PB2000-106679

RURAL PEDESTRIAN CRASH RATE: ALTERNATIVE MEASURES OF EXPOSURE

John N. Ivan
Paul J. Ossenbruggen
Xiao Qin
Jyothi Pendarkar

UNITED STATES DEPARTMENT OF TRANSPORTATION

REGION I UNIVERSITY TRANSPORTATION CENTER

PROJECT UCNR 11-10

FINAL REPORT

June 25, 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.
UCNR 11-10	N/A	N/A

4. Title and Subtitle	5. Report Date
Rural Pedestrian Crash Rate: Alternative Measures of Exposure	June, 2000
	6. Performing Organization Code N/A
7. Author(s) John Ivan; Paul J. Ossenbruggen; Xiao Qin; Jyothi Pendarkar	8. Performing Organization Report No. UCNR 11-10
9. Performing Organization Name and Address University of Connecticut Department of Civil and Environmental Engineering 191 Auditorium Road, Box U-37 Storrs, CT 06269	10. Work Unit No. (TRAIS) N/A
	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 Report
	14. Sponsoring Agency Code N/A

15. Supplementary Notes

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

16. Abstract

The objective of this research is to better understand how to prevent pedestrian fatalities by investigating new means of measuring pedestrian exposure to crashes. The study involves finding ways to estimate pedestrian volumes (pedestrians per week) on specific facilities, focusing on rural areas.

Part I investigates the relationship between the weekly pedestrian exposure in rural areas of Connecticut and factors such as population density, sidewalk system, number of lanes, area type, signal type and median household income. General Linear Regression (GLM) and Tukey or Duncan multiple comparison of means methods are used to identify the significant factors. Only the number of lanes, area type and sidewalk system are significant in the resulting model for pedestrian exposure. Other factors do not significantly explain the variation in the pedestrian exposure.

Part II identifies site characteristics, i.e., factors describing land use activity, roadside design, merging and crossing traffic, traffic control and vehicular speed, that can be used to predict roadway risk, the probability of a crash leading to injury and/or death. All variables, with the exception of crosswalks, street parking and paved shoulder width, proved significant. Typical village and residential sites proved to be the least hazardous and shopping sites the most hazardous.

The findings from this project could significantly change the way pedestrian crashes are reported and analyzed, and thus improve their usefulness and meaning. This new reporting format could help jurisdictions decide how to allocate funds for enhancing pedestrian safety by giving them more detailed information about where the enhancements are best applied. This research also has applications toward general travel demand forecasting, by providing insight into how to better predict the choice of walking as a travel mode.

17. Key Words	18. Distribution Statement
Pedestrian Exposure; Highway Safety; Rural; General Linear Regression (GLM); Logistic Regression; Duncan Multiple Comparison of Means; Prediction Model; Land Use, Pedestrians, Traffic Control, Roadway Design.	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	51	N/A

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 Institute Program, in the interest of information exchange. The U.S. Government assumes no liability for the concerns.

TABLE OF CONTENTS

PART I: Pedestrian Exposure Prediction Model in Rural Areas

ABSTRACT

Pedestrian exposure is defined as the risk of exposure to traffic or the probability of the chance of a crash occurrence. It is an important factors influencing pedestrian crashes. Since pedestrian exposure is not readily available, pedestrian volume counts or population density is used as substitute in the pedestrian crash prediction model in many cases. However, pedestrian volume counting is costly and time-consuming when the number of study sites is large. Also, the population density is not a good replacement of pedestrian exposure.

This study investigates the relationship between the weekly pedestrian exposure in rural areas of Connecticut and factors such as population density, sidewalk system, number of lanes, area type, signal type and median household income. General Linear Regression (GLM) and Tukey or Duncan multiple comparison of means methods are used to identify the significant factors. Only the number of lanes, area type and sidewalk system are significant in the resulting model for pedestrian exposure. Other factors do not significantly explain the variation in the pedestrian exposure. Ongoing research will take advantage of the model to estimate the pedestrian crash model in the rural areas of Connecticut.

INTRODUCTION

Pedestrians are extremely vulnerable in crashes with motor vehicles, and for this reason, although pedestrian crashes only make up 2 percent of highway injuries, they constitute 13 percent of the highway fatalities in 1997 in the U.S (1). For example, in Connecticut, pedestrian fatalities represented 13.7 percent of all traffic fatalities in 1998. Despite the tremendous progress made in transportation safety, especially the research on pedestrian travel patterns and the improvement of pedestrian facilities in the last three decades, more than 5,220 pedestrians were killed and 69,000 pedestrians were injured (1). On average, a pedestrian was killed in a traffic crash every 97 minutes and a pedestrian was injured in a traffic crash every 6 minutes in 1996 (2).

Nevertheless, pedestrian injuries have actually decreased since 1986. In 1996, 5,412 pedestrians were killed in traffic crashes in the United States — a decrease of 20 percent from the 6,779 pedestrians killed in 1986 (2). This decrease appears more pronounced when looking at the number of crashes per person in the population, since population increased from 237,626,036 in 1986 to 265,283,783 in 1996 (3). However, does this mean it was safer to walk in 1996 than in 1986? Actually, the decline may be due to the attendant reduction in walking as a travel mode rather than an improvement in pedestrian safety, because people tend to drive rather than walk with increasing motorization. It is more accurate to assess pedestrian safety by how many people actually walk on the streets if we want to learn what factors improve or worsen pedestrian safety.

Because the actual time spent walking is difficult to observe or the pedestrian volume is not readily available, population density is usually used as a substitute in pedestrian crash prediction models. However, population density does not necessarily relate directly to the actual number of people walking on the streets. For example, some tourist sites usually attract a large number of people who are not counted in the population density. In other words, there is unexpected high pedestrian volume that can not be represented by the population density in these areas. In some areas, the low pedestrian volume compared to the high population density may be attributed to the high vehicle-owner rate. Therefore, prediction models based on population density are probably unreliable.

The purpose of this study is to learn how to estimate pedestrian exposure in rural areas for more accurate reporting of pedestrian crash statistics. Many studies have investigated pedestrian safety in urban areas because pedestrian crashes occur there at a higher frequency. There is little research studying pedestrian activities or safety in rural areas despite the fact that 32 percent of fatalities apparently occur in rural or suburban areas rather than urban areas (1). Our study summarizes the findings of research regarding the effects of road features, neighborhood and land use, site characteristics, and demographic characteristics on pedestrian activities in rural areas of Connecticut. It also sets forth a model for estimating pedestrian exposure and travel patterns in rural areas.

LITERATURE REVIEW

Exposure is a concept in risk analysis describing the opportunity for a random event to occur, in other words, the number of trials. Consequently, identifying the appropriate measure of exposure for a particular risk event is extremely important for analyzing the likelihood of its occurrence (4). For pedestrian safety analysis, this exposure measure should account for the extent to which people place themselves at risk of being hit by a motor vehicle. If these criteria are met, the exposure metric can be a reliable explanatory variable for predicting pedestrian crashes.

Different measures of pedestrian exposure cause different results in risk analysis. Keall examined pedestrian crash data by using exposure measures "time spent walking" and "number of roads crossed" (5). These two measures of risk are more precise than the most common mode of presenting pedestrian crash statistics, number of crashes per person in the population. Crashes per person overestimated the risk of people under 30 years, underestimated the risks of people over 79 years, and underestimated the risks of males compared with females (5).

Many studies investigate how site characteristics are associated with both pedestrian exposure (represented by probability of choosing walking as a travel mode) and pedestrian crashes. Hess et al. studied the relationship between site design and pedestrian travel in a mixed-use, medium-density environment (6). He investigated site design characteristics, such as the mean block size, completeness and continuity of the sidewalk system, and on-street parking. The findings showed that these factors significantly influenced the likelihood of choosing walking as a travel mode. Shriver's results supported the conclusion that neighborhood transportation, land use and design characteristics affected pedestrian activities (7). Knoblauch et al. found that sites without sidewalks were more than twice as likely to have pedestrian crashes than sites with sidewalks (8).

A large number of studies have recognized the indirect relationship between pedestrian crashes and economic or demographic factors. For example, Bagley investigated the probability of sites to be hazardous given socioeconomic and crime data (9). Roberts et al. noted a relationship between economic and ethnic differences with the pedestrian crash rate (10). Epperson recognized that the economic status of a neighborhood significantly influenced the number of pedestrian crashes (11). McMahon et al. studied demographic variables such as the percentage of single parents with children, the percentage of housing stock built after 1980, whether 85 percent of the households were composed of families and whether the unemployment rate was less than 1.75 (12). The study showed that percentage of single parents with children and housing stock built after 1980 significantly influenced the "walking along the road" crashes. Finally, McMahon et al. concluded that factors contributing to "walking along the roadway" crashes included not only geometric characteristics of the sites but also demographics and neighborhood characteristics (12). The indirect economic or demographic factors may influence the pedestrian exposure that is directly related to pedestrian crashes.

McMahon et al. identified the risk to pedestrians who were walking along the roadway (12). However, different types of pedestrian activities bear quite different risks of experiencing conflict with motor vehicles: for example crossing a road as opposed to walking along a road, or walking on a sidewalk. Actually, 45 percent of all pedestrian crashes involve a pedestrian crossing a road, while only 14.1 percent involve a pedestrian walking along a road (13). Statistics suggest that crossing the street might be more dangerous than walking along the road, so that crossing exposure should be distinguished from roadside or sidewalk exposure.

Case-control methods to predict pedestrian crashes or exposure have been widely applied (7, 14, 15). For example, Hess selected sites by controlling the variables regarded by former research as affecting the pedestrian volume, such as gross population density, land use, income. The hypothesis is that when holding control variables constant, the other factors like the mean block size, completeness and continuity of the public sidewalk system and so on affected pedestrian volume (6).

In our study, the factors that may be important contributors to predict the pedestrian exposure, such as site characteristic, traffic control types, demographic data, land use characteristic and road site features are investigated using the General Linear Regression model. In addition, the weekly crossing pedestrian volume is applied as the measure of pedestrian exposure in the prediction model.

METHODOLOGY AND DATA PROCESSING

Study design

Variable selection

In order to estimate the pedestrian exposure in rural areas, 32 sites are selected for specific site characteristics from rural areas in Connecticut. The factors that may influence pedestrian activities are from the following categories.

1. Pedestrian Amenities
   Sidewalk is used as an independent variable to present the site characteristic because sidewalk is an important pedestrian-friendly design. In general, sites with sidewalks can attract more pedestrians than sites without sidewalks.
2. Traffic control
   Pedestrian-friendly traffic control designs are assumed to encourage people to walk, such as marked crosswalk, traffic signal and so on.
3. Demographic data

Median household income is an important demographic characteristic expected to be associated with the pedestrian exposure. Previous studies show that neighborhoods with high income usually have less pedestrian exposure than the neighborhoods with low income.

4. Land use characteristics
   Area type can greatly influence the pedestrian travel patterns. For example, pedestrian travel patterns in commercial areas are not the same as those in residential areas. Furthermore, pedestrian exposure in tourist and college campus areas is quite different from others.
5. Road site features
   Some road geometric features can influence the degree of walking such as the number of lanes, road width and so on. For example, wider highways negatively affect pedestrians by significantly increasing the distance that must be traversed to get to the other side.

