by the fact that gas stations are often well lit and increase the nighttime visibility, helping drivers to stay on the road (recall that the single-vehicle crash rate is greater in the evening). Similarly, in the vicinity of intersections, drivers might be more cautious and either reduce their speeds or increase their alertness levels. There is no obvious explanation for the positive effect of apartment driveways and shoulder width. The multi-vehicle crash rate increases with all types of driveway, but mostly with industrial and other (churches and campgrounds) driveways. This is probably because industrial driveways involve slow-moving vehicles entering and leaving the roadway, and drivers do not expect to see traffic entering and leaving the other driveways.

Time of day is significant for both types of crash, but in different ways. single-vehicle crashes occur most often in the evening and at night, which is consistent with most other research findings (Mensah and Hauer 1998). What is significant about this finding is that v/c was also considered, and both variables are still significant. This time of day is more dangerous probably because drivers are more likely to be drowsy and less alert (or driving under the influence) than at other times of day, and thus more likely to lose control of their vehicles. On the other hand, multi-vehicle crashes are more likely to occur under daylight conditions at midday and during the evening peak period. This is when traffic volumes are the heaviest, and there are more discretionary trips than in the morning peak period.

Hourly exposure was also significant for both types of crash, but represented differently. For single-vehicle crashes, there is a negative-exponential relationship with the segment v/c, indicating that crash rate is highest at low v/c, drops sharply to a point, and then levels off. This is consistent with previous findings. Conversely, for multi-vehicle crashes the segment v/c is not significant at all, probably because it has only to do with the intensity of traffic on the main road, and nothing to do with the intensity of conflicts between intersecting roads. Instead, the additive exponent on the multi-vehicle exposure is significant as a predictor variable. Also noteworthy is that when this estimated exponent on multi-vehicle exposure is added to the offset exponent (1.0), the result is very close to 0.5, indicating that the offset is actually the square root of the exposure measure. This is actually quite intuitive, given that the exposure measure is in units of vehicles squared; this issue is explored in a forthcoming paper by the first two authors.

Note that we have limited driveway data, especially for gas station, apartment and office driveways. Therefore, the driveway coefficients should be interpreted carefully. Their effects should be investigated with more data before putting much stock in their significance, particularly for single-vehicle crashes, since these effects are not intuitive.

On the other hand, the time of day effects are strong and quite easily explained. Traffic at different times of day is composed of travelers making different types of trips, and drivers have different levels of alertness at night and during the day. It appears that the morning peak period is the safest time to be on the road; perhaps the traffic stream consists primarily of commuters who are familiar with their travel routes and all act more predictably than drivers at other times of day.

CONCLUSIONS

Three issues were raised in the introduction, which were to have been addressed by this paper. These issues are restated below, along with what was learned about them:

1. The first issue was whether the single-vehicle crash rate is higher at night due to a time of day effect or to the lower traffic intensity at that time. The findings reported in this paper show that actually both of these factors appear to influence the single-vehicle crash rate. This demonstrates clearly that traffic intensity is extremely important for accurately predicting single-vehicle crash rates and analyzing the causes of high crash rate locations more intelligently.
2. The second issue was whether or not the new intersection-related traffic exposure variable would be a better predictor for the multi-vehicle crash rate than a traditional segment-related traffic intensity variable (segment LOS) apart from being used as the exposure offset. In fact, these findings show this to be the case – the segment LOS was not significant in any of the multi-vehicle crash models estimated, but the new exposure term was significant as an additive exponent. Knowing the actual traffic intensity is thus just as important for multi-vehicle crashes.
3. The third issue was whether or not the surrounding land use (represented by driveways of various types) would be significant for predicting single-vehicle or multi-vehicle crash rates. The findings suggest that the number of driveways of

different types is indeed significant for predicting both types of crash. However, due to the limited sample size (only seventeen sites) and variability in these variables, we do not advise transferring these findings to other sites. The findings do suggest, however, that this is a factor that warrants more investigation, as it shows promise in explaining why some highway segments have much higher crash rates than others that are identical in other ways.

ACKNOWLEDGMENTS

This paper is accepted for publication in *Accident Analysis & Prevention.* These results were also presented at the 24th International Forum on Traffic Records and Highway Information Systems in Minneapolis, July 26-29, 1998.

REFERENCES

Ivan, J. N. and O'Mara, P. J. Prediction of non-freeway traffic accident rates in Connecticut using Poisson regression. Presented at Transportation Research Board Annual Meeting, Washington DC, paper no. 970861; 1997.

Ivan, J. N., Pasupathy, R. K., Ossenbruggen, P. J. Differences in causality factors for single and multi-vehicle crashes on two-lane roads. *Accident Analysis and Prevention* **31(?):** 695-704; 1999.

Karlaftis, M. G. and Tarko, A. P. Heterogeneity considerations in accident modeling. *Accident Analysis and Prevention* **30(4):** 425-33; 1998.

Levinson, H. and Gluck, J. Safety benefits of access spacing. *Traffic Congestion and Traffic Safety in the 21st Century,* ASCE Conference, Chicago IL; 458-464; 1997.

Mensah, A. and Hauer, E. Two problems of averaging arising in the estimation of the relationship between accidents and traffic flow. *Transportation Research Record* **1635:** 37-43; 1998.

Persaud, B. N. and Mucsi, K. Microscopic accident potential models for two-lane rural roads. *Transportation Research Record* **1485**:134-139; 1995.

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

Transportation Research Board (TRB). *Highway Capacity Manual.* Washington DC: Special Report 209, National Research Council; 1994.

Vogt, A. and Bared, J. Accident models for two-lane rural segments and intersections. *Transportation Research Record* **1635:**18-29; 1998.

Zhou, M. and Sisiopiku, V. Relationship between volume-to-capacity ratios and accident rates. *Transportation Research Record* **1581:**47-52; 1997.

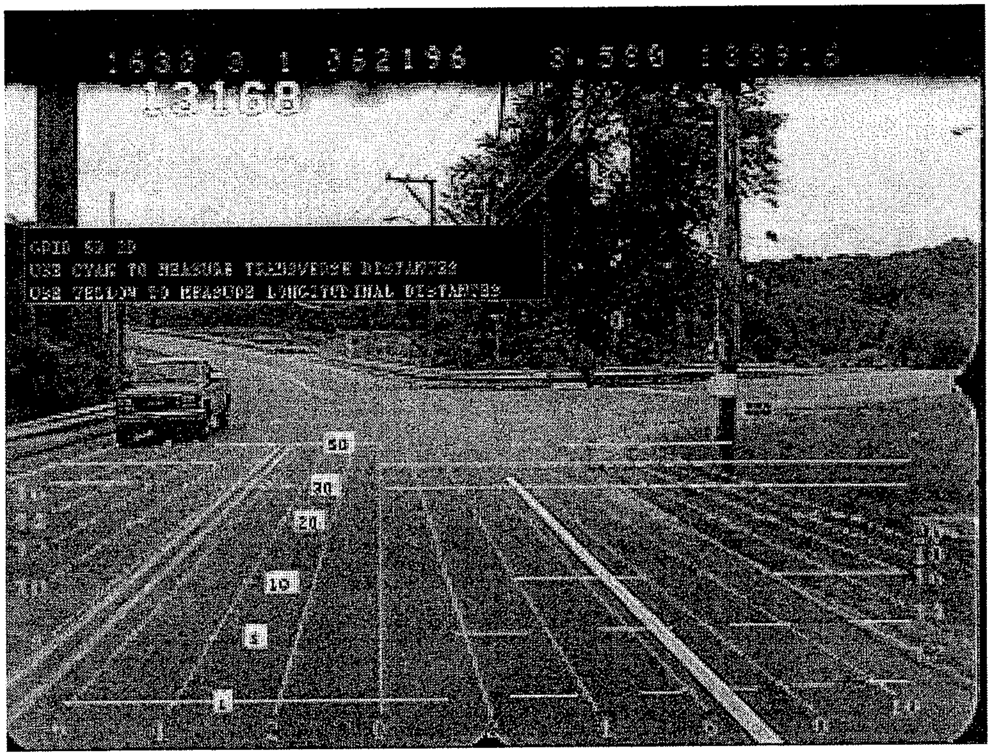

Figure 1. Sample ConnDOT Photolog Image

Table 1. Single-Vehicle Crash Counts by Study Segment

Segment Number	Number of Cases by Single-Vehicle Crash Count					Crash Count
	0	1	2	3	Total	
1	150	6			156	6
2	202	6	1		209	8
3	195	2			197	2
4	161	8			169	8
5	202	2			204	2
6	321	3			324	3
7	373	9		1	383	12
8	358	11	1		370	13
9	346	6	1		353	8
10	240	3	1		244	5
11	162	2			164	2
12	189	11			200	11
13	173	5	1		179	7
14	185	2			187	2
15	276	7			283	7
16	310	4			314	4
17	229	2			231	2
TOTAL	4072	89	5	1	4167	102

Table 2. Multi-Vehicle Crash Counts by Study Segment

Segment Number	Number of Cases by Multi-Vehicle Crash Count						Crash Count
	0	1	2	3	4	Total	
1	153	3				156	3
2	202	7				209	7
3	170	18	4	2	3	197	32
4	166	2	1			169	4
5	196	7	1			204	9
6	300	17	6		1	324	29
7	363	19	1			383	21
8	354	13	3			370	19
9	320	25	8			353	41
10	231	13				244	13
11	157	5	2			164	9
12	190	10				200	10
13	174	5				179	5
14	187	0				187	0
15	282	1				283	1
16	305	9				314	9
17	223	8				231	8
TOTAL	3973	162	26	2	4	4167	220

Table 3. Temporal Factor Models for Single-Vehicle Crashes

		Model 1		Model 2		Model 3		Model 4	
		Coeff.	t-value	Coeff.	t-value	Coeff.	t-value	Coeff.	t-value
Intercept		0.019	0.026	-0.44	-0.07	0.47	0.97	-0.73	-1.92
Ln(V) (10^6 veh.-miles)		-0.019	-0.12	-0.10	-0.70	-0.06	-0.52		
Light Condition	Dawn	1.06	1.49						
	Day	Base							
	Dusk	0.15	0.19						
	Dark	0.84	1.56						
Time of Day	6am-10am	Base		Base					
	10am-3pm	0.58	1.52	0.46	1.32				
	3pm-7pm	0.81	2.00	0.72	1.96				
	7pm-11pm	0.74	1.24	1.33	4.09				
	11pm-6am	0.74	1.23	1.31	3.80				
	6am-7pm					Base		Base	
	7pm-6am					0.97	4.49	1.22	5.47
V/C	0 - 0.1	Base		Base		Base			
	0.1 - 0.2	-0.80	-3.49	-0.83	-3.71	-0.79	-4.08		
	0.2 - 0.3	-1.75	-4.26	-1.81	-4.46	-1.66	-4.44		
	0.3 - 0.4	-0.59	-1.52	-0.68	-1.79	-0.64	-1.82		
	0.4 - 0.5	-1.50	-3.48	-1.54	-3.65	-1.44	-3.67		
	> 0.5	-1.68	-2.95	-1.78	-3.16	-1.54	-2.96		
Ln (V/C)								-0.33	-3.52
Residual Deviance		320,526		321,545		322,562		328,535	
Degrees of Freedom		4137		4140		4143		4148	

Table 4. Temporal Factor Models for Multi-Vehicle Crashes

		Model 1		Model 2		Model 3		Model 4	
		Coef.	t-value	Coef.	t-value	Coef.	t-value	Coef.	t-value
Intercept		3.72	2.96	4.81	4.24	5.24	4.47	3.15	2.42
Ln (V) (10^{12} veh^2)		-0.47	-8.04	-0.43	-8.24	-0.38	-7.09	-0.51	-8.21
Light Condition	Dawn	-0.68	-1.32	-1.09	-2.16				
	Daylight	Base		Base					
	Dusk	-1.24	-2.89	-0.99	-2.34				
	Dark	-0.95	-3.01	-0.32	-2.17				
Time of Day	6am-10am	Base				Base			
	10am-3pm	0.51	2.92			0.49	2.91		
	3pm-7pm	0.66	3.72			0.58	3.51		
	7pm-11pm	0.74	2.18			0.00	0.00		
	11pm-6am	1.30	3.63			0.58	3.11		
Light / time	Day, 6am-10am							Base	
	Dark, 6am-10am							-4.37	-1.22
	Dark, 3pm-7pm							-0.67	-1.15
	Dark, 7pm-11pm							-0.19	-0.86
	Dark, 11pm-6am							0.27	1.27
	Dawn, 6am-10am							-0.76	-1.46
	Day, 10am-3pm							0.53	2.95
	Day, 3pm-7pm							0.67	3.73
	Day, 7pm-11pm							-0.45	-0.50
	Dusk, 3pm-7pm							-0.70	-1.33
	Dusk, 7pm-11pm							-0.44	-0.59
Residual Deviance		345,102		351,016		348,911		343,936	
Degrees of Freedom		4142		4146		4145		4139	

Table 5. Site Factor Models for Single-Vehicle Crashes

	Driveway Types Combined		Driveway Types Separated	
	Coefficient	t-value	Coefficient	t-value
Intercept	-2.30	-4.15	-2.36	-4.26
Ln(v/c)	-0.36	-4.83	-0.24	-3.18
Dummy (6am-7pm)	-1.17	-6.25	-1.38	-7.42
Pct. No passing zone	0.03	5.20	0.03	5.50
Shoulder width	0.06	2.39	0.13	4.49
Intersections	-0.15	-2.57	-0.37	-6.23
Total driveways	-0.01	-1.38		
Gas station driveways			-0.65	-4.71
Apartment driveways			1.43	7.31
Other driveways			0.11	2.92
Residual deviance	354,831		342,561	
Degrees of freedom	4160		4158	

Table 6. Site Factor Models for Multi-Vehicle Crashes

		Driveway Types Combined		Driveway Types Separated	
		Coefficient	t-value	Coefficient	t-value
Intercept		-4.10	-12.55	0.81	1.42
Ln(V)		-1.01	-51.03	-0.65	-16.51
Light / time	Day, 6am-10am	Base		Base	
	Night, 6am-10am	-5.08	-1.91	-4.20	-1.76
	Night, 3pm-7pm	-1.31	-2.19	-0.86	-1.53
	Night, 7pm-11pm	-0.49	-2.13	-0.28	-1.31
	Night, 11pm-6am	-0.53	-2.66	0.06	0.32
	Dawn, 6am-10am	-1.51	-2.83	-0.98	-1.95
	Day, 10am-3pm	1.13	6.63	0.70	4.23
	Day, 3pm-7pm	1.12	6.29	0.80	4.72
	Day, 7pm-11pm	-1.46	-1.62	-0.73	-0.86
	Dusk, 3pm-7pm	-1.04	-1.95	-0.79	-1.57
	Dusk, 7pm-11pm	-1.30	-1.70	-0.68	-0.95
Intersections		0.17	3.60	-0.81	-8.21
Total driveways		0.08	11.02		
Residential driveways				0.04	3.74
Commercial driveways				0.14	13.28
Industrial driveways				0.26	3.01
Other driveways				0.31	9.54
Residual deviance		411,331		361,354	
Degrees of freedom		4153		4150	

PART II: Representing Traffic Exposure in Multi-Vehicle Crash Prediction for Two-lane Highway Segments

ABSTRACT

This paper describes a study of multi-vehicle crash potential on two-lane rural highways in Connecticut. Seventeen highway segments were studied over the time period October 1990 to October 1996. The effects of three temporal factors - traffic volume, time of day and light condition - on multi-vehicle crashes are investigated using Poisson regression. Special attention is given to the representation of traffic exposure for multi-vehicle crashes.