Variable description

The traffic signal categorical variable has five levels as follows:

1. no marked crosswalk and no traffic signal;
2. signal without marked crosswalk;
3. marked crosswalk without traffic signal;
4. marked crosswalk with yellow caution signal;
5. marked crosswalk with traffic signal ;

According to the sites available, we divide the area type into several levels (16).

1. *Downtown* areas are characterized by larger buildings abutting one another and abutting sidewalks.
2. *Compact residential* areas predominantly have houses close together and generally visible from the road, and often have sidewalks.
3. *Low-density residential* areas have houses that are spaced apart and often are not visible from the road. Sidewalks are rare in these areas. Areas with little to no development are included in this category.
4. *Village* areas consist of smaller buildings and residences set back from the road. Sidewalks may or may not be present.
5. *Medium and low-density commercial* areas have commercial development, often with sidewalks. This area type includes commercial development such as gas stations, fast food, and supermarkets. On-street parking is not likely to be found in this type of area.
6. *Tourist* areas usually include crosswalk and sidewalk without signal. Higher pedestrian exposure is expected and pedestrian's activities may be constant throughout the day with less pronounced peaks during commuting and lunch time than at other areas.
7. *Campus* areas usually include crosswalks, sidewalks without signal and narrow streets and speed limit. Much higher pedestrian exposure is expected and pedestrian's activities are greatly changed throughout the day with pronounced peaks during class and lunch or dinnertime.

Data collection and processing

Collection of Field Data

Pedestrian activity was observed in May, June, October and November of 1999, with the exception of one count in Storrs, which took place in November 1998. All data counts were carried out in non-inclement weather, therefore optimal conditions for walking were usually present. For each site, a weekday count as well as a weekend count was conducted. The weekday count was conducted on a typical weekday, the weekend count was mostly taken on Saturday. Sunday counts were conducted where, based on the characteristics of the site, they were expected to yield the more significant results, such as a town center with little retail activity. Table 2 shows the study sites and the count dates.

The data assessed included not only the count of the number of pedestrian exposures in general, but also pedestrian behavior, such as crossing highway behavior or walking on the highway, for each quarter hour of the observation period. The different activities observed were:

1. Walking along highway
2. Crossing highway: with dedicated pedestrian phase – walk
3. Crossing highway: with dedicated pedestrian phase – don't walk
4. Crossing highway: with signal but no dedicated pedestrian phase – green
5. Crossing highway: with signal but no dedicated pedestrian phase – red
6. Crossing highway: with crosswalk without signal
7. Crossing highway: without or outside crosswalk

Since each site featured different characteristics such as pedestrian crossing facilities or crosswalks, not all the categories were applicable to each site. However, with respect to this study, investigating alternative measures of pedestrian exposure to accidents, only the total number of crossing pedestrian exposures was of interest. Walking on the sidewalk was not included because we assume that these pedestrians are unlikely to be hit by an automobile. Observations generally took place in the time period from 8AM to 5:30PM.

Processing of Field Data

In order to obtain a single measure of pedestrian exposure for each observation site, the field data were processed by adding up all possible pedestrian exposures except for walking along the highway. As observation time did not always cover the total time period from 8AM to 5:30PM, the method of sample moment is applied to estimate the missing value. Usually the average number of pedestrian exposures for 15-minute periods during the observation time was used to predict pedestrian volume for the time during which no observation took place. The procedure is as follows (appendix A offers a detailed example):

Step 1: Group all the observation sites into small groups by site features such as land use and population density and so on.

Step 2: Obtain an average pedestrian volume distribution of every 15-minute observation period for each small group.

Step 3: Compute the percentage of each 15-minute observation period's pedestrian volume compared to the whole day observation.

Step 4: For each site, use the available 15-minute observation to divide corresponding percentage of that period and obtain the estimated whole day observation. If the number of the whole day observation is more than one, take the average.

Table 2 shows the pedestrian volume in weekday, weekend and the total weekly pedestrian volume with the following relationship.

$$V=5V_{wd}+2V_{we} \quad \textbf{(1)}$$

Where:

V is the total weekly pedestrian volume,
V_{wd} is the weekday pedestrian volume,
V_{we} is the weekend pedestrian volume.

Collection and Processing of Demographic Data

The next step was to assess the demographic characteristics of the vicinity of each site. The demographic data used in this study is median household income obtained from the 1990 census and customized for the geographic area of each site using digital street maps. The median household income is based on the households within a walking distance around the study site. And, the walking distance to the study site is defined as 3600 feet, which equals 20 minutes when walking at a speed of 3 feet per second. Thus the neighborhood within walking distance was defined as a polygon encompassing all areas that are within 3600 feet when walking on streets (Please refer to Appendix B for details).

Analysis methodology

Many previous studies have found a non-linear relationship between the pedestrian exposure with the corresponding variables. In fact, the choice of a model form is very arbitrary (18); the model form chosen here is:

$$V = P^{\alpha} e^{(\beta_0 + x_S\beta_S + x_D\beta_D + x_L\beta_L + x_R\beta_R + \varepsilon)} \quad \textbf{(2)}$$

where:

V is the total weekly pedestrian volume,
P is the computed population density in the walk area,
α is the exponent on population density to be estimated,
X_S, X_D X_L and X_R represent site characteristics, demographic characteristics, land use characteristics and road characteristics, respectively,
$\beta_0, \beta_S, \beta_D, \beta_L$ *and* β_R are parameters to be estimated, and ε is an error term.

After the natural Log transformation, the Formula (1) is turned into a simpler linear form as follows.

$$LnV = \alpha LnP + \beta_0 + x_S\beta_S + x_D\beta_D + x_L\beta_L + x_R\beta_R + \varepsilon \quad \textbf{(3)}$$

Statistical Analysis System (SAS) PROC GLM (generalize linear regression) is used to perform the linear regression and estimate the parameters for the covariates. Here, the counted weekly pedestrian volume crossing the street is regarded as the response variable. Most of the explanatory variables are categorical variables except for P (population density). These explanatory variables are assumed to be independent from each other and only the main effects are considered. In other words, the interaction effects between different predictor variables or explanatory variables are neglected. If the null hypothesis that our assumptions about the variables are significant was rejected, the corresponding factor (variable) can be discarded without influencing the predicted values' precision. Otherwise, the variable contributes toward predicting the weekly pedestrian exposure and should be kept in the model.

ANALYSIS AND RESULTS

In order to identify the patterns between pairs of continuous variables in a model, we examine the relationship between the variables by the matrix plot. There are three continuous variables in our primitive model: natural log of weekly pedestrian exposure (*LnV*), natural log of population density in the walking area (*LnP*) and the median household income (*M*). The relationships between *LnV* versus *LnP* and *LnV* versus income are identified. If the plots are randomly distributed and lack a linear pattern, the linear relationship between two variables is not obvious, such as the relationship between *LnV* and *M*. Otherwise the linear relationship between two variables is obvious, such as the relationship between *LnV* and *LnP*. Despite the fact that the selected sites cover a wide income range from $16,822/per year to $60,953/per year, after the test of between-subject effects, the income factor shows a non-significant role in the model prediction and should be discarded. Because the income is a significant demographic variable in some previous pedestrian exposure models, the inconsistent conclusion in our study may be due to the limited number of sites. However, we still need to be cautious when using the median household income (*M*) to predict the pedestrian exposure. Figure 2 shows the nonlinear relationship between variable *LnV* and *M* and the approximate linear relationship between *LnV* and *LnP*.

After discarding the variable income, the model estimation was undertaken with the rest of the variables, such as population density, area type, signal type, number of lanes and sidewalk. Only the population density is entered as a continuous variable. Others are all categorical variables. Following is the full model.

$$Ln\hat{V} = \alpha LnP + \beta_0 + \beta_a x_a + \beta_s x_s + \beta_l x_l + \beta_w x_w \qquad \textbf{(4)}$$

where the symbols V, α, P are as defined before, and subscripts a, s, l, w denote categorical variables area type, signal type, number of lanes and sidewalk, respectively. Table 3 lists all variables with brief definitions.

Regression coefficients were tested for each variable using two types of sum of square to test the significance of variables: Type I sum of square (SSI) and Type III sum of squares (SSIII). SSI, also called sequential sum of squares, is the incremental improvement in error SS as each variable is added to the model (17). Usually, the factor that we assume to be the most important in the model is added first. Thus, the SSI hypothesis depends on the order in which effects are specified in the model and the significance of the variables in the model is greatly influenced by this sequence. Unlike SSI, SSIII, also referred to as partial sum of squares, is considered by many to be the most desirable because the hypothesis for an effect does not involve parameters of other variables (17). Table 4 is an ANOVA table for the full model (Model 1) which does not consider the income factor.

As it is mentioned above, the different significance of *LnP* between SSI and SSIII can be attributed to the adding sequence. Based on the previous studies that population density has a significant effect on the pedestrian exposure, population density is temporarily kept in the model and to be tested further. Despite the discrepancy in the variable *LnP*, the two sum of squares are consistent for the other four variables. In other words, factors such as with or without sidewalk (*Xw*), number of lanes (X_l), area type (*Xa*) and signal type (*Xs*) are all significant in predicting the response variable *LnV* at the 95 percent significance level.

However, whether or not specific levels within each categorical variable are significantly different from each other or not is still unclear until a multiple means comparison method, such as Tukey or Duncan grouping, is undertaken. The significance of all levels in each categorical variable is open to the test. Table 5 gives the results of main effect of levels in each categorical variable.

In the Tukey and Duncan method results, levels indicated by the same letter code are not significantly different from each other. Table 5 shows that these differences for the sidewalk, number of lanes and signal type are consistent for both Tukey and Duncan methods. However, using Tukey methods, there are some overlaps between insignificantly different levels for area type, which means there is not a single way to group the seven area types, though there are none using the Duncan method. Because Duncan method is more sensitive in identifying the difference between the levels, we only use Duncan method to regroup the categorical variables. As can be seen in Table 5, the coefficients on tourist area and downtown are not significantly different from each other in Duncan method. Consequently, they are combined into a single category, as are the other levels. Therefore, we estimated Model 2, in which area type is reduced from seven levels to three levels and signal type from five to two. Table 6 shows the new categorical variables in Model 2.

The ANOVA table for Model 2 is given in Table 7. The output from SAS GLM procedure shows that the significance of all the variables in Model 2 are unchanged except that the signal factor changes from 95 percent significance level to 85 percent significance level. The result shows that there is confounding in the signal effect on pedestrian exposure. More likely, signal type effect is confounded with the others, i.e., it is not independent, but somewhat correlated with other factors. Thus, we discarded it from Model 2 and obtain Model 3. Table 8 shows the ANOVA table for Model 3.

Model 3 is a reduced model derived from Model 2 because signal type variable is not considered in Model 3. As can be seen in Table 9, in Model 2 and Model 3, the negative coefficient for the population density (*LnP*) means more population density cause lower pedestrian exposure when other factors are controlled, which does not make any sense. Therefore, population density in the walking area (*LnP*) should be discarded from the predicted model. Now, there are three categorical variables in the new restricted model (Model 4): with or without sidewalk, number of lanes, and area type. The SAS output is in Table 10.