The square root of the product of hourly volumes on the highway segment and the intersecting roads appears to be the best representation of multi-vehicle crash exposure. This product also contributes more to multi-vehicle crashes at night than during the day. In addition, using light and time combination gives better results than using the variables separately. Higher multi-vehicle crash risk is found in general irrespective of traffic volume from 10am to 7pm.

The impact of segment geometric characteristics appears to be much less important than for single-vehicle crashes when the temporal factors explain more causality of multi-vehicle crashes. Transportation engineers should realize that upgrading geometric characteristics might not reduce multi-vehicle crashes as much as single-vehicle crashes.

INTRODUCTION

Multi-vehicle (MV) crashes on rural two-lane highways have become a critical issue of highway safety. Rural two-lane highways constitute a substantial portion of the North American highway network; according to Kalakota *et al.*, approximately 2.5 million miles, or 63 percent of US highway mileages *(1)*. Furthermore, fifty percent of fatalities occur on two lane rural highways, giving this highway type a higher crash rate (per vehicle mile of exposure) than all others; for example, four to seven times higher than on rural interstate highways. Multi-vehicle crashes also accounted for over 70 percent of injuries, injury crashes, and all crashes *(2)*.

Estimation of the crash potential of road sections usually requires defining a relationship between crash rate and the exposure to crashes, traditionally, million vehicle miles traveled (VMT). This exposure metric has worked well for predicting segment-related crashes or single-vehicle (SV) crashes. However, past studies have revealed that MV and SV crashes relate to exposure differently. For example, Ivan *et al.* found that SV crashes tend to occur at better LOS, whereas LOS did not help predict the number of MV crashes at all *(3)*. Furthermore, Kulmala also found the risk of MV crashes to increase as the traffic volume of the intersecting road increases. Therefore, VMT used in predicting SV crash rates may not be appropriate in predicting MV crashes *(4)*.

Consequently, the focus of this study was to find the appropriate exposure term to represent the effect of traffic volume on MV crash experience for two-lane rural highways in Connecticut. We collected several variables describing traffic volume at 17 study sites, including Annual Average Daily Traffic (AADT) on the highway segment, the sum of the AADT's on roads intersecting the segment and the volume observed on the highway segment in that hour.

Several models for predicting MV crashes were estimated using Poisson regression to find the best way to account for the effect of traffic intensity. These models also controlled for light conditions, time of day and site effects to separate the effects of these variables and more accurately identify which help explain the MV crash rate. The exposure metrics investigated in our models were segment volume/capacity ratio, the aggregated hourly volume on the highway segment, the aggregated hourly volume on intersecting roads, and a new exposure term defined specifically for MV crashes (the aggregated product of hourly volume on the segment and the intersecting roads). Logistic and square root forms of these variables were also considered. The findings indicated that the new exposure term, which incorporates traffic volume on both the segment and the intersecting roads, performs the best.

METHODOLOGY

Poisson Regression

When events of a given group occur in discrete degrees, the probability of occurrence of a particular event in a specified number of trials may be described by the binomial distribution *(5)*. If in a given experiment the number of times a particular event occurs is small compared to the number of opportunities for an event to occur, and the average number of times the event occurs has a finite value, the Poisson distribution can adequately approximate the binomial distribution. The crash frequency along a highway segment, a non-negative discrete quantity, meets these requirements.

Using Poisson regression, it is possible to accommodate a nonlinear regression relationship between the dependent and independent variables. The dependent variable, crash frequency for this study, is defined as the product of the crash rate in crashes per unit exposure and a measure of traffic exposure. We define the dependent variable this way, that is, scaling the crash rate by exposure, because it is more accurate to assume that the number of crashes (rather than the crash rate) has a Poisson distribution *(6)*.

This definition may be written as:

$$N = V\lambda \tag{1}$$

where

N is the number of MV crashes,

V is a measure of traffic exposure, and

λ is the MV crash rate, in crashes per unit exposure.

Then, we relate crash rate to explanatory variables as follows:

$$\lambda = e^{x\beta} \tag{2}$$

where

$\boldsymbol{x}$ is a vector representing explanatory variables, and

β is a vector representing a set of parameters to be estimated.

The exposure metric V could be used not only for scaling the number of crashes with an exponent of 1.0 but also as a predictive variable to allow the exposure to have an additional contribution towards explaining MV crash variation. For example,

$$N = V^{\alpha} e^{\beta x} \tag{3}$$

or, by moving the unknown exponent α (to be estimated) into the exponential expression,

$$N = e^{\alpha \ln V + \beta x} \tag{4}$$

In this case, α is not restricted to any particular value.

One important feature of Poisson regression is that it assumes the variance of the dependent variable is equal to its mean value in the entire dataset. For crash data, this assumption is often violated, with a higher variance being observed. This problem is called over-dispersion. To deal with it, the regression coefficients are estimated using the quasi-likelihood estimation technique, as implemented in the S-PLUS statistical package *(7)*. With quasi-likelihood methodology, an over-dispersed Poisson regression model can be estimated by supplying the appropriate link and variance functions for the Poisson family.

Study Design

One of the complex issues in highway safety evaluation is how to incorporate a numerical measure of exposure, the opportunity for a crash to occur, in the analysis. For SV crashes, for example, exposure is usually defined as the vehicle-miles traveled (VMT). However, for MV crashes, it is more intuitive to investigate measures that incorporate intersecting road volumes, for example, the product of the AADT's on the intersecting roads, as suggested by Vogt and Bared for intersection crashes *(8)*. Moreover, exposure is a good predictor of crash rates. Finding volume/capacity to be significant for predicting SV crashes *(3)* inspired us to consider exposure for MV crash prediction too, but in a different form.

Four basic exposure metrics were considered to account for traffic intensity in this study: (a) v/c ratio, (b) hourly volume on the highway segment, (c) hourly volume on the intersecting roads, and (d) the product of the hourly volumes on the highway segment and the intersecting roads. The effects of these exposure metrics were investigated by changing the exposure metric while keeping the other significant variables, which represent the prevailing roadway conditions, the same in every model.

V/C ratio was computed as the observed hourly volume for each hour in the original data divided by the capacity of the segment. The capacity was calculated using procedures in the Highway Capacity Manual *(9)*. V/C ratio is represented in two forms: one is a categorical variable ranging from 0 to 1 in 0.1 increments, the second a continuous representation, the natural log of the v/c.

Connecticut Department of Transportation (ConnDOT) permanent count stations provided the hourly volume on each highway segment we studied for the six year period from October 1990 to October 1996. However, the hourly volume on intersecting roads is not available directly. Figure 1 illustrates how the hourly volume on intersecting roads was derived. It shows an example study site, which is a half-mile (0.8 km) long highway segment with three unsignalized intersections. Assuming the hourly variations in traffic volume on the intersecting road to be similar to those on highway segment, we estimated the total hourly volume on the intersecting roads to be

$$q_{I_h} = \frac{q_{s_h}}{Q_s} \sum_{i=1}^{n} Q_i \qquad (5)$$

where

q_{I_h} is the estimated sum of the hourly volumes on intersecting roads for that hour,

q_{s_h} is the observed hourly volume on the segment for that hour,

Q_i is the estimated AADT on intersecting road i,

Q_s is the observed AADT on the segment, and

n is the number of intersecting roads (in this case 3).

Note that this quantity is computed for every site and each hour of the study period.

To investigate the efficacy of exposure metrics for predicting MV crashes, we used equation (3) to find the appropriate exposure measure as the multiplier for MV crashes. Note that we do not restrict the exponent α to a value of 1.0. V_{M_t} (aggregated hourly volume on the segment), V_{M_t} and V_{I_t} (aggregated hourly volume on the intersecting roads), and V_{MV_t} (our new exposure metric) were examined as possible forms of V, and defined as follows:

$$V_{M_t} = \sum_{h \in H_t} q_{s_h} \quad (6)$$

$$V_{I_t} = \sum_{h \in H_t} q_{I_h} \quad (7)$$

$$V_{MV_t} = \sum_{h \in H_t} q_{s_h} q_{I_h} \quad (8)$$

where

t is the condition defined by site, year, light condition, time of day and v/c range,

H_t is the set of hours under which condition t is observed.

V_{M_t} is essentially equivalent to the VMT used as exposure for SV crashes since every study segment is of the same length. V_{M_t} was observed from the ATR data, but V_{I_t} is an estimate derived from the intersecting road AADT's and the hourly volume on the segment on the basis of the assumptions mentioned earlier. V_{MV} is the summation of the product of hourly exposure on the highway segment and all intersecting roads on the segment for a specific hour. This metric is inspired by a similar exposure measured by Vogt and Bared, who used the product of average daily traffic (ADT) on two intersecting roads (with different exponents) as the exposure for predicting crashes at the intersection of the two roads. It was significant with a exponent of 0.8 on main road ADT and 0.5 on minor road ADT in their models *(8)*.

The natural log of zero is undefined, causing a problem in computing V_{I_t} and V_{MV_t} for segments having no intersecting roads. We arbitrarily chose to set V_{I_t} in equation (7) and q_{I_h} in equation (8) equal to 1.0 for segments with no intersecting roads. Setting these values to 1.0 enables the log transformation to be taken without distorting the data in a meaningful way.

Based on the analysis results, the best representation was selected as the offset for further estimation. The effects of exposure during different times of day are also investigated through further refinement of this equation.

Data Compilation

Our analysis focused on two-lane, rural highway segments in Connecticut. We chose segments near Automatic Traffic Recorder (ATR) stations so we could get observed hourly traffic volume on highway segments directly. Then we identified segments near each ATR with unique geometric cross-sections, but could be assumed to have the same mainline traffic volumes. Consequently, no signalized intersections are included in the segments, since the traffic volume would likely differ significantly on either side of such junctions. This results in our identifying seventeen one-half mile (0.8 km) long highway segments to use as our study sites.

We defined three types of variables: temporal factors, geometric characteristics and crash experience. The temporal factors include traffic volume, time of day, and light conditions, and vary by time. We obtained hourly traffic volumes from ConnDOT for the time period October 1990 to October 1996 for all ATR stations located on two-lane highways corresponding to our study sites. However, traffic volumes on intersecting roads for each study site were not available. Instead, we estimated the AADT for each road on the basis of an Average Daily Traffic (ADT) observed using tube counters on each road over one day, and seasonal variation factors generated by ConnDOT for this purpose. Time of day was defined as a categorical variable according to typical variation in traffic volume and trip purposes. The five categories are: AM peak (6-10am), midday (10am-3pm), PM peak (3-7pm), evening (7-11pm) and nighttime (11pm-6am). Light conditions for each time period were classified as dusk, day, dawn and dark according to sunrise and sunset times estimated by a computer program, the Applied Environmentrics Meteorological Table developed by the National Bushfire Research Unit. Geometric characteristic data were represented indirectly using segment ids. Since there are too few sites to make statistically reliable inferences about geometric characteristic causality, we instead focused on temporal factors, and especially exposure, or traffic volume. The crash data for these study sites for the years from 1990 to 1996 were provided by ConnDOT from their Accident Experience database.

Once we had all the data we needed, the next step was to compile them into a single file for analysis. The data were combined so that each case represented a single hour of the six year period for each site with the hourly volume, light condition, segment id, time of day, and the number of crashes that occurred during that hour at the location. Obviously, most cases did not have any crashes and very few of them had more than one crash. Consequently, we aggregated the crash and volume data in this original dataset on site, study year, time of day, v/c ratio and light condition as indicated in Equations (6) through (8) so that most cases at least had some crashes. After aggregation, the original 850,000 observations were reduced to 4,167 cases, each representing a unique combination of site, study year, light conditions, time of day and v/c ratio. Table 1 is an example of aggregated data set. All the model estimations and tests were performed on the aggregated data set.