Based on F-extra test between full Model 1 and Model 2, the null hypothesis that the surplus variables in full model are nonsignificant at 95 percent level of significance can not be rejected. In other words, the evidence does not prove that the additional information in Model 1 helps to better predict the pedestrian exposure. Compared with Model 1, the alternative Model 2 is much simpler and the predicted accuracy is good because of the close R square and Root MSE value of the two models. The result of F-partial test or t-test between Model 2 and 3 suggests that the variable signal type is not an important contributor to predict the pedestrian exposure. Thus, Model 3 is better than Model 2 because it has few parameters. The model selection procedure is in Table 11.

Figure 2 to Figure 5 are plots of 90 percent confidence interval of the predicted pedestrian exposure versus actual pedestrian exposure for Model 1, Model 2, Model 3 and Model 4, respectively. Figure 3 shows that Model 1 (the most complex one) fits the best and the shapes of the other models are very similar. However, the small increases in residual deviance of Model 4 (the simplest) is offset by the reduction in model complexity. Furthermore, the more parameters in a model, the unstable it is. So, Model 4 has the best form.

Finally, the best model is as follows.

$$Ln\hat{V} = 3.78 + 1.33X_W + 1.87X_L + 2.62X_C + 0.88X_{TD} \qquad \textbf{(5)}$$

where:

X_W is sidewalk (X_W=1, with sidewalk; X_W=0, without sidewalk),

X_L is number of lanes (X_L=1, two-lane highway; X_L=0, four-lane highway),

X_C is campus factor (X_C=1, campus, X_C=0, others),

X_{TD} is tourist and downtown factor (X_{TD}=1, tourist and downtown areas; X_{TD}=0, others).

After the transformation, the formula is as follow.

$$V = e^{3.78+1.33X_W+1.87X_L+2.62X_C+0.88X_{TD}} \qquad \textbf{(6)}$$

Therefore, based on the three variables in the Model 4, the predicted values are shown in Table 12. Because of the limited number of sites and actual site type, some site types are not observed, such as the sites in tourist or downtown areas without sidewalk. However, the predicted values for these site types without actual data are still offered in Table 12. From the Table 12, the predicted values are very close to the observed values.

CONCLUSIONS

Pedestrian exposure is an important variable to predict the pedestrian crashes because it represents the chances of risk for pedestrians to crash with vehicles. There are many factors that might influence the pedestrian travel patterns and pedestrian volumes. These factors can be described as population density, demographic characteristics, site characteristics, land use characteristics and highway geometric characteristics.

In this study, on the one hand, several factors' effects do not conform to our expectation. For example, the factor population density in walk area is not significant. The result warns us to be cautious when using the population density as a parameter to predict pedestrian crashes or exposure. Despite the fact that, according to our observation, most of the pedestrians prefer to use crosswalk or wait for the signal when crossing the street, the signal type is still non-significant. The non-significance of the signal variable may be because the signal type effect is confounded with the others, i.e., it is not independent, but somewhat correlated with other factors. Thus, it can be regarded as a cause brought by high pedestrian volume rather than a reason to increase the pedestrian volume.

Furthermore, it is interesting to find that the demographic variable, median household income, is not significant either. The result is different from many previous studies. The discrepancy might be due to the limited number of sites. Actually, all of our data are from rural areas of Connecticut. The median household income varies greatly from site to site and no obvious pattern shows that there is a relationship between the variable and the pedestrian exposure.

On the other hand, some factors have expected effects. For example, sidewalk has positive effect. The provision of sidewalk system possibly encourages people to walk rather than drive. Two-lane highway attracts more people to cross than four-lane highway because it is less dangerous and less chances of risk for pedestrians to crash with vehicles. Our study results also show that campus factor has the greatest positive effect on pedestrian exposure with the coefficient 2.62. Tourist and Downtown areas are next to the campus factor with the coefficient 0.88. Therefore, area type is a significant contributor to predict the pedestrian volume.

The model offers us a simple and effective way to predict the pedestrian exposure in rural area of Connecticut. It verifies that the pedestrian activity varies according to sidewalk system, the number of lanes on the highway and the area type. It shows that pedestrian safety analyses based on population density may distort the true risk values. Furthermore, it can also be applied to identify potentially hazardous locations for pedestrians. It is helpful for us to do the further research and analyses on the pedestrian crash rate.

It also should be mentioned that our study is limited to pedestrian crossing volumes in rural areas of Connecticut. If the results are applied in other regions, the regional differences should be considered. In addition, the limited number of sites does not cover all the site features such as the no observed sites in Table 11. Future research includes collecting information of the no observed sites to test the predicted pedestrian exposure; collecting pedestrian information from other states to test the transfer probability of the model or to find out the state factor. Furthermore, the predicted pedestrian exposure to walking trips is prepared for the facilities studied to be used for analyzing pedestrian fatality and injury rate in our future research.

REFERENCES

1. Bureau of Transportation Statistics. *Transportation Statistics Annual Report 1999*. U.S. Department of Transportation, 1999.
2. Bureau of Transportation Statistics. *Traffic Safety Facts 1996 (Pedestrians)*. U.S. Department of Transportation, 1996.
3. http://www.census.gov/prod/3/98pubs/p23-194.pdf.
4. Hauer, Ezra. Traffic conflicts and exposure. Paper presented at the International Symposium on Risk-Exposure Measurement, Aarhus, Denmark, June 1980.
5. Keall, Michael D. Pedestrian Exposure To Risk of Road Accident In New Zealand. *Accident Analysis and Prevention,* Vol. 27, No. 27, 1995, pp. 729-740.
6. Hess, Paul M., A.V. Moudon, et al. Site Design and Pedestrian Travel. In *Transportation Research Record 1674*, TRB. National Research Council, Washington, D.C., 1999, pp. 9-19.
7. Shriver, Katherine. Influence of Environmental Design on Pedestrian Travel Behavior in Four Austin Neighborhoods. In *Transportation Research Record 1578*, TRB. National Research Council, Washington, D.C., 1997, pp.64-75.
8. Knoblauch, R. L., B. H. Justin, S. A. Smith, and M. T. Pretrucho. Investigation of Exposure-Based Pedestrian Accident Area: Crosswalks, Sidewalks, Local Streets, and Major Arterials. Report FHWA/RD-87-038. FHWA, U.S. Department of Transportation, Feb. 1987.

9. Bagley, C. The Urban Setting of Juvenile Pedestrian Injuries: A study of Behavioral Ecology and Social Disadvantage. *Accident Analysis and Prevention*. Vol. 24, No.6, 1992, pp.673-678.
10. Roberts, I., R. Norton, and B. Taua. Child Pedestrian Injury Rates —— The Importance of "Exposure to Risk" Relating to Socioeconomic and Ethnic Differences, in Auckland, New Zealand. Journal of Epidemiology and Community Health, 1989.
11. Epperson, B. Demographic and Economic Characteristics of Bicyclists Involved in Bicycle-Motor Vehicle Accidents. In Transportation Research Record 1502, TRB, National Research Council, Washington., D.C., 1995, pp. 58-64.
12. McMahon, Patrick J., C. Duncan, et al. Analysis of Factors Contributing to "Walking Along Roadway" Crashes. In *Transportation Research Record 1674*, TRB. National Research Council, Washington, D.C., 1999, pp.41-48.
13. North Carolina Department of Transportation. Bicycle and Pedestrian Safety Study. December 1997.
14. Moudon, A V., P.M. Hess, M. C. Snyder, and K. Stanilov. Effects of Site Design on Pedestrian Travel in Mixed-Use,Medium-density Environments. In *Transportation Research Record 1578*, TRB. National Research Council, Washington, D.C., 1997, pp. 48-55.
15. Carlin, J.B., Taylor, and T. Nolan. A Case Control Study of Child Bicycle Injuries: Relationship of Risk to Exposure. Accident Analysis and Prevention, vol.27, No.6 1995, pp.839-844.
16. Zajac, S. Sylvia. Factors Influencing Injury Severity of Motor Vehicle-Crossing Pedestrian Crashes in Rural Connecticut. MS thesis, Civil and Environmental Engineering, University of Connecticut, CT, 2000.
17. SAS/STAT User's Guide Volume 2, GLM-VARCOMP (Version 6, Fourth Edition). SAS Institute, Inc., Cary, N.C., 1990.
18. Valavanis, S., 1959. Econometrics: An Introduction to Maximum Likelihood Methods. McGraw-Hill, New York.
19. Jensen, S. U. Pedestrian Safety in Denmark. *In Transportation Research Record 1674*, TRB. National Research Council, Washington, D.C., 1999, pp. 61-69.
20. Weatherall, R. What we need to know about walking. Traffic Engineering and Control. Vol. 38, No. 7/8,1997, pp. 385-387.
21. Hauer, Ezra. On Exposure and Accident Rate. Traffic Engineering and Control. Vol. 36, No. 3, 1995, pp. 134-138.
22. Garber, Nicholas J. and Lienau, Torsten K., "Traffic and Highway Geometric Characteristics Associated with Pedestrian Crashes in Virginia," Virginia Transportation Research Council, Report No. VTRC 96-R29, March 1996.
23. Draper, Norman R. Smith, Harry. Applied Regression Analysis (Third Edition). Wiley-Interscience Publication. John Wiley & Sons, INC., 1998.