Variables and their definitions in Table 1 are listed as follows:

segmt_id	study segment id (1-17)		
cat_year	categorical year		
		1	10/90 – 09/91
		2	10/91 – 09/92
		3	10/92 – 09/93
		4	10/93 – 09/94
		5	10/94 – 09/95
		6	10/95 – 09/96
cat_time	categorical time		
		1	6am - 10am
		2	10am - 3pm
		3	3pm – 7pm
		4	7pm – 11pm
		5	11pm – 6am

lit_cond	light condition		
		0	dark
		1	dawn
		2	day
		3	dusk
cat_v_c	categorical v/c ratio		
multi	number of multi-vehicle crashes		
main	aggregated hourly volume on highway segment		
cross	aggregated hourly volume on intersecting roads		
newmv	aggregated product of hourly volume on highway segment and intersecting roads		

In the model estimation process we weighted each aggregated case by the number of original cases it represented in the unaggregated database. The effect of weighting by number of cases is to place greater emphasis on observed highway conditions with greater frequency than on those with less frequency. Our rationale is that we want the regression procedure to work harder at fitting the more commonly observed cases than those observed less frequently.

ANALYSIS RESULTS

Find the Appropriate Offset

The following models are designed to permit identifying the most effective exposure metric; their results are presented in Table 2.

Model 1. $N = {V_M}^{\alpha_M} \exp(\beta + S_i + L_j + T_k)$ (9)

Model 2. $N = V_M^{\alpha_M} V_I^{\alpha_I} \exp(\beta + S_i + L_j + T_k)$ (10)

Model 3. $N = {V_{MV}}^{\alpha_{MV}} \exp(\beta + S_i + L_j + T_k)$ (11)

Model 4. $N = {V_{MV}}^{\frac{1}{2}+\alpha_{MV}} \exp(\beta + S_i + L_j + T_k)$ (12)

where

S_i is the effect of site i,
L_j is the effect of light condition j,
T_k is the effect of time of day k, and
α_M, α_I, α_{MV} and β are the parameters to be estimated.

The values presented in Table 2 are the exponents of potential exposure metrics and coefficients for segment id, time of day and light condition. The t-values are statistics for testing whether or not each coefficient estimate is significantly different from zero. Coefficients in shaded bold face are significantly different from zero at 95% confidence level with a t-value larger than 1.96. The dispersion parameter for each model is estimated in S-PLUS to account for the over-dispersion problem in the crash dataset *(7)*. Null Deviance is twice the negative of the log likelihood ratio for the saturated model (defined as a model with a parameter for each record in the dataset, thus permitting perfect predictions). Residual deviance is twice the negative of the log likelihood ratio for the model being fitted. The residual deviance is a measure of fit; the larger the deviance, the worse the fit for models with the same null deviance. Each deviance has degrees of freedom equal to the difference between the number of parameters in the model and the number in the saturated model (the number of cases) *(10)*.

The results for model 1 show that the exponent of V_M is significantly different from 0 with a value less than 1. Not surprisingly, the exponent of V_M is significant again in model 2 when both V_M and V_I are used to account for traffic intensity. However, the exponent of V_I is not significantly different from 0 and has a negative sign. Recall that we derived V_I using equation (4) and (7), so model 2 could actually be rewritten as

$$\text{Model 2: } N = V_M^{\alpha_M + \alpha_I} \left(\frac{\sum_{i=1}^{n} Q_i}{Q_s} \right)^{\alpha_I} \exp(\beta + S_i + L_j + T_k) \tag{13}$$

Finding α_I to be not significantly different from 0 indicates that the ratio of AADT's on the intersecting roads to that on the highway segment does not explain the effect of the conflict between the vehicles on the segment and those on the intersecting roads. The negative exponent would also indicate the crash frequency decreases as the proportion of vehicles entering the junction from the intersecting roads increases. These findings are contrary to Kulmala, in which the share of minor-road traffic has the most importance for crossing accidents (an exponent of 0.8) *(4)*. The estimation error was probably caused by using estimates rather than observed values for the hourly volume on the intersecting roads, which are potentially suspect. The new metric, V_{MV}, attempts to represent the potential for intersecting vehicle conflicts on the roadway, which may help account for the occurrence of MV crashes. The estimated power of V_{MV} in model 3 is significantly different from 0 as expected with an exponent of 0.53.

Since V_{MV} is the product of the volumes on the highway segment and the intersecting roads, it is intuitive to instead consider the square root of V_{MV} as an offset, a scaling multiplier on the right hand side of the model. (Note that the estimated value for the exponent on V_{MV} is 0.53 in model 3.) When we used square root of V_{MV} as the offset and V_{MV} as a predictor in model 4, the additive exponent on V_{MV} is not significantly different from 0, indicating that square root of V_{MV} may be the appropriate multiplier for the MV crash rate. Although models 3 and 4 have a slightly higher residual deviance, the others (models 1 and 2) have key coefficients that are contrary to expected signs, or are insignificant. The square root of V_{MV} appeared to be the best offset for models predicting MV crash rate, and it is used as the offset in the following models.

Exposure/Time of Day Interactions

The time of day variable is related to several aspects of driver characteristics, which may also vary by trip purpose. It also reflects drivers' alertness due to circadian rhythms, which are considered to contribute significantly to motor vehicle crashes *(11)*. On the other hand, light conditions, while obviously related to time of day, are not constant with time of day through the year, particularly in a location as far north as Connecticut. Instead, this quantity captures variability in roadway visibility conditions. Poor visibility is more likely to cause drivers' errors by reducing advance warning time and not having time to respond to unexpected events.

We investigated time of day and light condition using categorical analysis to determine which categories are significant in predicting MV crashes. Studying time of day and light condition together is complicated because the same light condition could cross more than one time of day. These two variables cannot strictly be considered to be independent of one another, and thus, the estimates of their coefficients could be confounded. Consequently, we considered two model forms for exploring the relationship between these two quantities and the MV crash rate.

Model 3. $N = V_{MV}{}^{\alpha_{MV}} \exp(\beta + S_i + L_j + T_k)$ (11)

Model 5. $N = V_{MV}^{\alpha_{MV}} \exp(\beta + S_i + LT_{jk})$ (14)

where LT_{jk} indicates the observed combination of light condition j and time of day k.

The first formulation (model 3, repeated for convenience) includes light condition and time of day effects separately, while the second (model 5) combines them into one variable, with one category for each combination. Table 3 presents results from estimating these models in the format of the previous tables. In these models, the base condition is the category of the variable that is used as a reference in comparison to the other categories and is "day" and "6am-10am" for light condition and time of day, respectively. For the combined variable, it is "day, 6am-10am". The t-statistics test whether or not each coefficient estimated is significantly different from the base case (day and 6am-10am).

The paired t-statistics given in Table 4 and Table 5 test whether or not the differences between each pair of categories of light and time or light-time combination are significant. The test results in Table 4 show that 6am-10am is distinctly different from all other times of day, but the others are not. On the other hand, differences among light conditions are not as clear. However, the paired t-test statistics for the light-time combination in model 5 (Table 5) indicate different results. Only two levels of light-time combination are significantly different from the others with a positive sign, which together represent daylight conditions between 10am and 7pm. This suggests MV crash rate is higher midday and in the evening peak; both increase the rate of MV crashes. The lower residual deviance value of model 5 indicates that using light and time combination gives better results than using the variables separately, probably because it represents the interaction between light condition and time of day more accurately.

The effect of exposure may not be consistent during the entire day. In other words, the contribution made by the same amount of exposure to MV crashes may change by time of day. To investigate the effect of exposure during different times of day, we added unique V variables for all time of day categories except for the base case, 6am-10am, such that:

$$V_k = \begin{cases} V_{MV} & \text{if } T = k \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

where V_K is the exposure variable for time of day k, and T is the time of day for the case.

The hypothesis that exposure has different effects for different time of day category was tested by estimating model 6:

Model 6. $N = V_{MV}^{\alpha_{MV}} V_k^{\alpha_k} \exp(\beta + S_i + LT_{jk})$ (16)

which permitted a different exponent on V_k for each time of day category.

If the paired t-test procedure suggests the exponents on time-specific exposure variables are different from each other, it means that exposure indeed has a different effect at these two times of day. The t-test pairwise comparisons in Table 6 for model 6 indicate that the exponents on exposure for time periods from 6am to 11pm are not significantly different from each other. However, all of them (6am-10am, 10am-3pm, 3pm-7pm, 7pm-11pm) are different from 11pm-6am. We can therefore group the effect of exposure into two significantly different categories: nighttime and non-nighttime. The exponent on V_{MV} is larger from 11pm-6am than at other times of day, suggesting that the same magnitude of exposure causes more multi-vehicle crashes at night than at other times.

We also segregated exposure by the new light-time combination variable in model 7 to see how the effect of exposure varies by each category of these combined factors.

$$\text{Model 7. } N = V_{MV}^{\alpha_{MV}} V_{jk}^{\alpha_{jk}} \exp(\beta + S_i + LT_{jk}) \tag{17}$$

The paired t-test statistics in Table 7 indicate that there are two distinct groups of exposure effect segregated by light-time combination. The exponents on exposure during day 6am-10am, dawn 6am to 10am, day 10am to 3pm, day 3pm to 7pm and dark 7pm to 11pm are not significantly different from each other, while all of these are significantly different from the others. These five levels of combination together represent all times of the day except for the time period from 11pm to 6am. The result is essentially the same as we obtained when exposure is segregated by time of day.

The positive additive exponent on exposure at night from 11pm to 6am has shown that exposure influences MV crash more (same exposure will cause more MV crashes) at night than during the day. It is important to note, however, that many fewer cases of MV crashes are observed at night. This does not contradict with our findings, though, since the exposure would be considerably lower at night than during the day.

Final Models

Many factor levels in model 7 are not significantly different from one another, so it is misleading to report estimates for all of their effects. It is more appropriate to only include factor levels we know for certain are distinct from one another. Consequently, we removed insignificant factor levels one variable at a time, re-checking t-statistics at each step. Model 7 includes exposure represented by V_{MV} segmented by light-time combination, the segment ids, and light-time combination. As shown in Table 7, only the exposure for dark 11pm-6am has a significantly different effect from the exposure for the other categories of light-time combination. Therefore, we dropped all other segregated exposure categories, and estimated model 8.

$$\textit{Model 8. } N = V_{MV}^{\alpha_{MV}} V_{DK116}^{\alpha_{DK116}} \exp(\beta + S_i + LT_{jk}) \tag{18}$$

The paired t-tests for light and time combination in model 8 are presented in Table 9. Dark 11pm-6am, day 10am-3pm and day 3pm-7pm are significantly different from all the other light and time combination categories. Positive signs on day 10am-3pm and day 3pm-7pm indicate these two time periods are more dangerous times to be on the road. Model 9 was then estimated with all but these three categories removed.

$$\textit{Model 9. } N = V_{MV}^{\alpha_{MV}} V_{DK116}^{\alpha_{DK116}} \exp(\beta + S_i + LT_{DK11-6} + LT_{DY10-3} + LT_{DY3-7}) \tag{19}$$

Up to now, we have only discussed the effects of temporal variables such as traffic volume, light condition and times of day. Obviously, it is reasonable to expect that conditions of the site might also impact the occurrence of MV crashes. Consequently, Table 10 gives the result of paired t-tests for differences among the site effects in model 9. Only two of the 17 sites, Rte. 32 and Rte. 6, are significantly different from the others in a consistent manner. The chance of being involved in a MV crash when travelling on these two highways is several times higher than on the others.

Model 10 was then estimated with only these two segment effects (dropping the nonsignificant ones).

$$\textit{Model 10.}\ \ N = V_{MV}^{\alpha_{MV}} V_{DK11.6}^{\alpha_{DK11.6}} \exp(\beta + S_6 + S_8 + LT_{DK11-6} + LT_{DY10-3} + LT_{DY3-7}) \tag{20}$$

Only the significant categories of segregated exposure, light and time combination, and segments are left in this model. However, $V_{MV_{DK11.6}}^{\alpha_{11.6}}$ and $LT_{DK11.6}$ are no longer significant. Following is model 10 with the symbols replaced by the estimated exponents and coefficients.

$$N = V_{MV}^{0.48} V_{DK11.6}^{0.12} \exp(-9.89 + 2.48S_6 + 2.06S_8 - 1.22LT_{DK11-6} + 0.8LT_{DY10-3} + 0.99LT_{DY3-7}) \tag{21}$$

CONCLUSIONS

This paper estimates models for predicting MV crash potential by taking into consideration temporal factors associated with MV crash occurrence, especially the representation of traffic exposure. The following points are concluded on the basis of this study.

Results confirmed V_M, number of vehicles on the highway segment, does not seem to be appropriate as an exposure for MV crash rate, since its exponent is less than 1.0. This suggests the number of trials is less than the number of vehicles travelling on the segment, which is conjecture. However, the square root of the product of the hourly volume on the highway segment and the intersecting roads appears to be better as an offset for MV crash prediction. Because the product is in unit of vehicles squared, it makes sense to take square root to bring the offset into unit of vehicles, resulting in an easily explained means of defining the number of trials.

When the effect of V_{MV} is segregated by time of day or light and time combination, the effect of exposure could be divided into two groups: 6am-11pm and 11pm-6am. During different time intervals, the same magnitude of exposure has a different contribution to MV crashes. In other words, traffic contributes more to MV crashes at night than during the day. This may be because drivers do not expect there to be any other cars on the road when driving at night; so since they are not prepared for the situation, the chance to get involved in MV crashes increases significantly, especially with poor visibility when it is dark.

In addition, we found higher risk in general irrespective of traffic volume from 10am to 7pm. Drivers may be less alert at this time of day due to circadian rhythms. The considerable diversity of drivers (trip purposes, ages) on the road at that time also could contribute to a higher risk. Some drivers may be not familiar with the road, which increases the risk of having crashes.

An interesting finding is that the number of significant sites decreases as the exposure representation becomes more detailed in explaining causality. There are only two out of seventeen sites significant in the final model. The impact of site characteristics on MV crashes appears to be much less significant than was noted in SV crash studies.