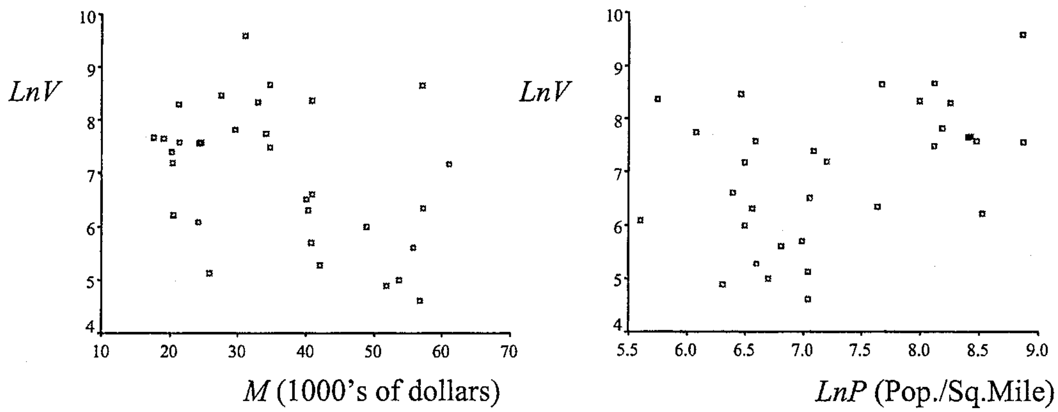

Figure 1. The Relationship Between Continuous Variables

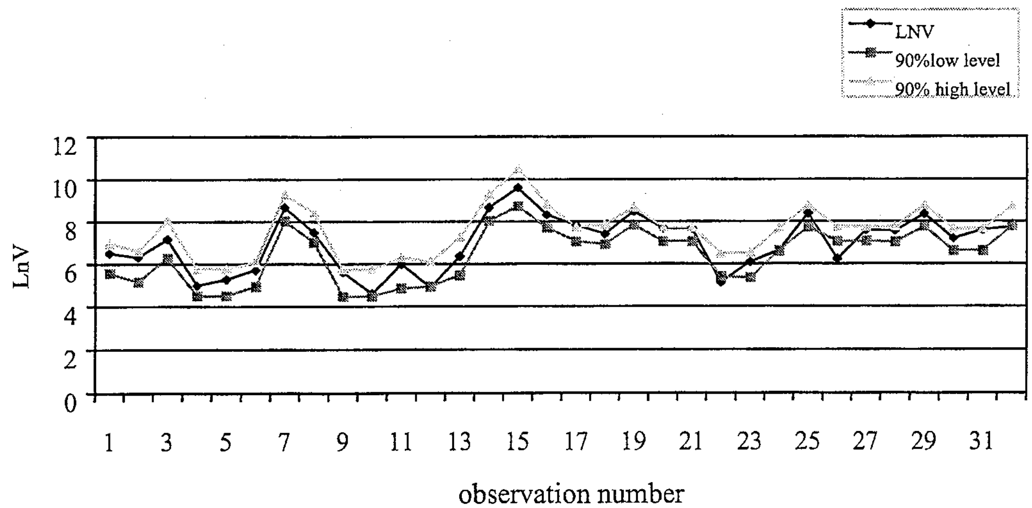

Figure 2. Actual *LnV* versus the range of predicted *LnV* (90%) in Model 1

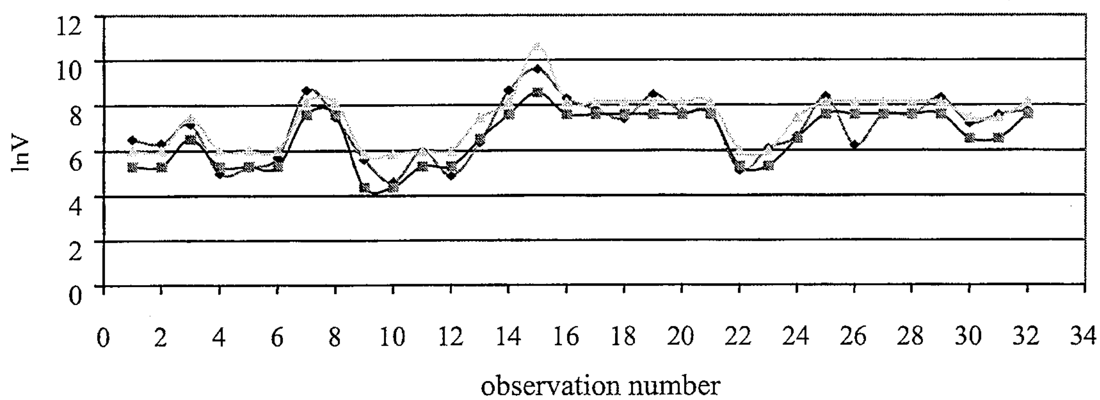

Figure 3. Actual *LnV* versus the range of predicted *LnV* (90%) in Model 2

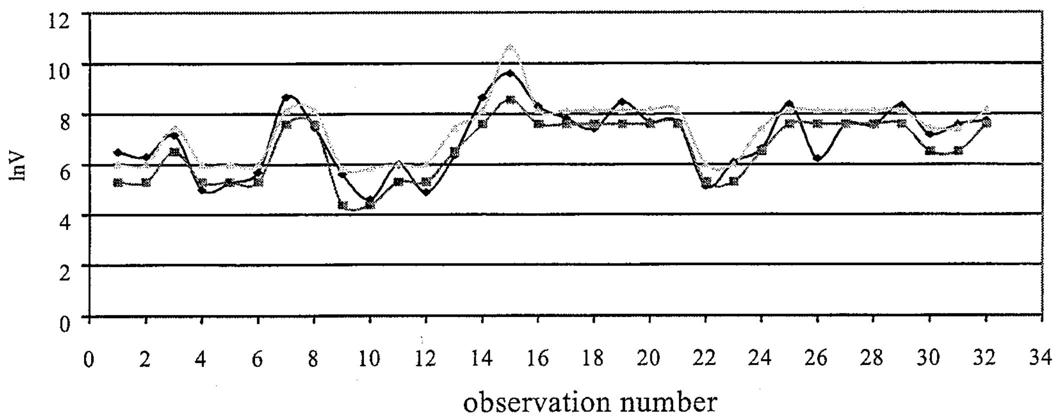

Figure 4. Actual *LnV* versus the range of predicted *LnV* (90%) in Model 3

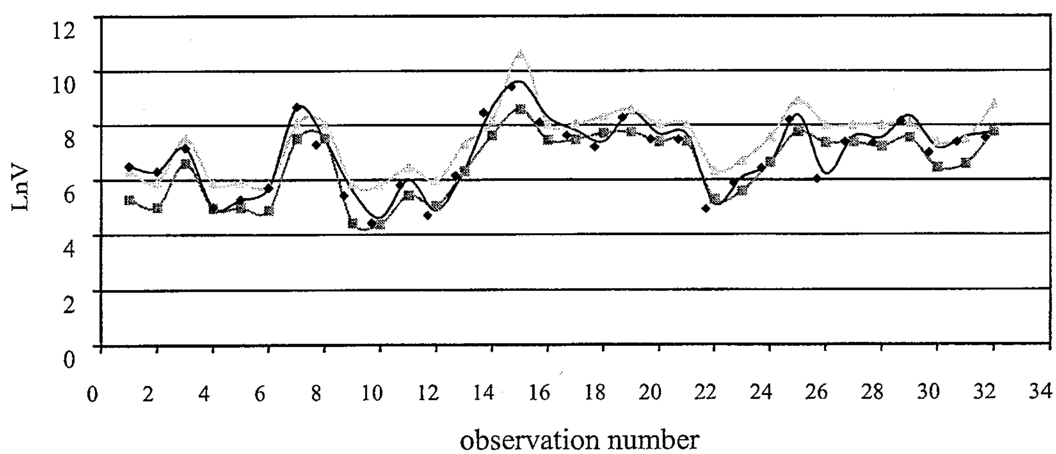

Figure 5. Actual *LnV* versus the range of predicted *LnV* (90%) in Model 4

Table 1. Sites and Traffic Count Dates List

Site number	Town	Weekday Count Date*	Weekend Count Date*
1	Coventry	5/17/99	5/22/99
2	Coventry	5/18/99	5/22/99
3	Tolland	5/27/99	6/6/99
4	Mansfield	5/25/99	6/12/99
5	Brooklyn	5/20/99	5/22/99
6	Brooklyn	5/21/99	5/22/99
7	Stafford	5/28/99	6/5/99
8	Stafford	5/28/99	6/5/99
9	Avon	6/3/99	6/5/99
10	Avon	6/3/99	6/5/99
11	Simsbury	6/2/99	-
12	Simsbury	6/2/99	-
13	Farmington	6/1/99	6/12/99
14	Farmington	6/1/99	6/12/99
15	Storrs	11/5/98	5/15/99
16	Groton	11/1/99	10/23/99
17	Pawcatuck	11/1/99	10/10/99
18	Canaan	10/27/99	-
19	Kent	10/29/99	10/10/99
20	Danielson	11/4/99	11/6/99
21	Jetwtt City	11/1/99	11/6/99
22	Durham	11/3/99	-
23	Rivertown	10/25/99	-
24	Lakeville	10/27/99	10/9/99
25	Salisbury	10/27/99	10/9/99
26	Winsted	10/25/99	-
27	Watertown	10/29/99	-
28	Rockville	11/10/99	-
29	Guiford	11/3/99	-
30	Baltic	11/5/99	-
31	Deep River	-	10/30/99
32	Essex	-	10/30/99

*: "-" represents no observation data at the site.

Table 2. Number of Pedestrian Exposures for Each Site

Site number	Town	Weekday Pedestrian Exposure	Weekend Pedestrian Exposure	Total Weekly Pedestrian Exposure
1	Coventry	98	89	668
2	Coventry	92	45	550
3	Tolland	187	185	1305
4	Mansfield	25	11	147
5	Brooklyn	26	32	194
6	Brooklyn	34	64	298
7	Stafford	878	715	5820
8	Stafford	263	231	1777
9	Avon	32	57	274
10	Avon	20	0	100
11	Simsbury	57	57	399
12	Simsbury	19	19	133
13	Farmington	102	32	574
14	Farmington	1128	31	5702
15	Storrs	2788	392	14724
16	Groton	398	1024	4038
17	Pawcatuck	410	215	2480
18	Canaan	239	223	1641
19	Kent	392	1402	4764
20	Danielson	327	263	2161
21	Jetwtt City	306	291	2112
22	Durham	27	16	167
23	Rivertown	71	42	439
24	Lakeville	111	92	739
25	Salisbury	327	1360	4355
26	Winsted	73	68	501
27	Watertown	285	266	1957
28	Rockville	278	260	1910
29	Guiford	335	1262	4199
30	Baltic	193	180	1325
31	Deep River	264	315	1950
32	Essex	363	243	2301

Table 3. Variable Information of Model 1

Variable name	Variable type	Variable feature	Variable Description	
P	Continuous	Population density		Population Density in walk area
Xw	Categorical	Site characteristics	1	Site with sidewalk along the highway
			0	Site without sidewalk along the highway
X_l	Categorical	Highway cross-section characteristics	1	Site on 2-lane highway
			0	Site on 4-lane highway
Xa	Categorical	Land use characteristics	1	Campus
			2	Tourist area
			3	Downtown
			4	Village area
			5	Medium, low density commercial area
			6	Compact residential area
			7	Low density residential area
Xs	Categorical	Traffic control type	1	No crosswalk and no signal
			2	signal without cross walk
			3	crosswalk without signal
			4	crosswalk with yellow cautious signal
			5	crosswalk with signal

Table 4. ANOVA Table of Model 1

Variable	DF	SSI	F-value (SSI)	P-value (SSI)	SSIII	F-value (SSIII)	P-value (SSIII)
LnP	1	9.846	37.65	<.0001	0.00026	0	0.9754
Xw	1	13.614	52.05	<.0001	2.067	7.9	0.0116
X_l	1	11.569	44.23	<.0001	2.8	10.71	0.0042
Xa	6	8.771	5.59	0.002	10.058	6.41	0.001
Xs	4	2.985	2.85	0.0541	2.985	2.85	0.0541

Table 5. the Multiple Means Comparison

Categorical variable	Levels number	Description	Tukey Method		Duncan Method
X_W	1	With sidewalk	A		A
	0	without sidewalk	B		B
X_l	1	2-lane highway	A		A
	0	4-lane highway	B		B
Xa	1	Campus	A		A
	2	Tourist area	A	B	B
	3	Downtown	C	B	B
	4	Village area	C	D	C
	5	Medium and low density commercial area	C	D	C
	6	Compact residential area	C	D	C
	7	Low density residential area		D	C
Xs	1	No crosswalk and no signal	A		A
	2	signal without cross walk	A		A
	3	crosswalk without signal	B		B
	4	crosswalk with yellow cautious signal	B		B
	5	crosswalk with signal	B		B

Table 6. Variable Information of Model 2

Categorical variable	Levels number	Description
Xw	1	With sidewalk
	0	without sidewalk
X_l	1	2-lane highway
	0	4-lane highway
Xa	1	Campus
	2	Tourist area and Downtown area
	3	Others
Xs	1	No crosswalk
	2	crosswalk

Table 7. ANOVA Table of Model 2

Variable	DF	SSI	F-value (SSI)	P-value (SSI)	SSIII	F-value (SSIII)	P-value (SSIII)
LnP	1	9.846	28.5	<.0001	0.856	2.48	0.1281
Xw	1	13.614	39.4	<.0001	2.805	8.12	0.0086
X_l	1	11.569	33.48	<.0001	4.928	14.26	0.0009
Xa	2	7.333	10.61	0.0005	7.275	10.53	0.0005
Xs	1	0.492	1.42	0.2441	0.492	1.42	0.2441

Table 8. ANOVA Table of Model 3

Variable	DF	SSI	F-value (SSI)	P-value (SSI)	SSIII	F-value (SSIII)	P-value (SSIII)
LnP	1	9.846	28.04	<.0001	0.899	2.56	0.1217
Xw	1	13.614	38.77	<.0001	6.159	17.54	0.0003
X_l	1	11.569	32.95	<.0001	4.927	14.03	0.0009
Xa	2	7.333	10.44	0.0005	7.334	10.44	0.0005

Table 9. Preliminary models for Weekly Pedestrian Exposure In Walk Area

Variable	Model 1	Model 2	Model 3	Model 4
Intercept	3.88	5.59	5.28	3.78
LnP	0.005	-0.23	-0.24	—
Without sidewalk	1.19	1.13	1.39	1.33
With sidewalk	Base	Base	Base	Base
2-lane highway	2.06	1.86	1.86	1.87
4-lane highway	Base	Base	Base	Base
Campus	2.42	3.09	3.1	2.62
Tourist area	2.36	1.09	1.1	0.88
Downtown	1.48			
Village area	1.24	Base	Base	Base
Medium and low density commercial area	0.86			
Compact residential area	-0.83			
Low density residential area	Base			
No crosswalk and no signal	-1.33	-0.47	—	—
signal without crosswalk	-0.85			
Crosswalk without signal	-1.28	Base		
Crosswalk with yellow caution signal	-0.95			
Crosswalk with signal	Base			
R square	0.909	0.832	0.823	0.805
Root MSE	0.511	0.588	0.593	0.609

Table 10. ANOVA Table of Model 4

Variable	DF	SSI	F-value (SSI)	P-value (SSI)	SSIII	F-value (SSIII)	P-value (SSIII)
Xw	1	22.306	60.05	<.0001	5.686	15.31	0.0006
X_l	1	12.585	33.88	<.0001	4.996	13.45	0.0011
Xa	2	6.573	8.85	0.0011	6.573	8.85	0.0011

Table 11. Model Selection*

Model selection	F-Value	F critical value	Conclusion
Model1~Model2	2.14	F(7,18 | 95%)=2.58	F-value<F-critical value Model 2 is better
Model2~Model3	1.42	F(1,25 | 95%)=4.24	F-value<F-critical value and *Xs* is discarded Model 3 is better
Model3~Model4	2.56	F(1,26 | 95%)=4.23	F-value<F-critical value and *LnP* is discarded Model 4 is better

*: 95 percent significance level is used in the model selection.

Table 12. Predicted Weekly Pedestrian Exposure Value List

Factor1	Factor2	Value	Campus	Tourist/Downtown	Other
Sidewalk	Two-lane highway	Predicted Value	14765	2592	1075
		Observed Average	14724	3047	1178
		Sample Size	1	15	5
	Four-lane highway	Predicted Value	2276	399	166
		Observed Average	N.O*.	N.O*.	187
		Sample Size	0	0	2
Without sidewalk	Two-lane highway	Predicted Value	3905	685	285
		Observed Average	N.O*.	N.O*.	255
		Sample Size	0	0	9
	Four-lane highway	Predicted Value	602	106	44
		Observed Average	N.O*.	N.O*.	N.O*.
		Sample Size	0	0	0

*: No = no sites observed in this cell

Appendix A: Estimate the Daily Pedestrian Crossing Volume

Counts	Barkhanmsted	Durham	Lakeville	conventry1	conventry2	Average pedestrian volume of 15-minute observation	Percentage of the 15-min observation of the whole day	Estimated the whole day observation Barkhanmsted	Durham	Lakeville
8:15				1	1	1	0.014925373			
8:30				0	0	0	0			
8:45				0	4	2	0.029850746			
9:00				0	0	0	0			
9:15				0	0	0	0			
9:30				2	0	1	0.014925373			
9:45				0	1	1	0.014925373			
10:00				2	1	2	0.029850746			
10:15				1	2	2	0.029850746			
10:30				3	0	2	0.029850746			
10:45				3	0	2	0.029850746			
11:00				3	1	2	0.029850746			
11:15				2	1	2	0.029850746			
11:30				2	0	2	0.029850746			
11:45			1	1	0	1	0.014925373			67
12:00			1		0	1	0.014925373			67
12:15			7		0	4	0.059701493			117.25
12:30	2		11		0	4	0.059701493	33.5		184.25
12:45	0		4		0	2	0.029850746			134
13:00	2		5			4	0.059701493	33.5		83.75
13:15	1		3			2	0.029850746	33.5		100.5
13:30	0	0	2			1	0.014925373			134
13:45	4	0				2	0.029850746	134		
14:00	3	1				2	0.029850746	100.5	33.5	
14:15	4	1		4	4	3	0.044776119	89.33333333	22.333333	
14:30		0		3	0	1	0.014925373			
14:45		1		1	3	2	0.029850746		33.5	
15:00		1		10	0	4	0.059701493		16.75	
15:15		0		1	0	1	0.014925373			
15:30				2	2	2	0.029850746			
.	.	.	.	.	.	.	.			
.	.	.	.	.	.	.	.			
Average								**71**	**27**	**111**

Appendix B Processing of Demographic Data

The demographic data used in this study was obtained from the 1990 census and customized for the geographic area of each site using digital street maps. For each study site an area encompassing the neighborhood within walking distance of the site was defined. The distance from which people were still considered to be likely to be walking was defined as 3600 feet, which equals 20 minutes when walking at a speed of 3 feet per second. Thus the neighborhood within walking distance was defined as a polygon encompassing all areas that are within 3600 feet when walking on streets.

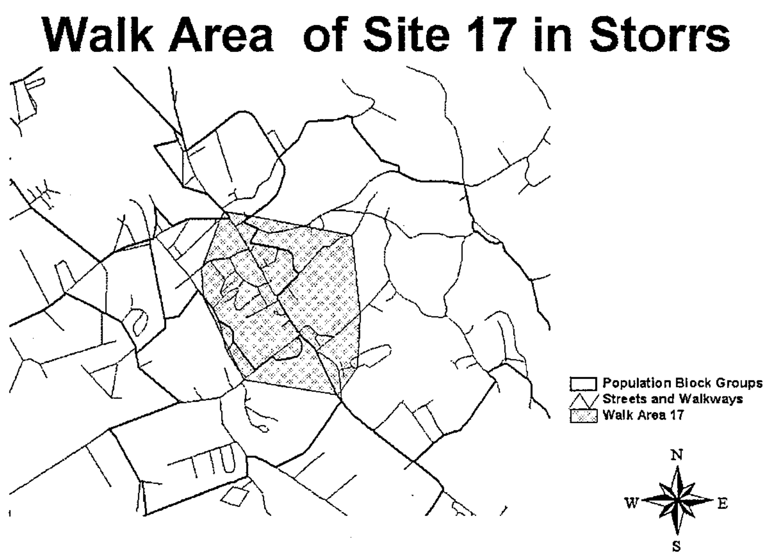

Figure B.1. Map of walk area in Storrs

For the most part, the service areas of the different sites intersected with two or more census block groups. The specific demographic data which was needed for this study, was obtained for each of these block groups from the 1990 census. Next, land use data for the different towns was used to determine the location of medium- and high-density residential and commercial development within each block group. The ratio R_{den} of developed area within each block group, which at the same time fell into the walk area of a particular site under investigation ($A_{den(bg\&sa)}$), to the developed area in the total block group ($A_{den(bg)}$) was established. The GIS procedures applied to perform this analysis are described in the Appendix B "GIS Procedures for Demographic Data Processing".

$$R_{den} = A_{den(bg\&w)} / A_{den(bg)}$$

Where:

$A_{den(bg\&sa)}$ = Densely developed area in block group that at the same time falls into the walk area of the site,

$A_{den(bg)}$ = Densely developed area in block group,

Developed area in a rural context mostly also represents populated area. Therefore this ratio was used to estimate the population and other demographic data in the vicinity of each site based on the demographic data of the adjacent block groups. The process of computing an estimate of demographic data for the vicinity of each study site, which is described above, is exemplified for the study site in Storrs in Table Data Collection and Processing 3. Thus the demographic data needed for this study, total population, age groups and median household income in 1989, was assessed for each block group in the vicinity of the study sites and translated to the scale of the different service areas using land use data. Age group and median household income data of the different block groups was weighed according to the population of the block group that lives within the walk area of the site.

Appendix C. GIS for Demographic Data Processing

The ArcView GIS Processing of demographic and geographic data is exemplified by the procedures employed to process the data for the site in Storrs, Mansfield. The other sites were treated the same way, only the numbering of towns, block groups and walk areas differs.

Files Necessary for Data Processing

The town code for Mansfield is 078. The necessary data and maps can be downloaded from the magic homepage http://magic.lib.uconn.edu. The necessary files are:

- Land use
- Population block group
- Streets

These files are converted to shape files:

- Lu078.shp
- Pg078.shp
- St078.shp

Extracting the Densely Developed Areas

The query builder tool is employed to extract those areas in Mansfield that feature Medium and High density residential and commercial development from the land use file. These polygons are converted to a shape file (den078.shp).

Establishing the Walk Area

The ArcView Network Analyst is employed to establish the area that is within a walking distance of 3600 feet of the site. The network used is the street network of Mansfield (st078.shp). The walk area is converted to a shape file (wa17.shp).

Establishing the Densely Developed Area in Each Block Group

Encompassed by the Walk Area

In Mansfield the block groups with the internal ID 7, 8 and 12 fall in the vicinity of the site and have to be included in the data processing. The data processing of block group 7 data serves as an example for the other block groups.
The necessary data files are:

- Den078.shp
- Pg078.shp
- Wa17.shp

Establishing the densely developed area in the total block group (7)

1. Select block group 7 from file pg078.shp and create a new shape file pg078_7.
2. Clip the densely developed areas in Mansfield (den078.shp) to the boundaries of block group 7 (pg078_7). This will yield a new file containing the densely developed areas within the block group. Name file den(pg078_7).