One could argue then, that improving geometric characteristics can reduce the SV crash occurrences considerably, but that the resulting flatter curves, wider lane and wider shoulder widths encourage drivers to go faster, exacerbating the possibilities for not being prepared for interactions with other vehicles. The safety benefits achieved in SV crash reduction by better geometric features may not be achievable for MV crashes. Agencies responsible for transportation should be aware of this situation – the improvement of geometric characteristics may not help reduce MV crashes. They need to make informed decisions as opposed to guesses about the influence of geometric elements on traffic accidents.

This phenomenon also calls for a further examination of the relationship between crash rate and geometric characteristics. Future study aimed at stratifying MV crashes into categories such as head-on, rear end and angle collisions may help explain the influence of geometric characteristics better. Models with geometric variables should be estimated and evaluated for each type of MV crash to learn which characteristics really contribute to MV crashes. We would not expect variables good for SV crash estimation such as lane width and shoulder width to also be good for predicting all types of MV crashes, though they might be good for head-on crashes. Instead, features such as number and type of driveways and intersections may play a more important role for other MV crashes. This suggests that we should concentrate on correcting the site conditions that contribute to the kinds of crashes experienced at a given highway location, rather than just upgrading the highway design blindly, hoping this will reduce MV crashes. This could also mean focusing on the land use environment as well as the highway design, or installing devices to control vehicle speeds, such as traffic calming.

ACKNOWLEDGMENT

This paper was presented at the 79th Annual Meeting of the Transportation Research Board in Washington DC, January 2000.

REFERENCES

1. Kalakota, K. R., Seneviratne, P. N., and Islam, M. N. Influence of Geometric Design Variables on Accident Rates on Two-lane Rural Highways. *VTI Rapport* **380A**: 207-226; 1992.
2. Haskin, A. F., Fearn, K. T., Kao, L., Miller, T., Porretta, K., Adams, and J., Munao, J. Accident Facts, 1995 Edition. National Safety Council, Statistics Department: 1995.
3. Ivan, J. N., Pasupathy, R. K., and Ossenbruggen, P. J. Differences in Causality Factors for Single and Multi-vehicle Crashes on Two-lane Roads. *Accident Analysis and Prevention* (in press); 1999.
4. Kulmala, R. Safety at Highway Junctions Based on Predictive Accident Models. *Third International Symposium on Intersections without Traffic Signals, Portland, Oregon U.S.A.*; 151-157; 1997.
5. Greenshields, B. D., and Weida, F. M. *Statistics with Applications to Highway Traffic Analysis.* The Eno foundation for highway traffic control, Connecticut, 1952.
6. Bernardo, N. R. and Ivan, J. N. Predicting Number of Crashes Versus Crash Rate using Poisson Regression. Presented at Transportation Research Board Annual Meeting, Washington DC, paper no. 980995; 1997.
7. S-PLUS User's Manual, Version 3.3 for Windows. Seattle: Statistical Science Division, Mathsoft, Inc.: 1995.
8. Vogt, A. and Bared, J. Accident Models for Two-lane Rural Segments and Intersections. *Transportation Research Record* **1635**: 18-29; 1998.
9. Highway Capacity Manual, 2nd Edition. Transportation Research Board Special Report; 1994.

10. Clayton, D., and Hills, M. *Stastical Models in Epidemiology.* Oxford University Press, 1993.
11. Lin, T. D., Jovanis, P. P. and Yang, C. Z. Time of Day Models of Motor Carrier Accident Risk. *Transportation Research Record* **1467**: 1-8; 1994.

Table 1. Aggregated Data Example

	segmt_id	cat_year	cat_time	lit_cond	cat_v_c	multi	main	cross	newmv
146	1	6	3	2	5	0	715	1	715
147	1	6	3	3	1	0	19326	142	19326
148	1	6	3	3	2	0	13785	59	13785
149	1	6	3	3	3	0	1987	5	1987
150	1	6	4	0	1	0	61953	1136	61953
151	1	6	4	0	2	0	200	1	200
152	1	6	4	2	1	0	10260	107	10260
153	1	6	4	2	2	0	1921	9	1921
154	1	6	4	3	1	0	13674	129	13674
155	1	6	4	3	2	0	2960	14	2960
156	1	6	5	0	1	0	23697	2443	23697
157	2	1	1	0	1	0	14367	2782.17	243820.3
158	2	1	1	0	2	0	1522	295.65	56281.12
159	2	1	1	1	1	0	21793	4173.64	408591.7
160	2	1	1	1	2	0	19824	3834.15	778672.7
161	2	1	1	2	1	0	86162	16517	2502796
162	2	1	1	2	2	0	56607	10817.45	2223632
163	2	1	1	2	3	0	1276	242.8	103975.7
164	2	1	2	2	1	0	100595	19291.85	3108643
165	2	1	2	2	2	0	266444	51127.75	11846343
166	2	1	2	2	3	0	4251	814	349474.9
167	2	1	2	2	4	0	2781	533.73	372078.7
168	2	1	2	2	5	0	778	148.04	115172.4
169	2	1	3	0	1	0	20879	4041.64	591910.2
170	2	1	3	0	2	1	16866	3293.83	740500.6

Table 2. Evaluating Potential Exposure Offsets								
	Model 1*		Model 2*		Model 3*		Model 4**	
	Value	t-value	Value	t-value	Value	t-value	Value	t-value
(Intercept)	**-14.15**	**-11.63**	**-14.83**	**-11.03**	**-10.85**	**-10.82**	**-10.85**	**-10.82**
$\ln(V_M)$	**0.84**	**9.58**	**0.97**	**6.68**				
$\ln(V_I)$			-0.13	-1.17				
$\ln(V_{MV})$					**0.53**	**8.96**	0.03	0.46
segment2 (Rt. 1)	**1.73**	**2.18**	**2.13**	**2.47**	0.09	0.11	0.09	0.11
segment3 (Rt. 1)	**3.91**	**5.13**	**4.47**	**4.96**	1.61	1.94	1.61	1.94
segment4 (Rt. 81)	0.07	0.08	-0.06	-0.06	0.42	0.47	0.42	0.47
segment5 (Rt. 81)	**1.95**	**2.48**	**2.44**	**2.74**	-0.28	-0.32	-0.28	-0.32
segment6 (Rt. 32)	**2.52**	**3.25**	**2.28**	**2.85**	**2.98**	**3.83**	**2.98**	**3.83**
segment7 (Rt. 124)	1.31	1.64	**1.76**	**1.99**	-0.96	-1.09	-0.96	-1.09
segment8 (Rt. 6)	**2.17**	**2.79**	**1.94**	**2.43**	**2.57**	**3.29**	**2.57**	**3.29**
segment9 (Rt. 6)	**2.57**	**3.32**	**2.96**	**3.52**	0.51	0.60	0.51	0.60
segment10 (Rt. 5)	**2.28**	**2.93**	**2.58**	**3.16**	0.84	1.02	0.84	1.02
segment11 (Rt. 12)	**2.10**	**2.70**	**2.51**	**2.94**	0.37	0.44	0.37	0.44
segment12 (Rt. 30)	**2.00**	**2.56**	**2.30**	**2.80**	0.60	0.73	0.60	0.73
segment13 (Rt. 8)	**1.97**	**2.45**	**2.37**	**2.72**	0.25	0.29	0.25	0.29
segment14 (Rt. 8)	-3.88	-1.21	-3.90	-1.22	-3.95	-1.19	-3.95	-1.19
segment15 (Rt. 66)	-1.43	-1.16	-1.58	-1.27	-1.17	-0.94	-1.17	-0.94
segment16 (Rt. 66)	**1.67**	**2.11**	**2.04**	**2.40**	-0.18	-0.21	-0.18	-0.21
segment17 (Rt. 217)	**1.73**	**2.19**	**2.28**	**2.48**	-0.77	-0.88	-0.77	-0.88
light (dusk)	-0.75	-1.69	-0.74	-1.68	**-1.24**	**-2.89**	**-1.24**	**-2.89**
light (dawn)	-0.35	-0.68	-0.34	-0.65	-0.68	-1.33	-0.68	-1.33
light (dark)	**-0.99**	**-2.53**	**-0.98**	**-2.50**	**-0.95**	**-3.02**	**-0.95**	**-3.02**
time (10am-3pm)	0.29	1.57	0.28	1.52	**0.51**	**2.92**	**0.51**	**2.92**
time (3pm-7pm)	**0.67**	**3.81**	**0.66**	**3.77**	**0.66**	**3.73**	**0.66**	**3.73**
time (7pm-11pm)	0.62	1.50	0.63	1.53	**0.74**	**2.18**	**0.74**	**2.18**
time (11pm-6am)	**0.86**	**2.02**	**0.91**	**2.15**	**1.30**	**3.63**	**1.30**	**3.63**
Dispersion Parameter	209		207		212		212	
Null Deviance (df)	574877	4166	574877	4166	574877	4166	497181	4166
Residual Deviance (df)	337960	4142	337665	4141	345133	4142	345133	4142
	Coefficients in shaded bold face are significant at 95%							
	* no offset							
	** offset is $V_{MV}^{1/2}$							

Table 3. Segregating New Exposure Offset by Time of Day and Light Condition								
	Model 3		Model 5		Model 6		Model 7	
	Value	t-value	Value	t-value	Value	t-value	Value	t-value
(Intercept)	**-10.85**	**-10.82**	**-10.44**	**-10.10**	**-10.39**	**-6.70**	**-10.89**	**-6.87**
ln(V)	**0.53**	**8.96**	**0.49**	**7.95**	**0.52**	**5.01**	**0.55**	**5.23**
ln($V_{10.3}$)					0.02	0.20		
ln($V_{3.7}$)					0.10	1.09		
ln($V_{7.11}$)					0.05	0.43		
ln($V_{11.6}$)					**0.41**	**3.13**		
ln($V_{DK6.10}$)							-0.33	-0.19
ln($V_{DK3.7}$)							0.05	0.17
ln($V_{DK7.11}$)							-0.02	-0.18
ln($V_{DK11.6}$)							**0.38**	**2.88**
ln($V_{DW6.10}$)							-0.31	-1.33
ln($V_{DY10.3}$)							-0.02	-0.20
ln($V_{DY3.7}$)							0.06	0.65
ln($V_{DY7.11}$)							0.66	0.76
ln($V_{DU3.7}$)							0.24	0.67
ln($V_{DU7.11}$)							0.99	1.03
segment2 (Rt. 1)	0.09	0.11	0.24	0.28	-0.15	-0.18	-0.13	-0.16
segment3 (Rt. 1)	1.61	1.94	**1.80**	**2.13**	1.30	1.50	1.32	1.57
segment4 (Rt. 81)	0.42	0.47	0.46	0.51	0.36	0.40	0.37	0.43
segment5 (Rt. 81)	-0.28	-0.32	-0.08	-0.09	-0.65	-0.72	-0.63	-0.72
segment6 (Rt. 32)	**2.98**	**3.83**	**3.03**	**3.85**	**2.89**	**3.70**	**2.91**	**3.82**
segment7 (Rt. 124)	-0.96	-1.09	-0.74	-0.83	-1.32	-1.43	-1.30	-1.45
segment8 (Rt. 6)	**2.57**	**3.29**	**2.61**	**3.31**	**2.51**	**3.19**	**2.52**	**3.30**
segment9 (Rt. 6)	0.51	0.60	0.72	0.83	0.15	0.17	0.17	0.19
segment10 (Rt. 5)	0.84	1.02	1.01	1.20	0.55	0.65	0.58	0.69
segment11 (Rt. 12)	0.37	0.44	0.55	0.65	0.08	0.09	0.10	0.12
segment12 (Rt. 30)	0.60	0.73	0.75	0.90	0.31	0.37	0.34	0.41
segment13 (Rt. 8)	0.25	0.29	0.37	0.43	0.03	0.04	0.06	0.07
segment14 (Rt. 8)	-3.95	-1.19	-3.93	-1.17	-3.90	-1.18	-3.92	-1.21
segment15 (Rt. 66)	-1.17	-0.94	-1.12	-0.89	-1.20	-0.96	-1.19	-0.98
segment16 (Rt. 66)	-0.18	-0.21	0.00	0.00	-0.54	-0.61	-0.52	-0.60
segment17 (Rt. 217)	-0.77	-0.88	-0.54	-0.61	-1.23	-1.33	-1.22	-1.35
light (dusk)	**-1.24**	**-2.89**						
light (dawn)	-0.68	-1.33						
light (dark)	**-0.95**	**-3.02**						
time (10am-3pm)	**0.51**	**2.92**						
time (3pm-7pm)	**0.66**	**3.73**						
time (7pm-11pm)	**0.74**	**2.18**						
time (11pm-6am)	**1.30**	**3.63**						
dark, 6am-10am			-4.37	-1.22	-4.34	-1.21	-0.14	-0.01
dark, 3pm-7pm			-0.67	-1.15	-1.99	-1.38	-1.36	-0.29
dark, 7pm-11pm			-0.19	-0.86	-0.92	-0.52	0.16	0.09
dark, 11pm-6am			0.27	1.27	**-5.17**	**-2.85**	**-4.67**	**-2.56**
dawn, 6am-10am			-0.76	-1.46	-0.72	-1.36	3.46	1.13
day, 10am-3pm			**0.53**	**2.95**	0.23	0.18	0.74	0.57
day, 3pm-7pm			**0.67**	**3.73**	-0.91	-0.66	-0.35	-0.25
day, 7pm-11pm			-0.45	-0.50	-1.06	-0.57	-9.53	-0.76
dusk, 3pm-7pm			-0.70	-1.33	-2.09	-1.46	-4.24	-0.77
dusk, 7pm-11pm			-0.44	-0.59	-1.08	-0.59	-14.54	-1.02
Dispersion Parameter	212		218		214		202	
Null Deviance (df)	574877	4166	574877	4166	574877	4166	574877	4166
Residual Deviance (df)	345133	4142	343936	4139	340872	4135	339860	4129
	* Coefficients in shaded bold face are significant at 95%							

Table 4. Paired T-Test for Light and Time in Model 3

	6am-10am	10am-3pm	3pm-7pm	7pm-11pm	11pm-6am	day	dusk	dawn	dark
6am-10am	0.00	**-3.40**	**-4.06**	**-2.21**	**-3.66**				
10am-3pm	**3.40**	0.00	-0.60	-0.52	-1.91				
3pm-7pm	**4.06**	0.60	0.00	-0.14	-1.54				
7pm-11pm	**2.21**	0.52	0.14	0.00	-1.12				
11pm-6am	**3.66**	1.91	1.54	1.12	0.00				
day						0.00	**3.07**	1.42	**3.17**
dusk						**-3.07**	0.00	-0.86	-0.60
dawn						-1.42	0.86	0.00	0.43
dark						**-3.17**	0.60	-0.43	0.00

Table 5. Paired T-Test for Light-Time combination in Model 5

	day, 6am-10am	night,6am-10am	night,3pm-7pm	night,7pm-11pm	night,11pm-6am	dawn,6am-10am	day,10am-3pm	day,3pm-7pm	day,7pm-11pm	dusk,3pm-7pm	dusk,7pm-11pm
day, 6am-10am	0	1.23	1.15	0.85	-1.48	1.45	**-3.17**	**-3.87**	0.49	1.33	0.58
night,6am-10am	-1.23	0.00	-1.03	-1.17	-1.30	-1.30	-1.37	-1.41	-1.07	-1.02	-1.08
night,3pm-7pm	-1.15	1.03	0.00	-0.76	-1.55	0.12	**-1.96**	**-2.21**	-0.22	0.04	-0.25
night,7pm-11pm	-0.85	1.17	0.76	0.00	-1.61	1.00	**-2.56**	**-3.04**	0.26	0.89	0.31
night,11pm-6am	1.48	1.30	1.55	1.61	0.00	1.88	-0.94	-1.50	0.79	1.75	0.93
dawn,6am-10am	-1.45	1.00	-0.12	-1.00	-1.88	0.00	**-2.34**	**-2.61**	-0.32	-0.08	-0.36
day,10am-3pm	**3.17**	1.37	**1.96**	**2.56**	0.94	**2.34**	0.00	-0.64	1.06	**2.21**	1.25
day,3pm-7pm	**3.87**	1.41	**2.21**	**3.04**	1.50	**2.61**	0.64	0.00	1.23	**2.48**	1.44
day,7pm-11pm	-0.49	1.07	0.22	-0.26	-0.79	0.32	-1.06	-1.23	0.00	0.26	0.00
dusk,3pm-7pm	-1.33	1.02	-0.04	-0.89	-1.75	0.08	**-2.21**	**-2.48**	-0.26	0.00	-0.29
dusk,7pm-11pm	-0.58	1.08	0.25	-0.31	-0.93	0.36	-1.25	-1.44	0.00	0.29	0.00

Table 6. Paired T-Test for Exposure Segregated by Time of Day in Model 6

	V6.10	V10.3	V3.7	V7.11	V11.6
V6.10	0.00	-0.20	-1.09	-0.43	**-3.13**
V10.3	0.20	0.00	-0.66	-0.24	**-2.55**
V3.7	1.09	0.66	0.00	0.30	**-1.99**
V7.11	0.43	0.24	0.31	0.00	**-2.03**
V11.6	**3.13**	**2.55**	**1.99**	**2.03**	0.00

* Coefficients in shaded bold face are significant at 95%

Table 7. Paired T-Test for Exposure Segregated by Combination of Light Condition and Time of Day in Model 7

	VDY6.10	VDK6.10	VDK3.7	VDK7.11	VDK11.6	VDW6.10	VDY10.3	VDY3.7	VDY7.11	VDU3.7	VDU7.11
VDY6.10	0.00	0.19	-0.17	0.18	**-2.88**	1.33	0.20	-0.65	-0.76	-0.67	-1.03
VDK6.10	-0.19	0.00	-0.22	-0.18	-0.40	-0.01	-0.18	-0.22	-0.51	-0.32	-0.66
VDK3.7	0.17	0.22	0.00	0.22	-0.94	0.93	0.21	-0.02	-0.66	-0.38	-0.92
VDK7.11	-0.18	0.18	-0.22	0.00	**-2.24**	1.10	-0.04	-0.54	-0.78	-0.69	-1.04
VDK11.6	**2.88**	0.40	0.94	**2.24**	0.00	**2.56**	**2.53**	**1.99**	-0.32	0.37	-0.63
VDW6.10	-1.33	0.01	-0.93	-1.10	**-2.56**	0.00	-1.19	-1.48	-1.08	-1.29	-1.32
VDY10.3	-0.20	0.18	-0.21	0.04	**-2.53**	1.19	0.00	-0.62	-0.77	-0.69	-1.04
VDY3.7	0.65	0.22	0.02	0.54	**-1.99**	1.48	0.62	0.00	-0.69	-0.48	-0.96
VDY7.11	0.76	0.51	0.66	0.78	0.32	1.08	0.77	0.69	0.00	0.45	-0.25
VDU3.7	0.67	0.32	0.38	0.69	-0.37	1.29	0.69	0.48	-0.45	0.00	-0.73
VDU7.11	1.03	0.66	0.66	1.04	0.63	1.32	1.04	0.96	0.25	0.73	0.00

* Coefficients in shaded bold face are significant at 95%

Table 8. Pruning Insignificant Factors								
	Model 7		Model 8		Model 9		Model 10	
	Value	t-value	Value	t-value	Value	t-value	Value	t-value
(Intercept)	**-10.89**	**-6.87**	**-11.14**	**-10.44**	**-11.84**	**-12.39**	**-9.89**	**-19.09**
NLNEWMV	**0.55**	**5.23**	**0.57**	**8.34**	**0.62**	**10.12**	**0.48**	**14.58**
MVDK6.10	-0.33	-0.19						
MVDK3.7	0.05	0.17						
MVDK7.11	-0.02	-0.18						
MVDK11.6	**0.38**	**2.88**	**0.38**	**3.41**	**0.41**	**3.92**	0.12	1.46
MVDW6.10	-0.31	-1.33						
MVDY10.3	-0.02	-0.20						
MVDY3.7	0.06	0.65						
MVDY7.11	0.66	0.76						
MVDU3.7	0.24	0.67						
MVDU7.11	0.99	1.03						
segment2 (Rt. 1)	-0.13	-0.16	-0.23	-0.27	-0.46	-0.58		
segment3 (Rt. 1)	1.32	1.57	1.21	1.4	0.92	1.14		
segment4 (Rt. 81)	0.37	0.43	0.34	0.38	0.28	0.33		
segment5 (Rt. 81)	-0.63	-0.72	-0.74	-0.83	-1.05	-1.26		
segment6 (Rt. 32)	**2.91**	**3.82**	**2.86**	**3.66**	**2.78**	**3.77**	**2.48**	**13.89**
segment7 (Rt. 124)	-1.30	-1.45	-1.44	-1.57	**-1.76**	**-2.06**		
segment8 (Rt. 6)	**2.52**	**3.30**	**2.47**	**3.15**	**2.41**	**3.26**	**2.06**	**10.3**
segment9 (Rt. 6)	0.17	0.19	0.04	0.04	-0.28	-0.34		
segment10 (Rt. 5)	0.58	0.69	0.46	0.54	0.21	0.26		
segment11 (Rt. 12)	0.10	0.12	-0.02	-0.02	-0.29	-0.36		
segment12 (Rt. 30)	0.34	0.41	0.23	0.27	-0.01	-0.01		
segment13 (Rt. 8)	0.06	0.07	-0.04	-0.04	-0.23	-0.28		
segment14 (Rt. 8)	-3.92	-1.21	-3.89	-1.18	-3.91	-1.26		
segment15 (Rt. 66)	-1.19	-0.98	-1.23	-0.98	-1.29	-1.09		
segment16 (Rt. 66)	-0.52	-0.60	-0.64	-0.73	-0.92	-1.12		
segment17 (Rt. 217)	-1.22	-1.35	-1.34	-1.46	**-1.69**	**1.97**		
dark, 6am-10am	-0.14	-0.01	-4.2	-1.18				
dark, 3pm-7pm	-1.36	-0.29	-0.56	-0.98				
dark, 7pm-11pm	0.16	0.09	-0.15	-0.66				
dark, 11pm-6am	**-4.67**	**-2.56**	**-4.67**	**-3.15**	**-4.81**	**-3.46**	-1.22	-1.06
dawn, 6am-10am	3.46	1.13	-0.65	-1.24				
day, 10am-3pm	0.74	0.57	**0.44**	**2.43**	**0.54**	**3.55**	**0.8**	**6.14**
day, 3pm-7pm	-0.35	-0.25	**0.61**	**3.38**	**0.73**	**4.83**	**0.99**	**7.25**
day, 7pm-11pm	-9.53	-0.76	-0.28	-0.32				
dusk, 3pm-7pm	-4.24	-0.77	-0.63	-1.22				
dusk, 7pm-11pm	-14.54	-1.02	-0.3	-0.41				
Dispersion Parameter	202		214		190		210	
Null Deviance (df)	574877	4166	574877	4166	574877	4166	574877	4166
Residual Deviance (df)	339860	4129	341409	4138	343181	4145	413608	4159
	* Coefficients in shaded bold face are significant at 95%							

Table 9. Paired T-Test for Light-Time Combinations in Model 8

	day, 6am-10am	night,6am-10am	night,3pm-7pm	night,7pm-11pm	night,11pm-6am	dawn,6am-10am	day,10am-3pm	day,3pm-7pm	day,7pm-11pm	dusk,3pm-7pm	dusk,7pm-11pm
day, 6am-10am	0.00	1.18	0.98	0.66	**3.15**	1.24	**2.43**	**3.38**	0.32	1.22	0.41
night,6am-10am	-1.18	0.00	-1.01	-1.14	0.12	-0.99	-1.31	-1.35	-1.07	-0.99	-1.08
night,3pm-7pm	-0.98	1.01	0.00	-0.67	**2.58**	0.11	-1.65	-1.94	-0.27	0.09	-0.28
night,7pm-11pm	-0.66	1.14	0.67	0.00	**3.01**	0.88	**-2.04**	**-2.63**	0.15	0.86	0.20
night,11pm-6am	**-3.15**	-0.12	**-2.58**	**-3.01**	0.00	**-2.56**	**-3.42**	**-3.54**	**-2.55**	**-2.57**	**-2.63**
dawn,6am-10am	-1.24	0.99	-0.11	-0.88	**2.56**	0.00	**-1.97**	**-2.28**	-0.36	-0.02	-0.38
day,10am-3pm	**2.43**	1.31	1.65	**2.04**	**3.42**	**1.97**	0.00	-0.68	0.80	**1.96**	0.97
day,3pm-7pm	**3.38**	1.35	1.94	**2.63**	**3.54**	**2.28**	0.68	0.00	1.00	**2.27**	1.19
day,7pm-11pm	-0.32	1.07	0.27	0.15	**2.55**	0.36	-0.80	-1.00	0.00	0.35	0.02
dusk,3pm-7pm	-1.22	0.99	-0.09	-0.86	**2.57**	0.02	**-1.96**	**-2.27**	-0.35	0.00	-0.37
dusk,7pm-11pm	-0.41	1.08	0.28	-0.20	**2.63**	0.38	-0.97	-1.19	-0.02	0.37	0.00

* Coefficients in shaded bold face are significant at 95%

Table 10. Paired T-Test for Site Dummies in Model 9

	seg. 1	seg. 2	seg. 3	seg. 4	seg. 5	seg. 6	seg. 7	seg. 8	seg. 9	seg. 10	seg. 11	seg. 12	seg. 13	seg. 14	seg. 15	seg. 16	seg. 17
segment 1 (Rt. 7)	0.00	0.58	-1.14	-0.33	1.26	**-3.77**	**2.06**	**-3.26**	0.34	-0.26	0.36	0.01	0.28	1.26	1.09	1.12	**1.97**
segment 2 (Rt. 1)	-0.58	0.00	-1.22	-0.64	0.51	**-2.98**	1.11	**-2.64**	-0.16	-0.60	-0.15	-0.40	-0.21	1.07	0.58	0.40	1.04
segment 3 (Rt. 1)	1.14	1.22	0.00	0.55	1.71	-1.71	**2.29**	-1.36	1.01	0.63	1.07	0.83	1.01	1.50	1.55	1.60	**2.22**
segment 4 (Rt. 81)	0.33	0.64	-0.55	0.00	1.12	**-2.23**	1.70	-1.89	0.47	0.06	0.49	0.25	0.44	1.30	1.08	1.02	1.64
segment 5 (Rt. 81)	-1.26	-0.51	-1.71	-1.12	0.00	**-3.45**	0.60	**-3.11**	-0.66	-1.10	-0.65	-0.90	-0.71	0.89	0.17	-0.11	0.54
segment 6 (Rt. 32)	**3.77**	**2.98**	1.71	**2.23**	**3.45**	0.00	**4.04**	0.36	**2.77**	**2.38**	**2.82**	**2.58**	**2.77**	**2.09**	**2.93**	**3.36**	**3.96**
segment 7 (Rt. 124)	**-2.06**	-1.11	**-2.29**	-1.70	-0.60	**-4.04**	0.00	**-3.70**	-1.25	-1.69	-1.25	-1.51	-1.32	0.67	-0.33	-0.71	-0.06
segment 8 (Rt. 6)	**3.26**	**2.64**	1.36	1.89	**3.11**	-0.36	**3.70**	0.00	**2.43**	**2.03**	**2.47**	**2.23**	**2.42**	**1.98**	**2.66**	**3.02**	**3.63**
segment 9 (Rt. 6)	-0.34	0.16	-1.04	-0.47	0.66	**-2.77**	1.25	**-2.43**	0.00	-0.43	0.01	-0.24	-0.05	1.13	0.70	0.55	1.19
segment 10 (Rt. 5)	0.26	0.60	-0.63	-0.06	1.10	**-2.38**	1.69	**-2.03**	0.43	0.00	0.45	0.20	0.39	1.28	1.05	0.99	1.63
segment 11 (Rt. 12)	-0.36	0.15	-1.07	-0.49	0.65	**-2.82**	1.25	**-2.47**	-0.01	-0.45	0.00	-0.25	-0.06	1.13	0.70	0.54	1.19
segment 12 (Rt. 30)	-0.01	0.40	-0.83	-0.25	0.90	**-2.58**	1.51	**-2.23**	0.24	-0.20	0.25	0.00	0.19	1.22	0.90	0.80	1.44
segment 13 (Rt. 8)	-0.28	0.21	-1.01	-0.44	0.71	**-2.77**	1.32	**-2.42**	0.05	-0.39	0.06	-0.19	0.00	1.15	0.75	0.61	1.25
segment 14 (Rt. 8)	-1.26	-1.07	-1.50	-1.30	-0.89	**-2.09**	-0.67	-1.98	-1.13	-1.28	-1.13	-1.22	-1.15	0.00	-0.79	-0.93	-0.69
segment 15 (Rt. 66)	-1.09	-0.58	-1.55	-1.08	-0.17	**-2.93**	0.33	**-2.66**	-0.70	-1.05	-0.70	-0.90	-0.75	0.79	0.00	-0.26	0.28
segment 16 (Rt. 66)	-1.12	-0.40	-1.60	-1.02	0.11	**-3.36**	0.71	**-3.02**	-0.55	-0.99	-0.54	-0.80	-0.61	0.93	0.26	0.00	0.65
segment 17 (Rt. 217)	**-1.97**	-1.04	**-2.22**	-1.64	-0.54	**-3.96**	0.06	**-3.63**	-1.19	-1.63	-1.19	-1.44	-1.25	0.69	-0.28	-0.65	0.00