Establishing the densely developed area in the block group (7) that at the same time falls into the walk area of site 17

1. Union the block group coverage for the town of Mansfield (pg078.shp) and the walk area coverage for site 17 (wa17.shp). Name the new file unpgwa78_17 or union1.
2. From the new file (union1) select the area that falls into block group 7 and into the walk area 17 at the same time. Convert this area to a new shape file (pg078_5 & wa17).
3. Clip the densely developed areas in Mansfield (den078) to the boundaries of pg078_5 & wa17. The new file will show densely developed areas in block group 7 that are encompassed by the walk area boundary too. Name it den(pg078_7 & wa17).

Summing up densely developed areas in Excel

The tables of the shape files den(pg078_7) and den(pg078_7 & wa17) can be opened in Excel and the sum of the densely developed area in the block group 7 and the densely developed area in 7 that at the same time falls into the walk area of site 17 can be established. This number is used in the block group spreadsheet to estimate the population and demographic data in the walk area of each site based on the demographic data of the block group and the distribution of developed area in the block group.

PART II: HIGHWAY SAFETY ON TWO-LANE, UNDIVIDED ROADWAYS ON RURAL AND SUBURBAN LOCATIONS

ABSTRACT

Police Accident Reports reveal that in a five year period between 1993 and 1997 that there were 892 crashes on two lane, undivided roadways in Strafford County, New Hampshire, a county consisting of suburban and rural communities. Among them are 300 motor vehicle crashes leading to 423 injuries and 33 deaths. The purpose of the study is to identify site characteristics, i.e., factors describing land use activity, roadside design, merging and crossing traffic, traffic control and vehicular speed, that can be used to predict roadway risk, the probability of a crash leading to injury and/or death. The site is characterized by its land-use activity, a residential, shopping, or village zone. Residential and shopping zones are single purpose zones consisting mostly of single-family dwelling units and roadside shopping areas with ample off-street parking. A village zone is a multi-purpose zone permitting a combination of activities found in residential, shopping and commercial areas. Sites in a typical village site are characterized as being pedestrian friendly, that is, having sidewalks, crosswalks and permitting on-street parking. The presence or absence of stop signs, yield signs and signals, sidewalks, crosswalks and street parking (a surrogate measure of speed) and crash type (single vehicle or multiple vehicle crash) as well as land use activity are used as categorical variables and paved shoulder width (a surrogate measure of speed) and intersecting roadway traffic flow are used as continuous variables in a logistic regression. All variables, with the exception of crosswalks, street parking and paved shoulder width, proved significant. Typical village and residential sites proved to be the least hazardous and shopping sites the most hazardous. The probability of a crash at a site in a shopping zone is about four times as great as a crash of a site in either a village or residential zone. The implications of the findings are discussed.

INTRODUCTION

The primary source of motor-vehicle accident information used in this study comes from Police Accident Reports (PAR). Each accident record from an annual PAR contains information about the vehicle involved in a crash. The crash location and personal injury information are most important for this study. These records are used to calculate crash counts n_c and injury counts n_i for 82 undivided, two-lane roadway sites in rural and suburban communities in Strafford County, New Hampshire. The n_c and n_i counts are used to calibrate two logistic models:

1. to predict the probability of a crash ϖ_c, and
2. to predict the conditional probability of a injury given a crash has occurred $\varpi_{i|c}$.

A key feature of this modeling approach is to identify statistically significant factors associated with the probability of a crash at the site. The factors include land use activity, roadside design, merging and crossing traffic, traffic control and vehicular speed.

Before introducing the models, a brief description of how the PAR data are compiled and used to calculate n_c and n_i. This is followed by a discussion on why it is important to introduce roadway exposure e into the risk assessment process. The concept of individual lifetime risk is also presented and used to identify hazardous sites. Finally, the site characteristic factors are defined, logistic regression models of ϖ_c and $\varpi_{i|c}$ are calibrated, and then used in a risk assessment of village, residential and shopping sites.

RISK ASSESSMENT METHODS

Crash and Injury Counts: A PAR record exists for each vehicle involved in the crash. An accident report is filed for a crash with a minimum property damage of $1,500 and/or personal injury. For example, a single-vehicle crash will contain one PAR record and a multi-vehicle crash involving three vehicles will contain three PAR records. The number of PAR records have no bearing on the n_c crash count. Single- and multi-vehicle crashes each count as one crash in determining n_c.

A crash leading one or more personal injuries and/or one or more deaths is counted as one injury crash. For example, a crash involving one vehicle with a motor-vehicle occupant injury and a pedestrian death is counted as one personal injury crash in the determination of n_i. A crash involving three vehicles with no personal harm is not counted in the determination of n_i.

Over a five year period, there were a total of 892 crashes with a total of 300 injury crashes leading to 58 pedestrian injuries, 3 pedestrian deaths, 365 occupant injuries and 30 occupant deaths. In order to give a feel for the magnitude of n_c and n_i counts for individual sites, three sites have been selected as shown in Table 1.

Risk Rankings: The site with the largest count is considered to be the most hazardous, the second largest count is the second most hazardous, and the smallest count is the least hazardous. The ranking from most to least hazardous site for the n_c counts is: {A, B, C}. Given the magnitude of the n_c counts, site A appears to be much more hazardous than either B or C. The magnitude of the n_c counts for site B and C suggests that site B is more hazardous than site C.

When the n_i counts are used, a slightly different perspective is obtained. The site ranking for the n_i and n_c counts are the same. Like the n_c counts, site A appears to be the most hazardous site because the n_i count for site A is so much greater than either site B or C. Unlike the n_c counts, the magnitude of the n_i counts for sites B and C suggest that these sites are equally hazardous.

Roadway Exposure and Risk: The n_c and n_i counts and hazardous rankings fail to explain why site A is more hazardous than the other two sites. The traffic volumes for the three sites, measured as an average daily traffic (ADT) in units of vehicles per day (vpd), are given in Table 2. Intuitively, since site A has a largest traffic volume, more crashes are expected here than at the other two sites. In other words, any vehicle that passes through a study site can potentially be involved in a crash. For a five year period, the risk exposure e is estimated with the ADT volumes as e = (5 years) (365 days per year) ADT. Clearly, site A has the greatest exposure and greatest potential for motor-vehicle crashes.

Individual Lifetime Risk: A key question is whether or not these sites are considered hazardous. The question is addressed with the concept of individual lifetime risk (Ossenbruggen, in press).

The probability that a person will die prematurely (before the age of 70 years) in a motor vehicle crash is called the individual lifetime risk and is designated as θ^*. Given θ^* = 1 chance in 1,000 or 10^{-3} and a person makes one motor-vehicle trip each day, the probability of a fatal crash in one daily trip is calculated as $\theta = (10^{-3}) / (70 \text{ years} \times 365$ days per year$) = 3.91 \times 10^{-8}$.

The value of θ is treated as an acceptable risk [1] and the definition of θ is broadened to include probability that motor vehicle crashes will lead to injury as well as death. The expected number of crashes in a five year period is the product θ_e where e is the five year exposure. The acceptable risk expressed as a expected value for a five year period an is $R^* = \theta\_e$. A site is deemed safe if n_i _ R^* and hazardous if $n_i > R^*$. Since $n_i > R^*$ for sites A, B and C as shown in Table 3, all sites are deemed hazardous. Clearly, the goal to achieve an acceptable risk of $\theta^* = 10^{-3}$ is a stringent requirement.[2]

Probabilistic Methods: Proportions[3] have important properties that make them useful for risk assessment. Crash and injury proportions are easily calculated with crash and roadway exposure data as $p_c = n_c / e$ and $p_i = n_i / e$, respectively. The proportions for sites A, B and C are given in Table 4.

Properties considered important for risk assessment are summarized as follows:

1. Motor-vehicle crash and injury events are rare. Given $p_c = 19.06 \times 10^{-7}$ _ 2.0 x
   10^{-6}, the largest value from Table 4, for example, a person observing 1,000,000 vehicles pass through site A would expect to see two crashes.
2. The p_c and p_i values, which incorporate roadway or risk exposure e, can give different perspectives on the severity and nature of the roadway risk

---

[1] If a person makes x trips per day, then the acceptable risk is reduced to $3.91 \times 10^{-8} / x$.

[2] Risk is often expressed a crash rate in units of vehicle miles traveled (VMT) where $VMT = e\_d$, the product of exposure e and roadway segment length d. The crash rate θ^{**} and lifetime risk p are related; therefore, like θ or θ^*, θ^{**} can be treated as an ôacceptableö risk or roadway safety goal. The product p_θ^{**} is the probability a crash occurs in a roadway segment of length d. Assuming $\theta = \theta^{**}\_d$, the segment length d is calculated as $d = \theta / \theta^{**}$ or d = 3.91 miles.

Risk expressed as θ is a point measure which is most suitable for describing risk at a crash site at a roadway intersection or within close proximity of an intersection. In this paper, θ is the preferred measure because the crashes under investigation are reported at roadway intersections. The important point, given either assignment $\theta = 3.91 \times 10^{-8}$ at an intersection or $\theta^{**} = 10^{-8}$ per VMT for a 3.91 mile road segment, is that the two assignments are equally challenging goals to achieve in practice. None the sample sites given in Table 3 satisfy the ôsafety standardö of n_i _ R^*.

[3] The terms, proportions and probability, are used synonymously.

as compared to values of n_c and n_i counts. The hazardous site ranking using p_c and n_c values of Tables 1 and 2 are the same, {A,B,C}, but the ranking for the p_i values is {A,C,B} and for the n_i values is {A,B,C}. The injury crash proportion is almost twice as great at site C as at site B.

3. The p_i and p_c values are approximately the same magnitude; therefore, it is advantageous to discern the relationship between crashes and injury outcomes with conditional probability. Conditional probability $p_{i|c}$, the likelihood of an injury given a motor-vehicle crash has occurred, is estimated as $p_{i|c} = n_i / n_c$. The $p_{i|c}$ values given in Table 5 show that there is about a one in four chance of an injury at sites A and B and certain chance of an injury at site C, i.e., $p_{i|c} = 1$. In other words, every crash at site C will result in one or more injuries or deaths. Clearly, this conclusion is too strong for a calculation based on a small sample count, $n_i = n_c = 2$. By introducing n_c, n_i and e from all 82 sites where there are a total of 300 injury crashes into the logistic regression analysis, the difficulty of small samples size is resolved.
4. The relative risk is an effective method for evaluating pair-wise events. The relative risk for crashes at sites A and B is calculated as the ratio of the p_c at site A to p_c at site B, or 19.06 x 10^{-7}/4.09 x 10^{-7} = 4.65. In other words, the probability of a crash at site A is 4.65 times more likely than a crash at site B. The relative risk for sites A and C is 7.84, the probability of a crash at site A is 7.84 times more likely than at site C.
5. The odds ratio and relative risk give similar values when the proportion of crashes is small. The odds ratio for sites i and j is $\Omega_{i,j} = \frac{n_i (e_j - n_j)}{n_j (e_i - n_i)}$. The odds ratios for sites A and B is $\Omega_{A,B} = 4.65$ and sites A and C is $\Omega_{A,C} = 7.84$, the same values calculated for relative risks.