* Coefficients in shaded bold face are significant at 95%

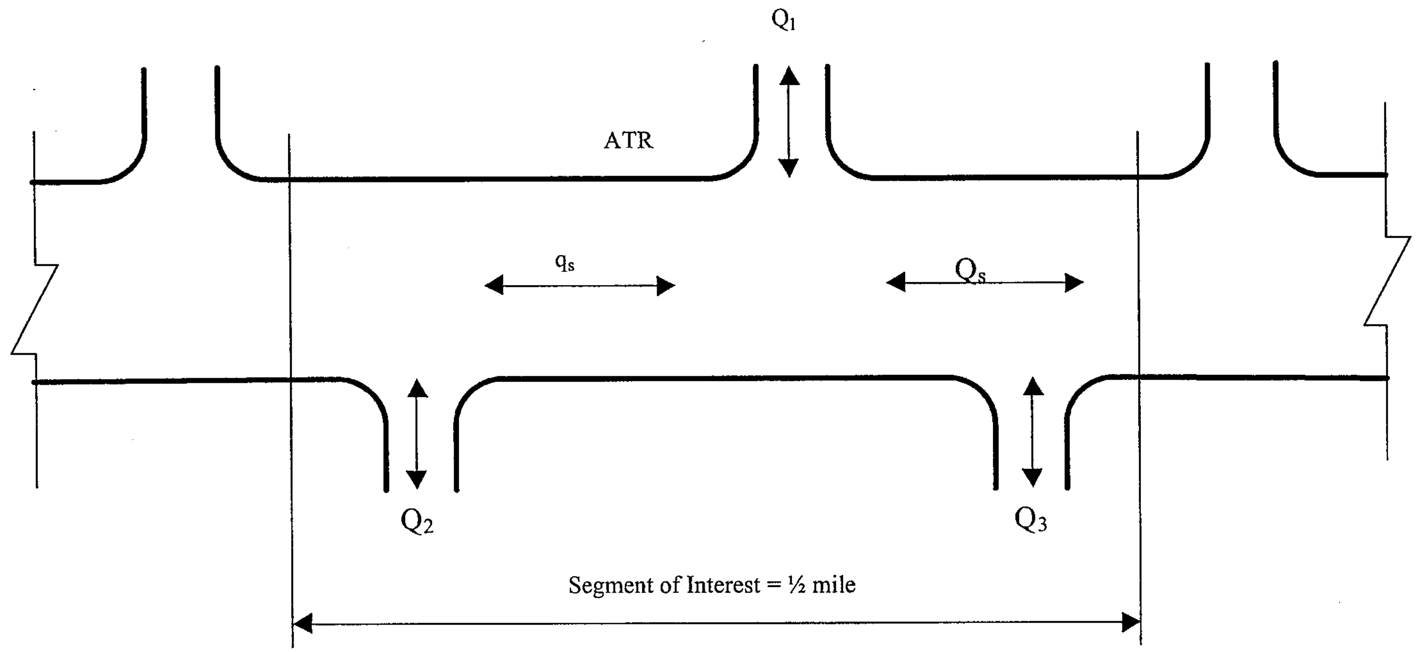

Figure 1. Illustration of traffic volumes on a study site

PART III: The Impacts of Using a Safety Compliance Standard in Highway Design

INTRODUCTION

To objectively test and evaluate safety, I introduce a compliance standard into the highway design process. The principle of individual lifetime risk is used to establish the standard.[1] A highway design location or site s, is defined to be operating at an acceptable risk when an individual's chance of being involved in a fatal crash over a lifetime of motor vehicle travel at s is equal to or less than 1 in 1000, or $\theta^* = 10^{-3}$. Site s is defined "hazardous" if it fails to meet this criterion.

The highway design process as presented in this paper uses a design algorithm derived from basic concepts of:

- highway design, or level of service (*LOS*) considerations,
- risk analysis, or the principle of individual lifetime risk, and
- statistical modeling.

The algorithm will be described and a case study will demonstrate how it is applied and the design algorithm will be critiqued.

OVERVIEW OF THE DESIGN ALGORITHM

The design algorithm is formulated as a constrained optimization model using well-established principles of traffic flow and accepted highway design guidelines:

Objective:	*maximize*	$\bar{u}$		(1)
Constraint set:	*subject to*	$\pi \leq \varpi$	(fatal)	(2)
		$\pi_I \leq \varpi_I$	(injury)	(3)
		$\pi_P \leq \varpi_P$	(property damage)	(4)

The objective is to maximize average operating speed $\bar{u}$ because speed is considered to be a most important *LOS* measure in design. "Speed and travel time are fundamental measurements of traffic performance of the existing highway system, and speed is the key variable in the redesign or design of new facilities."[2] "Except for local streets where speed controls are included intentionally, every effort should be made to use as high a design speed as practical to attain a desired degree of safety, mobility and efficiency while under the constraints of environmental quality, economics, esthetics and social and political impacts."[3]

A design is considered safe when the safety compliance constraint set is satisfied. That is, the predicted crash probabilities resulting in fatality π, injury π_I and property damage π_P are less than or equal to the corresponding compliance probabilities for fatality ϖ, injury ϖ_I and property damage ϖ_P.

---

[1] Paul J. Ossenbruggen, ôA Method of Identifying Hazardous Location Using the Principle of Individual Lifetime Riskö, **9 Risk: Health, Safety & Environment 83**, (winter 1998).

[2] Adolf D. May, **Traffic Flow Fundamentals** 116 (1990).

[3] Am. Ass'n State Highway and Transportation Officials (AASHTO), **A Policy on Geometric Design of Highway and Streets** 62 (1994).

In this paper, logistic regression is used to calibrate a crash prediction model for injury π_I. The data set comprised of police accident reports, traffic volume and speed records for a six-year period at eight different, undivided two-lane highways in urban and rural Connecticut. The constrained optimization model for injury is:

$$\textit{maximize} \quad \bar{u} \tag{1}$$

$$\textit{subject to} \quad \pi_I \leq \varpi_I. \tag{3}$$

The model development, discussion and case study are focused on this model. Each of the $\bar{u}$, ϖ_I and π_I models are presented in turn.

THE AVERAGE OPERATING SPEED MODEL

The objective to maximize average operating speed $\bar{u}$ is the concept used by highway designers. Average operating speed $\bar{u}$ is a function of the free-flow speed u_f measured in miles per hour (mph), traffic flow v measured in vehicles per hour (vph) and highway capacity c (vph). It is calculated as:

$$\bar{u} = 0.5u_f\left(1 \pm \sqrt{1 - v/c}\right) \tag{5}$$

The equation is derived from Greenshield's linear speed-density model and flow-density-speed relationship.[4]

Figure 1 shows that the average operating speed $\bar{u}$ model with u_f = 60 mph and c = 2,800 vph does a nice job of representing the Highway Capacity Manual[5] *LOS* letter rating system for a two-lane, undivided highway under ideal traffic conditions. An ideal condition is passenger cars traveling at an average operating speed of no less than 60 mph on level terrain with a 100% passing zone and with a 50/50 directional traffic flow split. A 50/50 split means that there are an equal number of passenger cars in each lane.

An ideal two-lane, undivided highway has a bidirectional flow capacity of c = 2,800 vph. If one or more conditions are not met, then the capacity is reduced. Adjustments are made for grades > 3 percent, directional distributions other than a 50/50, heavy vehicle usage, lane widths < 12 ft, and shoulder widths < 6 ft.

Design Optimization: For design optimization, free-flow speed u_f is used as the control variable. A solution satisfying the conditions of the optimization model, *maximizing* $\bar{u}$ *subject to:* $\pi_I \leq \varpi_I$, is designated as an optimum solution u_f^*. For design, u_f is used as a design specification and u_f^* refers to a design specification that satisfies the objective function and the safety compliance constraint.

---

[4] See Nicholas A. Garber & Lester S. Hoel, **Traffic And Highway Engineering,** 184-85 (1997). Greenshield's Model is $\bar{u} = u_f\left(1 - \frac{k}{k_j}\right)$ where k = traffic density in vehicles per mile (vpm) and k_j = jam density (vpm). The flow-speed-density relationship is $v = u.k$.

[5] Transportation Research Board, **Highway Capacity Manual** (HCM), Special Report 209. National research Council, Washington, DC. (1994).

The free-flow speed u_f is a function of driver sight distance as determined by horizontal and vertical roadway curvature, right-of-way dimension, lane and shoulder widths, or in other words, the *geometric alignment of the highway*.[6] The design specification u_f affects average operating speed $\bar{u}$ and highway capacity
c. For example, the average speeds for two highways designed for u_f = 60 and 45 mph given the same traffic flow $v = 1000$ vph are $\bar{u}$ = 54 and 49 mph, respectively.

Highway Capacity: Using Greenshield's linear speed-density relationship, highway capacity is calculated as:

$$c = {u_f \cdot k_j}/{4} \tag{6}$$

where vehicles $k_j =$ jam traffic density measured in vehicles per mile (vpm). Given c = 2,800 and u_f = 60 mph, the jam density for ideal conditions is estimated to be k_j = 187 vpm. This value of jam density k_j is assumed to be the same for all design specifications u_f for both ideal and non-ideal traffic conditions.

For example, given k_j = 187 vpm for the two design specifications above, highway capacities are calculated to be c = 2,100 and 2,800 vph for u_f = 45 and 60 mph, respectively.[7] Furthermore with k_j assumed to be a constant, the highway capacity formula reduces to a linear function, $c = 48.5 \cdot u_f$.

Speed Maximization: Given a traffic flow v, the solid line speeds in Figure 1 are calculated as $\bar{u} = 0.5 \cdot u_f\left(1+\sqrt{1-{v}/{c}}\right)$ and the broken line speeds are calculated as $\bar{u}_F = 0.5 \cdot \left(1-\sqrt{1-{v}/{v}}\right)$. Since $\bar{u} > \bar{u}_F$ and the design objective is to *maximize* $\bar{u}$, the solutions given by $\bar{u}_F$ are not of interest for design optimization. Furthermore, since $c = 48.5 \cdot u_f$, the objective function is written as the function of the control variable u_f exclusively:

$$maximize\ \bar{u} = 0.5 \cdot u_f\left(1+\sqrt{1-{v}/{48.5 \cdot u_f}}\right) \tag{7}$$

This function draws attention to the fact that the u_f specification directly affects the highway speed and capacity.

---

[6] In this paper, specific values of the geometric factors will not be given. The important point is that the value of u_f can be achieved by specifying one or more geometric alignment factors. For example, if u_f < 60 mph, then a particular value of u_f can be obtained by specifying a lane width < 12 ft or by specifying a combination of lane width < 12 ft and shoulder width < 6 ft. Of course, other combinations can also lead to the desired value of u_f.

[7] The Highway Capacity Manual uses adjustment factors to adjust the capacity c for non-ideal conditions. Consider a design specification u_f^* = 45 mph. An adjustment factor for narrow lanes and restricted shoulder width is f_w = 0.75. The capacity is $c = 2800 \cdot f_w$ = 2,100 vph.

Safety Considerations: Driver convenience and speed are often sacrificed by reducing the average operating speed $\bar{u}$. Theoretically, this can be achieved through (1) the geometric alignment of the highway and through (2) speed limit control.

Geometric Alignment and Traffic Calming: Highway designers and planners are under pressure to construct a high-speed highway system because there is an insatiable worldwide desire for mobility.[8] The construction of "big roads", that is, wide, straight roads with geometric alignment to maximize driver sight distance are favored.

The design algorithm puts less emphasis on mobility (reducing congestion and delay [9]) for purposes of improved highway safety and more emphasis on controlling speed through geometric design. It permits the use of narrow highway lanes and reduced sight distance to control speed by forcing drivers to slow down.

According to traffic calming advocates, "Bigger roads increase people's mobility" is a myth.[10] They claim that straight, wide roads encourage greater speed and encourage motorists to take greater risks. They show that measures to force drivers to slow down are effective in controlling speed and reducing injury and fatal crashes in cities and residential communities. They claim that the crash toll can be reduced by more than 40 percent when traffic-calming methods are implemented.

The design algorithm and the traffic calming methods have a similar goal, but the manner in which the goal is achieved differ. Traffic calming measures are generally employed in residential communities where quality of life from high-speed traffic is threatened. Roads are calmed by employing geometric alignment and other techniques, such as speed tables, chicanes, neck-downs and interrupted sight lines, to impede a motorist's ability to speed. These impediments dramatically reduce the average operating speed, so much so, as to cause traffic diversion. Of course, this is achieved through purposeful design. Through traffic motorists are encouraged to find alternate roadways; therefore, both speed $\bar{u}$ and traffic volume v are reduced.

The geometric alignment methods contemplated for the design algorithm reduce speed less dramatically without causing traffic diversion. Furthermore, introducing traffic diversion into the design algorithm would greatly complicate the mathematics.

Speed Limit Control: Speed limit control is often employed to reduce speed $\bar{u}$ at sites where crashes occur due to excessive speed or where excessive speed is considered a hazard. For example, it is not uncommon to observe a highway with a posted speed limit $s_P = 30$ mph with a design specification $u_f = 60$ mph. Clearly, the speed restriction would be unnecessary if the

---

[8] Andreas Schafer and David Victor, "The Past and Future of Global Mobility", **Scientific American,** (October 1997) pages 58 - 61.

[9] Congestion and delay often accompanied with a long waiting line are a possibility when the traffic flow v approaches the highway capacity c. In Figure where $u_f = 60$ mph and when v approaches $c =$ 2,800 vph, the average operating speed is about one-half u_f or $\bar{u} \cong 30 mph$. Under the same condition for $u_f = 45$ mph, v approaches $c = 2{,}100$ vph at $\bar{u} \cong 23 mph$. Either equation 5 or 7 can be used to calculate $\bar{u}$ for $u_f = 45$ mph.

[10] Citizens Advocating Responsible Transportation (CART), **Traffic Calming, The Solution to Urban Traffic and a New Vision for Neighborhood Livability**, Sensible Transportation Options for People (STOP), 15405 S.W. 116th Avenue #202B, Tigard OR 997224-2600. (1989) Page 12.

highway was safe at operating speeds that approach the design specification speed u_f. The signage is an attempt to control the average operating speed $\bar{u}$ for a highway design specification deemed to be too fast and hazardous for site *s*.

In the context of design optimization, this design specification given in the example does not satisfy the safety compliance constraint $\pi_I \leq \varpi_I$ for u_f = 60 mph. It is an infeasible solution; therefore, $u_f^* \neq u_f = 60mph$. The speed limit control method, while used in practice, is not applicable to the philosophy or methods espoused. The aim of this paper is to design a safe highway the first time, and to avoid the use of corrective traffic control schemes and costly roadway reconstruction at hazardous sites.

THE ALLOWABLE SAFETY LIMIT MODEL

The allowable limit for fatality crashes ϖ is determined from the individual lifetime risk model[11],

$$\theta = 1 - \exp(-70 \cdot \eta \cdot \varpi) \qquad (8)$$

and the assignment of an acceptable lifetime risk $\theta = 10^{-3}$. The annual exposure η = 664 trips per year per person. Given these assignments, the allowable limit for fatal crashes is calculated to be $\varpi = 2.2 \times 10^{-8}$. Given that one in about 55 serious injury crashes result in death, the allowable limit for injury crashes is $\varpi_I = 1.2 \times 10^{-6}$. The assignment of ϖ_P is based on property damage costs and is independent of individual lifetime risk considerations; therefore, it is outside the scope of this paper.

Figure 2 shows the effect of annual trip exposure η on ϖ_I for a constant acceptable risk equal to $\theta = 10^{-3}$. If incentives to travelers to reduce the annual individual trip exposure η can be found, then ϖ_I can be relaxed. In other words, the allowable limit of $\varpi_I = 1.2 \times 10^{-6}$ can be increased. In the U.S. exposure η is increasing among a growing driver population. If an acceptable lifetime risk equal to $\theta = 10^{-3}$ is to remain constant over time, then ϖ_I should be decreased to account for the increased individual exposure to highway risk.

AN INJURY CRASH PREDICTION MODEL

Traffic volume counts, police accident reports and other descriptive materials for eight sites in Connecticut for a period from 1990 to 1995 formed a data set for model calibration and validation. Each site listed in Tables 1 and 2 are locations of continuous traffic counting stations that have been grouped by posted speed limit s_p. The characteristics given under the headings of Land Use, Traffic Control and Geometric Design Factors in Table 2 show a variety of land use and roadside activity adjacent to the sites and highway designs located in rural and urban areas in Connecticut.

Exploratory Data Analysis: Exploratory data analysis is an intuitive and effective means to identify patterns and trends in the data and it often helps to identify statistically significant factors prior to performing model calibration and validation testing.

---

[11] See *supra* note 1. Page 86. The model was derived from basic principles of probability using the geometric and Poisson distributions. A premature death is considered to be a person who dies before the age of 70 years. According to the National Personal Transportation Survey, in 1990 a person made an average of 664 trips per year.

The total, injury and fatal crashes and the estimates of the probability of crashes resulting in property damage $\tilde{\pi}_P$ and in injury $\tilde{\pi}_I$ are given in Table 1. The probability $\tilde{\pi}_I$ is estimated as the ratio of the number of injury crashes to annual trip count.[12] Similarly, $\tilde{\pi}_P$ is estimated as the ratio of the number of property damage crashes to annual trip count. For example, the estimates for Hebron are $\tilde{\pi}_I = 10/1.3x10^{-6} = 7.7\ x\ 10^{-6}$ and $\tilde{\pi}_P = (23 - 10)/1.3x10^{-6} = 10.x10^{-6}$, respectively.

If $\tilde{\pi}_I$ estimates are used to rank sites, Hebron is the most hazardous location because it has the largest $\tilde{\pi}_I$ value in Table 1. Darien and East Windsor with the two highest annual traffic volumes and Darien and Waterford with the maximum number of injury and total crash counts are relatively safe when comparing their $\tilde{\pi}_I$ values to other sites. Annual trip count, a measure of exposure to highway risk at a site s, plays a critical role in the $\tilde{\pi}_I$ calculation and also in the site ranking.

Speed Limit Control: When comparing speed limit groups, the total, injury and fatal crashes and the values of $\tilde{\pi}_I$ and $\tilde{\pi}_P$ in Table 1 are the largest at sites where the most restrictive speed limit controls are used. *Highway risk is the greatest at sites with posted speed limits of 35 and 40 mph than at sites with the least restrictive speed limit of 45 mph.*

Land Use and Roadside Activity: The characteristics given in Table 2 show a diverse set of land use, traffic control and geometric design characteristics for the sites in each of the speed limit groups. However, no single characteristic seems to stand out in explaining why one site has a greater crash probability than another one. Sorting the data set in different ways and using contingency tables and scatter plots proved revealing.

The contingency table, Table 3, suggests that time-of-day and *LOS* rating may be important explanatory variables. Comparing $\tilde{\pi}_I$ and $\tilde{\pi}_P$ values in the two time-of-day categories by the same *LOS* rating show a pattern that suggests that there is a greater chance of being involved in a crash during dusk than at any other time period of the day.

Comparing $\tilde{\pi}_P$ and $\tilde{\pi}_I$ values by *LOS* rating within the dawn, day and night category and to a lesser degree within the dusk category suggests that the probability of being in a crash is dependent on *LOS* rating. Travelers experiencing driving conditions rated as *LOS* A and B, are more likely to be involved in a crash than at poorer *LOS* ratings. *This suggests that the average operating speed $\bar{u}$ is related to the crash probability.*

Logit scatter plots, which are not shown in this paper, suggest location s, posted speed limit s_P, time-of-day t and shoulder width may be significant explanatory variables. A logit is calculated as the natural logarithm of the ratio of the number of injury crashes to annual trip count or $\log[\hat{\pi}_I]$. The scatter plot for shoulder width suggests that shoulder widths of three feet or more tend to reduce the probability of a crash resulting in injury.

The exploratory data analyses suggest that:

- *LOS* rating or vehicular speed is an important factor in explaining the number of crashes,
- a posted speed limit has little effect in minimizing highway risk, and
- time-of-day and shoulder width may be important factors in predicting crash probability.

Modeling Calibration Results: LOS rating, expressed as capacity utilization *v/c*, posted speed limit s_P, time period t and the characteristics listed under the headings of Table 2 were

[12] Annual trip count is treated as a measure of highway risk exposure.

introduced as candidate variables in logistic regression model calibration and testing. The method of maximum likelihood was used to calibrate models and to estimate the variances and covariances of their model parameters. Models were tested using the likelihood-ratio (Wilk's statistic) and Wald tests.[13]

The following crash prediction model[14] satisfied validation testing:

$$\pi_I = \frac{\exp\left[-8.34 - 0.12 \cdot s_P - 0.34 \cdot t - 1.36 \cdot {}^{v}\!/_{c}\right]}{1 + \exp\left[-8.34 - 0.12 \cdot s_P - 0.34 \cdot t - 1.36 \cdot {}^{v}\!/_{c}\right]} \quad (9)$$

The time period variable t is a discrete variable where $t = -1$ for dusk and $t = 1$ for dawn, day or night (D/D/N). The variables s_P and v/c are continuous variables with ranges of $35 \le s_P \le 45$ mph and $0 \le v/c \le 1$, respectively. All model parameters are significant at $\alpha = 5\%$.

Shoulder width, which showed promise in the exploratory data analyses, when treated as a continuous variable was insignificant at $\alpha = 5\%$. When introduced as a discrete variable, it proved to be a significant variable; however, the model was considered unsuitable for the general concepts presented in this paper.

Model Properties: For purposes of crash prediction and highway design, a model should, at minimum, be a function of variables reflecting the travel demand, land use and roadside activity, and geometric design features at the site s. The crash prediction model π_I satisfies these minimum requirements with the following variables serving various purposes:

- v, a travel demand input parameter,
- s_P, a surrogate land use and roadside activity variable,
- c, a principle design variable,
- v/c, a measure of design performance *LOS*, and
- t, an indicator signifying that crash probability is a function of time-of-day.

Travel Demand: The affect of travel flow v on π_I is most easily discerned with the plots given in Figures 3 and 4. The major difference in the two figures is the designation of s_P. Comparing the two plots denoted dusk indicates that the probability of a crash is larger at a site designated $s_P = 35$ mph than the one designated $s_P = 45$ mph. The same relationship holds for plots denoted dawn, day and night.

The safety compliance constraint, $\pi_I \le \varpi_I$, is satisfied for all traffic volumes v except at dusk for $v < 500$ vph for site designation $s_P = 45$ mph shown in Figure 3. The safety compliance constraint is violated at dusk for all v and for dawn, day or night when $v < 1{,}500$ vph for site designation $s_P = 35$ mph. Clearly, the highway risk is greatest at a site designated $s_P = 35$ mph than at a site designated $s_P = 45$ mph shown in Figure 4.

---

13 Alan Agresti, **Categorical Data Analysis**, Wiley Interscience. (1990) pages 112-117.

14 Kopl Halperin, *A Comparative Analysis of Six Methods for Calculating Travel Fatality Risk*, **4 Risk: Health, Safety & Environment 14**, (Winter 1993). Traffic engineers report fatality rate in the number of fatalities per vehicle miles traveled (*VMT*). *VMT* is considered to be an inappropriate measurement for public health hazards.

The Surrogate Land Use and Roadside Activity Variable: The variable s_P is an indicator of how hazardous it is to drive at a site s. As a result, the variable s_P is considered to be a site characteristic variable.

Interpreting s_P to be a traffic control measure leads to the claim that the probability of an injury crash will decrease by increasing the posted speed limit s_P at a site s. Of course, this is nonsense and a naive claim. Speed limits are imposed to reduce the probability of a crash not to increase it. The only meaningful interpretation is that more restrictive speed limits are imposed at more hazardous sites. A site designated s_P = 35 mph has a greater highway risk than a site designated s_P = 45 mph and likewise, sites designated s_P = 35 and 40 mph have greater risk than a site designated s_P = 45 mph.

The s_P variable is not considered as traffic control measure in the crash prediction model. In fact, the model suggests that a posted speed limit is an ineffective means of improving highway safety. This finding is consistent with the findings of a study of raising and lowering posted speed limits on 83 comparison sites over increments of 5, 10, 15 and 20 mph. The signs had no practical significance in controlling speed.[15]

LOS : As the v increases, both average operating speed $\bar{u}$ and crash probability π_I decrease. This result suggests that a loss in *LOS* is coupled with an improvement in highway safety. Stated another way, it suggests that $\bar{u}$ and π_I are positively correlated. More simply stated, *faster speed is associated with greater highway risk.*

Time-of-day Considerations: An individual is not exposed to the same travel volumes each hour of the day, h = 1, 2, 3, ... , 24. Traffic flow varies by hour of the day. A key point in the design algorithm considers this range of hourly traffic volume v_h exposure with the use of marginal and condition probabilities.

The constrained optimization model for injury written as a function free-flow speed u_f becomes:

$$\text{Maximize} \qquad \bar{u} = \sum_h p_h \cdot \bar{u}_h \tag{10}$$

$$\text{subject to} \qquad \pi_I = \sum_h p_h \cdot \pi_{I/h} \leq \varpi_I \tag{11}$$

where the conditional probability for average operating speed given hour h is:

$$\bar{u}_h = 0.5 \cdot u_f \left(1 + \sqrt{1 - {}^{v_h}\!\!\big/_{48.5 \cdot u_f}} \right), \tag{12}$$

the conditional crash probability for injury given hour h is:

$$\pi_I = \frac{\exp\left[-8.34 - 0.12 \cdot s_P - 0.34 \cdot t - 0.028 \cdot {}^{v}\!/_{c}\right]}{1 + \exp\left[-8.34 - 0.12 \cdot s_P - 0.34 \cdot t - 0.028 \cdot {}^{v}\!/_{c}\right]}, \tag{13}$$

[15] Federal Highway Administration, "Effects of Raising and Lowering Speed Limits on Selected Roadway Sections", Publication Number FHWA-RD-92-84. (January,1997)

and p_h = probability that an individual is traveling in hour h. The values of p_h are estimated to be the ratio of the hourly to daily traffic counts, $p_h = v_h \big/ \sum_h v_h$. The summation can be interpreted as the average daily traffic (*ADT*).