Simple proportions and statistical summaries give a more insightful interpretation than counts alone.

LOGISTIC REGRESSION

Once again, it leads one to ask, why is site A more hazardous than sites B or C? There are site characteristics that suggest an answer to this question and to other questions dealing with the relationship between crash and injury outcomes.

Site Characteristics: Sites A, B and C, which are situated within the same town boundary, Durham, NH, are classified as being shopping, village and residential zones, respectively.

Site A, known as Gasoline Alley, caters to convenience shopping and automotive needs. Five service stations, a used car dealership and a convenience store are in the vicinity of the site. There is ample off-street parking with no curbing in some places and driveways with wide curb cuts in other places. The parking lots extend directly to the roadway edge. There is a white line painted on the pavement that separates the travel lane and paved shoulder. Without benefit of clearly delineated sidewalks, pedestrians are observed walking in the paved shoulder region. Vehicles are often observed driving in the paved shoulder region to avoid stopping behind vehicles in the travel lane that are stopped making left-hand turns.

Site B, on the other hand, has a stop sign, a cross walk, granite curbs, sidewalks and on-street parking. It is considered to be pedestrian friendly. The restaurants, grocery stores and other shops that surround the site rely heavily on pedestrian traffic for business.

Roadside design appears to affect pedestrian street traffic and to be a critical in explaining why the probability of a crash is 4.65 times more likely at site A than at site B. Site B is pedestrian friendly and site A is not.

However, roadway design does not explain why the probability of a crash at site A is 7.84 times more likely than at site C. Sites A and C, which is zoned residential, contains no traffic controls, curbed sidewalks and other roadway design features that would make them pedestrian friendly. Other factors are needed to explain the difference in roadway risk for these two sites.

Six factors considered important and investigated in this study are:

1. *Land Use Activity:* Are sites located in village and residential zones inertly safer than sites located in shopping zones?
2. *Roadside Design:* Do any, some or all the factors cited for a pedestrian friendly environment contribute to roadway risk?
3. *Traffic Control:* How much do stop signs and other traffic control devices contribute to safety?
4. *Merging and Crossing Traffic*: Do conflicts and delay caused by merging and crossing vehicles from side streets contribute to roadway risk?
5. *Speed:* To what extent does vehicular speed play a role in roadway risk?
6. *Crash Type:* Is the probability of single- and multi-vehicle crashes and injuries the same or different?

Justification for introducing these factors into study are as follows.

Land Use Activity: A site is characterized as being located in a village, residential or shopping zone. The characterization are based on the land use ordinance and on on-site evaluation. A village permits multi-purpose activities including residential, shopping and business activities. Compared to residential and shopping zones, both single-purpose zones, pedestrian street traffic in village sites is most active.

Many single purpose shopping sites have the attributes of strip malls. Virtually all pedestrian traffic is observed in parking lots built in front of the buildings and little pedestrian street traffic is observed or encouraged.

Residential zones have low population densities starting with one single dwelling unit per quarter acre and larger. Pedestrian street traffic is relatively low consisting mainly of dog walkers, joggers and strollers. The statistics given in Table 6 compare crash and crash injury counts by land use activity.

Roadside Design and Traffic Control: The percentage of sites with various pedestrian amenities and using traffic control devices are listed in Table 7. Given the high percentage of village sites with crosswalks, sidewalks, traffic control devices and on-street parking, a typical village site is considered to be pedestrian friendly and typical shopping and residential sites are not.

Merging and Crossing Traffic: When traffic flow increases or the speed on the main road increases, drivers attempting to merge and to cross from intersecting streets and driveways find it difficult. The time gap between arriving vehicles on the main roadway decreases with increasing speed. Drivers making a left-hand turn from a side street or driveway can find it difficult because they must cross a lane of on-coming traffic and then merge into another one. Similarly, drivers on the main road attempting to make left turns often must stop in the main travel lane and wait for a sufficient time gap in the on-coming traffic stream before completing their turns. Under these circumstances, some drivers become impatient and often take chances; thereby increasing the crash risk.

In addition, when the traffic flow from a side street or driveway increases, there are more merging conflicts and longer delays. Intuitively, the greater the traffic flow on the main and intersecting roadways, the greater the number of conflicts, the longer the delay and the greater the risk for a crash. Consequently, the traffic flows on the main and intersecting roadways are used as surrogate measures of the merging and crossing traffic.

Table 8 gives the main and intersecting traffic volumes by land use activity. Given the relatively large main and intersecting traffic volumes, shopping zones have the most traffic conflicts and longest delays as compared to residential zones which have the least conflicts and shortest delays.

Speed: Paved shoulder width and on-street parking are introduced into the study as surrogate measures of speed.

Sites with on-street parking tend to have lower average operating speeds than sites without parking. Speed is correlated with driving lane and paved shoulder width. The driving lane width at all sites ranged between 11 and 12 feet; therefore, it is considered a constant and not introduced into this study. The average paved shoulder width, on the other hand, shown in Table 9, have a great deal of variability. Given the relatively large average shoulder width, a typical shopping zone is expected to have a higher speeds than a typical village or residential site. Given restricted speed signs of 45 mph or less were observed at many residential and shopping sites, there is a concern for excessive speeding.

Crash Type: The vehicle that crashes into pedestrians and/or object is called a single-vehicle crash and a vehicle that crashes into one or more vehicles is called a multi-vehicle crash. The responsibility for the crash rests entirely on the driver involved in a single-vehicle crash and is shared among drivers involved in a multi-vehicle crash. Drivers involved in single-vehicle crashes are more often found violating traffic laws. They are generally considered imprudent drivers.

Logistic Regression Models: Logistic regression for crash ϖ_c and injury crash $\varpi_{i|c}$ models have the same model form, $\varpi = e^u/(1+e^u)$. The subscript is dropped for simplicity. The probability of no crash or injury is $1-\varpi = 1/(1+e^u)$. The symbol u is a linear predictor function of explanatory variables. The list of variables, which are hypothesized to be significant, are given in Table 10. Shoulder width S_w and the traffic flow on the main F_m and intersecting F_i roadways are continuous variables. All remaining variables are categorical.

Calibrated Models: The method of maximum likelihood is used for model calibration and the likelihood ratio and t-tests are used to test the significance of model coefficients (Hosmer and Lemshow, 1991, Agresti, 1990). Table 11 contains a list of the variables and their significance levels. The linear predictors for the crash model is:

$$u_c = a_0 + a_1\_T + a_2\_M + a_3\_W + a_4\_R\_F_i + a_5\_S\_T\_F_i$$

and the injury crash model is

$$u_{i|c} = b_1\_T + b_2\_F_i \, .$$

The crash model contains main effects and interactive factors and the injury crash model contains main factors only. The main effects factors are: T, M, W and F_i and the interactive effects factors are: R_F and $S\_T\_F_i$. With the exception of crosswalk C and the surrogate measures of speed , on-street parking P and sidewalks S_w, all hypothesized site characteristic factors are found in the crash model, including variables describing land use activity R, S and V, a roadway design feature W, traffic control M, merging traffic flow F_i and crash type T.

Statistically Insignificant Factors: Crosswalk C, on-street parking P and shoulder width S_w proved to be statistically insignificant variables for the vehicular crash study conducted here. In contrast, crosswalks and parking proved to be significant in pedestrian safety studies in urban regions.

Evidence, however, shows crosswalks to have limited safety benefits for pedestrians. In a five year period in San Diego, CA, there were 177 pedestrians hit in 400 marked crosswalks compared to 31 pedestrians hit in 400 corresponding crosswalks. The ineffectiveness of crosswalks in reducing pedestrian risk appears to be a reflection of pedestrian attitude and lack of caution (Herms, 1970). Furthermore, law enforcement efforts directed at motorist violating crosswalk laws failed to demonstrate that an increase in the number of drivers willing to stop for pedestrians (Britt, Bergman and Moffat, 1995, Tidwell and Doyle, 1995). The fact that C is an insignificant variable appears to be inconsistent with this reported evidence. Regardless, crosswalks are assumed to have no measurable affect on roadway risk in this study.

The association of vehicles parked on the street with injury risk proved significant for resident neighborhoods in Orange County, CA (Agran, Winn, Anderson, Tran and Valle, 1996). Midblock dart-out emerged as the single most common trait for pedestrian and bicycle injuries in Long Beach, CA (Kraus, Hooten, Brown, Peek-Asa, Heye and McArthur, 1996). The small amount of pedestrian street traffic observed at residential and shopping sites might explain why P is insignificant in this study.

Paved shoulder width S_w proved to be a significant variable in a study of undivided, two-lane roadways at urban, suburban and rural sites in CT (Ivan, Pasupathy, and Ossenbruggen, in press). For single-vehicle crashes, the crash rate decreases with an increase in S_w and for multi-vehicle crashes, the crash rate decreases with a decrease in S_w. These results are consistent with the notion that drivers involved in single-vehicle crashes act imprudently and use excessive speed. Increasing the paved shoulder width gives these drivers an extra margin of safety to avoid off roadway collisions with fixed objects and rear-end collisions (Ogden, 1997). Narrowing the paved shoulder width has the affect of reducing operating speed therefore reducing the risk of a multi-vehicle crash (Shankar, Mannering and Barfield, 1994).

While the surrogate measures of vehicular speed P and S_w prove to be insignificant, there is evidence to support the claim that vehicular speed does play an important role. Consider speeding or driving at a speed in excess of a speed that is considered reasonable and safe. When speeding is judged to endanger, posted speed limit signs are erected. Given the number and location of posted speed limit signs in the study region, speeding is judged to be more of a problem in residential and shopping zones than in village zones. Speeding and how it is incorporated in the models is addressed in the Discussion subsection, *Speed and Speeding*.

A Risk Assessment: For simplicity, the crash and injury crash models are used to assess sites classified as typical village, residential and village sites. If 50% or more of the sites in a land use activity class have sidewalks, then $W = 1$; otherwise, $W = 0$. Similarly, if 50% or more of the sites in a class have traffic control devices, then $M = 1$; otherwise, $M = 0$. Simplified linear predictor function by site class are given in Table 12. The linear predictor functions for the crash and injury crash models for each class reduce to functions of crash type T and the intersecting traffic flow F_i. The average traffic flows by site class are assigned to be F_m and F_i of Table 8.

The risk assessment for a typical site class is similar to the one used for the three sample sites A, B and C. As before, the concept of individual lifetime risk is used. Recall that a site is deemed safe if $n_i \_ R^*$ and is deemed hazardous if $n_i > R^*$. In lieu of the n_i count, the expected number of injury crashes is used where the expected number of injury crashes is calculated as $N_i = e\_\varpi_i$ with e equal to five-year exposure for a typical site class or $e = (5)\,(365)\,F_m$ where the probability of an injury crash is:

$$\varpi_i = p(T=0)\_\varpi_{i|c}(T=0,F_i)\_\varpi_c(T=0,F_i) + p(T=1)\_\varpi_{i|c}(T=1,F_i)\_\varpi_c(T=1,F_i).$$

Similarly, the expected number of crashes for a typical site class is calculated as $N_c = e\_\varpi_c$ where the probability of a crash is:

$$\varpi_c = p(T=0)\_\varpi_c(T=0,F_i) + p(T=1)\_\varpi_c(T=1,F_i)$$

The proportion of the single and multi-vehicle crashes for a typical site class are denoted as $p(T=0)$ and $p(T=1)$. The values of F_i, $p(T)$, $\varpi_c(T,F_i)$ and $\varpi_{i|c}(T,F_i)$ are given in Table 13 and the ϖ_c, ϖ_i, N_c, N_i and R^* are given in Table 14 by typical site class.