Consequently, the design process using marginal probabilities considers all hours of the day, incorporating among other things the effects of (1) high speed on risk π_I and (2) high traffic volume on *LOS* as measured by $\bar{u}$.

Odds: Since the crash probabilities π_I are small numbers and can be difficult to comprehend, the odds are summarized in Table 4.

In the case of time-of-day, odds = $\pi_I(A)/\pi_I(B) = \pi_I(t=-1)/\pi_I(t=1)$ where t = 1 (dusk) and t = 1 (dawn, day or night) where *v/c* and s_P are assigned the same values for t = 1 and t = -1. The odds of a crash resulting in injury is twice as great during dusk than during dawn, day or night.

In the case of land use and roadside activity, the odds of a crash resulting in injury is 3.3 times greater at site s_P = 35 mph than at site s_P = 45 mph. Clearly, a site designated as s_P = 35 mph will be expected to pose the greatest highway design challenges.

Since average operating speed is a most important *LOS* measure and it is a function of *v/c*, $\bar{u}$ was used in the odds table with capacity c = 2,800 vph. The odds of a crash resulting in an injury is 2.7 times greater at $\bar{u}$ = 60 mph (*LOS* A) than at $\bar{u}$ = 45 mph (*LOS* D).

CASE STUDIES

The design specification u_f affects c, π_I, $\bar{u}$, *v/c* and the *LOS* rating. Assigning it is critical in design optimization. In this section, graphs of π_I are plotted as functions of u_f where the travel demand v is held constant.

Figure 5 contains π_I plots for sites designated s_P = 45 mph for traffic volumes v = 400 and 2,000 vph at dusk and at dawn, day and dusk. For simplicity, the subscript h is not shown. Figure 6 contains π_I plots for sites designated s_P = 35 mph for the same traffic volumes and times-of-day as in Figure 5. Inspection of Figure 5 shows that the safety compliance constraint, $\pi_I \leq \varpi_I$, is satisfied for a wide range of u_f values at almost any time of the day and at both traffic volumes. In contrast, Figure 6 shows that the safety compliance constraint is satisfied for a narrow range of conditions.

Table 5 contains case study results for sites designated s_P = 35, 40 and 45 mph. In each case, the same annual trip count of $5.7x10^6$ or *ADT* = 16,000 trips per day is assumed. The hourly traffic volume v_h = 400 vph is assumed for all hours of the day except for a two-hour dawn period and for a two-hour dusk peak period. During these two-hour periods, v_h = 2,000 vph. The π_I and $\bar{u}$ values are calculated as marginal probabilities given by equations 12 and 13. The candidates for u_f^* are given under column heading u_f.

Sites s_P *= 40 and 45 mph:* Inspection of case study results show that optimal solutions are obtained for sites designated $s_P = 40$ and 45 mph. That is, *maximum* $\bar{u} = 52$ mph *subject to:* $\pi_I \leq \varpi_I$ is achieved.

Sites s_P *= 35 mph:* The four candidate solutions for $s_P = 35$ mph proved non-optimal. The reasons for non-optimality are:

- safety non-compliance because $\pi_I > \varpi_I$,
- traffic congestion and delay because $v_h > c$, and
- a combination of these reasons.

Reducing the free-flow speed u_f is marginally effective in reducing π_I. Inspection of any one of the π_I plots given in Figure 6 shows that their slopes are slight. As a result, design specifications $u_f < 60$ mph reduce π_I values to a relatively small degree.

Consider design specification $u_f = 40$ mph where $\pi_I > \varpi_I$ and $v_h > c$ are cited as reasons for non-optimality.

The safety constraint is in non-compliance when π_I is calculated as a marginal probability, even though the site meets the safety compliance constraint during dawn and dusk. Figure 6 shows $\pi_{I/h} = \varpi_I$ for $v_h = 2{,}000$ vph at dusk and $\pi_{I/h} < \varpi_I$ for $v_h = 2{,}000$ vph at dawn. A significant portion of the *ADT*, however, occurs during the day and night when $v_h = 400$ vph and $\pi_{I/h} > \varpi_I$.

In addition, the capacity at $u_f = 40$ mph is $c = 1{,}870$ vph; therefore, $v_h > c$ during the dawn and dusk. This design specification is also unacceptable for a reason of traffic congestion and delay.

Relaxing the Allowable Limit: An option that remains to be explored is to increase the allowable limit ϖ_I. This can be accomplished by reducing an individual's exposure η to highway risk. Suppose at site $s_P = 35$ mph, an alternative is found to reduce individual exposure from η = 664 to 400 trips per person per year. The allowable limit is increased from $\varpi_I = 1.2\text{x}10^{-6}$ to $2\text{x}10^{-6}$ as shown in Figure 2. Now, the design specification $u_f = 60$ mph satisfies the safety compliance constraint; thus $u_f^* = 60$ mph!

Given the heavy reliance on the private motor-vehicle in our daily lives, many motor-vehicle trips are made out of necessity and not out of choice, an outcome of urban sprawl. The most mundame tasks, such as buying a newspaper or a loaf of bread, require a trip to the store by automobile. Convenience stores are outlawed by local zoning ordinances in many suburban communities. Through coordinated transportation and land use planning efforts, both individual exposure to the private motor-vehicle η and traffic volume v can be reduced. In addition to promoting highway safety, attractive alternatives, like public transportation, pedestrian and bicycle friendly communities, have far reaching social, public health and environmental benefits.

DISCUSSION

The purpose of this paper is to draw attention to a new outlook that the concept of individual lifetime risk can bring to the highway design process. The design algorithm that was developed is structured as a constrained optimization problem, with an objective to maximize average operating speed subject to a safety compliance constraint. Case studies were analyzed using the design algorithm and its models, average operating speed, allowable safety limit and crash prediction models. Case study results, exploratory data analysis, and individual models used individually and collectively give insightful meaning to the highway design process. For example, the crash prediction model gives insights as to *why* a design may not satisfy the safety compliance constraint at a site; and when this insight is introduced into the larger framework of constrained optimization, this additional information gives further insights as to *how* an optimal design can be achieved. The various analyses and case study results suggest that:

- highway risk or π_I is highly dependent on the land use and roadside activity adjacent to the site s,
- reducing highway risk by geometric alignment is marginally effective,
- reducing highway risk by speed limit control is not considered to be a viable solution in the context of design optimization,
- reducing highway risk by reducing individual exposure can be effective if attractive alternatives to divert motor-vehicle users can be found,
- the concept of individual lifetime risk, odds and crash probability are useful for ranking risks and for effective risk communication.

The crash prediction model, average operating speed and allowable safety limit models, which are fundamental to the design process using a safety compliance standard, have shortcomings. But, it must be realized that no model is "perfect" and no solution is without criticism. No model is capable of incorporating all the multifaceted demands of the driver, the neighborhood, and the various public and private organizations concerned with transportation service, environment, public health, and financing. The models, even when used for design optimization, are unable to address most of these demands.

The crash prediction model is the most severely flawed. Obviously, it can not be used to address issues associated with fatal crashes. This limitation, the use of a surrogate variable for land use and roadside activity, and the questionable result suggesting that "reducing highway risk by geometric alignment is marginally effective", are all directly linked to shortcomings in the data set. There are practical difficulties in constructing the data set for model calibration.

First, the choice of the eight Connecticut sites was motivated by the need to obtain high-quality annual traffic counts and speed data. Without it for instance, the exploratory analysis of *LOS* rating and time-of-day and their significance as explanatory variables would not have been discovered. It was unknown when the data set was compiled that the selected sites would only have one fatal crash in six year period at the eight sites with an annual traffic count of 50 million. As a result, model calibration had to be limited to crashes resulting in injury only.

Secondly, the posted speed limit, a surrogate variable, does not describe the characteristics of the site. Expanding the data set to more sites with additional land use characteristics would be beneficial. Recall that population and road class were the only characteristics available in the data set. Of course, adding sites and land use characteristics must be done with care assuring that the count data can be used to obtain reliable annual traffic counts, an essential measure of highway risk exposure.

Thirdly, the suggestion that "reducing highway risk by geometric alignment is marginally effective" was obtained by extrapolation. Specifically, the crash predictions and average operating speeds calculated for design specifications u_f = 30 and 40 mph for site designation s_P = 35 mph given in Table 5 are suspect. The sites used in this study have highway design speeds of 50 mph and greater. The highways have good sight distances and adequate lane and shoulder widths as given in Table 2. The model calibration did not include data for highway designs of 30 and 40 mph; therefore, the predictions for these highway design speeds are not supported by observation. A data set consisting of sites with highway design speeds of 40 mph and less and sites where traffic calming measures have been used is desirable. Models calibrated with this data set will clarify whether or not geometric alignment is an effective method in reducing highway risk and speed. The predictions given by the crash prediction model for design specifications u_f = 30 and 40 mph seem inconsistent with the results given by CART.[16]

Regardless of imperfections, the overall benefits of the crash prediction model outweigh its shortcomings. Especially, when it is introduced into constrained optimization model, its benefits and the potential usefulness of the design algorithm for highway design are demonstrated. The crash prediction model, in its current stage of development, is considered to be a concept model.

[16] See *supra* note 9.

Table 1

Annual Trip Volume and Crash Counts

Site s	Posted Speed Limit S_P (mph)	Annual Trip Volume (x10^6)	Crash Counts Total	Injury	Fatal	π_P (x10^{-6})	π_I (x10^{-6})
Darien	35	17.7	56	25	0	1.8	1.4
Killingly	35	8.1	34	17	0	2.1	2.1
Hebron	35	1.3	23	10	0	10.	7.7
	Totals	27.1	113	52	0	2.3	1.9
Waterford	40	4.7	72	21	0	11.	4.5
Kent	40	3.4	19	9	0	2.9	5.6
Colebrook	40	4.6	7	4	1	2.9	1.5
	Totals	12.7	98	34	1	5.0	2.7
East Windsor	45	11.0	6	0	0	0.5	0.0
Clinton	45	7.1	6	4	0	0.3	0.6
	Totals	18.1	12	4	0	0.4	0.2

Table 2

Site Characteristics

Site s	S_P (mph)	Land Use Population (x1000)	Road Class	Heavy Vehicles (%)	Traffic Control Number of Signals	Number of Stop Signs	Geometric Design Lane Width (feet)	Shoulder Width (feet)	Sight Distance (feet)
Darien	35	50-200	UPA	3	0	0	11	2	1,500
Killingly	35	< 5	RMA	2	0	0	12	4	1,350
Hebron	35	< 5	RPA	3	1	0	12	8	1,350
Waterford	40	50-200	UPA	1	1	0	12	3	1,500
Kent	40	< 5	RMA	1	0	0	12	1	1,350
Colebrook	40	< 5	RMA	1	0	0	11	2	1,200
East Windsor	45	> 200	UPA	2	0	0	12	5	1,500
Clinton	45	< 5	RPA	1	0	0	12	1	1,500

UPA = Urban Principal Arterial,
RMA = Rural Minor Arterial,
RPA = Rural Principal Arterial

Table 3[17]

Contingency Table of Annual Traffic Volume and Crash Counts

Time-of-day	LOS[18]	Annual Traffic Count (x10^{-6})	Counts Total	Injury	π_P (x10^{-6})	π_I (x10^{-6})
Dawn/Day/Night	A	6.1	85	41	7.2	6.7
	B	17.9	62	17	2.5	1.0
	C	7.7	17	4	1.7	0.5
	D	4.6	12	6	1.3	1.3
	E	6.4	16	4	1.9	0.6
Dusk	A	3.0	12	8	1.3	2.7
	B	2.0	8	3	2.5	1.5
	C	0.6	3	3	0.0	5.0
	D	0.5	1	1	0	2.0
	E	1.2	7	3	3.3	2.5

Table 4

Odds Table

	B	A	$Odds = \dfrac{\pi_I(A)}{\pi_I(B)}$
Time-of-day	$t = -1$	$t = 1$	2.0
Land Use and Roadside Activity	$s_P = 45$ mph	$s_P = 40$ mph	1.8
		$s_P = 35$ mph	3.3
Operating Speed	$\overline{u} = 45$ mph	$\overline{u} = 50$ mph	1.3
		$\overline{u} = 55$ mph	1.8
		$\overline{u} = 60$ mph	2.7

Table 5

Case Study Results (Annual trip count = 5.7x10^{6})

Site s_P	u_f (mph)	π_I	$\overline{u}$	$u_f^* = u_f$	Comments
45	60	0.5 x10^{-6}	52	Yes	$u_f^* = 60$ mph
40	60	1.0 x10^{-6}	52	Yes	$u_f^* = 60$ mph
35	60	1.8 x10^{-6}	52	No	$\pi_I > \varpi_I$
	50	1.6 x10^{-6}	41	No	$\pi_I > \varpi_I$
	40	1.4 x10^{-6}	—	No	$\pi_I > \varpi_I$, $v_h > c$
	30	1.1 x10^{-6}	—	No	$v_h > c$

[17] David W. Hosmer and Stanley Lemeshow, **Applied Logistic Regression**, Wiley Interscience. (1989) pages 25-37.

[18] The *v/c* ratios for each site were calculated using the highway geometric design characteristics given in Table 2. The data for each site were then sorted by *LOS* rating and then combined to form this table.