All sites are deemed hazardous because $N_i > R^*$. Comparing the expected values N_c and N_i and the magnitudes of the difference between the expected number of injury crashes and acceptable risk, $N_i - R^*$, the most hazardous site is a typical shopping site, next most hazardous is a typical residential site, and the least hazardous by a typical village site.

The adjusted odds ratios offer insight into the nature of crashes at the various sites. The adjusted odds ratios for factor j is simply calculated as $\Omega = e^{a_j}$ where a_j is the model coefficient for factor j. In other words, all factors with the exception of j are the assigned the same value; therefore, the affect of factor j on the odds is adjusted to consider it alone.

Odds Ratios for Crashes: When the probability of an event is small as in the case of crash events, the odds ratio is a good estimate of relative risk.

Crash Type: The probability of a multi-vehicle crash is two times (2.04 from Table 15) more likely than a single-vehicle crash.

Traffic Control: The probability of a crash is almost two times (1.91) more likely at a site using traffic control than a site without control. This result suggests that a site requiring traffic control is more hazardous than a site not requiring it. It does not suggest that introducing a traffic control will increase the crash risk or removing it will reduce the risk. It merely suggests that a site requiring traffic control is inherently more hazardous than a site not requiring control.

Roadside Design: The probability of a crash with a sidewalk is approximately one-half (0.49) times as likely than a site without a sidewalk.

Merging and Crossing Traffic: The linear predictor u_c for a typical village site is independent of intersecting traffic flow F_i, therefore, a village site is considered the baseline site. The probability of a crash at a typical residential site with intersecting traffic flow F_i = 1,604 vph is 1.15 times more likely at a typical village site. Similarly, the probability of a crash at a typical shopping site with intersecting traffic flow F_i = 4,291 vph is 1.43 times more likely than a typical village site. Increasing flow F_i causes reduces speed increasing traffic delay and congestion. It suggests that increased traffic delay and congestion are associated with increased crash risk.

Adjusted Odds Ratio for Injury Crashes: Since the conditional probabilities of an injury given a crash are large, the odds ratio is not a good estimate of the relative risk.

Crash Type: The odds of an injury are 1.76 (1/0.57) times higher for a single-motor vehicle crash than for a multi-vehicle crash.

Merging and Crossing Traffic: Injuries are 0.93 times less likely to occur in residential site with a traffic flow of F_i = 1,604 vpd than for a site with no intersecting traffic flow or F_i = 0 vpd. From Table 15, the odds ratios decrease with increasing intersecting flow F_i. It suggests that traffic congestion and delay which accompanied with lower speed F_i reduces the severity of injury.

DISCUSSION

Preconceived Notions and Reality: During the on-site evaluation and prior to performing the modeling study, it was predicted that shopping and village sites would be most hazardous and the residential sites to be the least hazardous. More traffic delay and merging conflicts were observed at shopping and village sites than at residential sites. In addition, the pedestrian traffic flow was substantially less at shopping and residential sites as compared to the village sites. It was felt that the greater pedestrian street traffic would be compensating factor thereby making village sites as hazardous as shopping sites.

As a result, it was a surprise that the model predictions indicate a typical village site to be least hazardous. The expected number of crashes N_c and injury crashes N_i for a typical shopping site are almost four times larger as the expected number of crashes

N_c and injury crashes N_i for a typical village site. In addition, since the difference between the expected number of injury crashes and the safety standard, N_i - R^*, is substantially larger for a typical shopping site than the differences for either a typical residential or village site, it suggests a typical shopping site is in more serious violation of the acceptable risk standard than either a typical residential or village site.

The model helps explain why a typical shopping site is the most hazardous. It (1) has the greatest motor-vehicle traffic flow F_m, therefore the greatest risk exposure, (2) has the greatest traffic flow F_i, therefore the greatest number of merging conflicts and longest delays, and (3) lacks sidewalks making it the least pedestrian friendly. In hindsight, shopping sites are the most hostile to pedestrian street traffic. With the exception of individuals walking to and from their vehicles to shops, little pedestrian street traffic was observed. Most shops were too far from residential neighborhoods for walking. Owing to high traffic flow and vehicular speed, they are uninviting to walkers.

The model helps explain why a typical village site is the least hazardous. It (1) has sidewalks where neither a typical residential nor shopping site have them, (2) has multiple land use activities with substantially more pedestrian street traffic than observed at any residential or shopping sites, and (3) has the infrastructure that is useful in promoting safety through the reduction of driver speeding (de Ward, Jessurum, Steyvers, Raggatt and Brookhuis, 1995; Garder, 1989). Village sites are pedestrian friendly!

Speed and Speeding: While surrogate measures of speed proved insignificant in the model building process, there is evidence that speed and excessive speed are associated with roadway risk. Furthermore, there is a general theme in the traffic engineering literature that traffic control enhance safety, but definitive evidence is difficult to generate (Evans, 1991). Posted speed and yield and stop signs are erected at sites judged to be hazardous. No doubt they provide safety benefits, but they unable to reduce risk to the same level as a typical village site. Consider the following evidence:

First, a site using a traffic control device according to the adjusted odds ratio is about twice (Ω = 1.91) as likely to have a crash than a site without one.[4]

Second, a model building study for urban, suburban and rural crash sites in CT shows the probability of a crash to be higher at sites with more restrictive posted speed limits than at sites with less restrictive posted speed limits (Ossenbruggen, in press). For example, the probability of a crash is greater at a site with a 35 mph speed limit sign than one with a 45 mph speed limit sign.

Third, the same result is found in this study. As mentioned previously, speed limits are posted at sites where speeding is observed and judged hazardous. Using the number of restrictive speed limit signs erected as a means of ranking sites, village zones are deemed least hazardous, residential zones are more hazardous, and shopping zones are most hazardous. This ranking is the same ranking using the ϖ_c and ϖ_i predictions of typical village, residential and shopping sites of Table 14. Speeding was four times more likely to be near a convenience store, gas station or fast food store (Kraus, *etal*,

---

[4] A traffic control device is a stop sign, yield sign or signal. If one of these devices are present, M = 1, otherwise, M = 0. While a posted speed limit sign is a traffic control device, it is not considered when determining the value of M.

1996). It is interesting to note that the crash risks ϖ_c and ϖ_i are also about four times more likely to occur at a shopping site than at a residential or village site.

Four, village sites, which are generally free of posted speed signs, have the greatest impediments to speeding. Drivers at a typical village site must contend with: (1) pedestrian street traffic, (2) vehicles parked on the street, and (3) a traffic control device, usually a stop sign that interrupts traffic flow. In contrast, a driver at a typical residential site sees little pedestrian traffic, no traffic control or no on-street parking and a driver at a typical shopping site sees little pedestrian traffic and on-street parking. Drivers maintain higher average operating speeds through residential and shopping sites than through village sites. The distinguishing feature among the sites is village sites are pedestrian friendly and residential and shopping sites are not.

The Ideal Site: An ideal site is defined to be a site that has the maximum average operating speed $\overline{u}$ while satisfying the acceptable risk constraint, N_i _ R^* or ϖ_i _ θ. In other words, *maximize* $\overline{u}$ subject to ϖ_i _ θ = 3.19 x 10^{-8}. The optimum solution to this problem is: $V^* = W^* = 1$, $R^* = S^* = T^* = M^* = 0$ and $F_i^* = 3{,}000$ vpd. This is a description of village site with no traffic control and sidewalks involving only single-vehicle crashes. That is, $p(T^* = 0) = 100\%$.

Clearly, this is a theoretical solution. The variables of T and F_i are not controllable. Crashes are probabilistic events; therefore, specifying $p(T^* = 0) = 100\%$ is an impossibility. Similarly, imposing a traffic flow $F_i^* = 3{,}000$ vpd to satisfy the constraint is another impossibility. Traffic flow is a function of demand and can not be controlled.

The optimum solution, while theoretical, can be used as a policy, planning and design goal. The model contains five controllable variables, the land use activity variables R, S and V, the roadside design variable W, and a traffic control variable M. When feasible, (1) make a site have the characteristics of a village, (2) provide sidewalks, (3) use no traffic controls, and (4) reduce risk exposure e.

Practical Application 1: Consider reconstructing Gasoline Alley, a strip mall, into an ideal site. First, adding sidewalks to Gasoline Alley is expected to derive safety benefits. Sidewalks will protect pedestrians who currently use the paved shoulder as a walkway. Second, Gasoline Alley has the potential to be transformed into a village site. With incentives, it may be possible to convert some existing businesses to shops and restaurants. Single family homes are contiguous to the site and the site is within a 10 minute walk of the central business district; therefore, it has the potential to be pedestrian friendly area. Steps 3 and 4 are considered infeasible.

Practical Application 2: Now consider reconstructing seventeen shopping sites without sidewalks. First, reconstructing these sites with new sidewalks should achieve additional safety benefits. Second, unlike Gasoline Alley, most of these shopping zones are isolated, making it imperative to drive to them. There is no supportive infrastructure to increase volume of pedestrian street traffic. Third, many of these sites have posted speed limit signs but no traffic controls. According to crash model, there is no safety advantage to introduce traffic controls but there is an advantage to reduce the crash risk ϖ_c by reducing F_i. Four, by reducing risk exposure e, or equivalently F_m, the expected number of crashes N_c and injuries N_i can be reduced. Possibly, redirecting existing traffic through better highway network management or improving public transit service will reduce F_i and F_m.

These examples show that some of the features of an ideal site can be achieved. However, once the land has been developed, altering land use activity to resemble a village site as illustrated in Practical Application 2 can be very difficult or impossible.

CONCLUSIONS

Logistic regression analysis is used to identify statistically significant factors that are associated with roadway risk, the probability of a crash and injury. Factors that proved to be significant define the crash site. There are categorical factors that describe the site by land use activity, the presence of sidewalks, the use of traffic control devices and crash type and a continuous factor that describes intersecting traffic flow causing conflict and delay. The association of speed and roadway risk is discussed. The models, including the use of odds ratios, are used to describe the nature of roadway risk, to perform risk assessments, and to define an ideal site.

Shopping sites are the most hazardous. Typical shopping sites proved to be most hazardous with a probability of a crash about four times greater than either a typical residential or village site. Village sites , which are considered to be pedestrian friendly, proved to be the least hazardous and therefore come closest to satisfying the conditions of an ideal site. Typical residential sites are slightly more hazardous than typical village sites in spite the fact village sites have relatively more traffic delay and merging conflicts than residential sites.

REFERENCES

1. Agran, P.F. Winn, D.G., Anderson, C.L., Tran C. and Valle, C.P. (1996) The role of the physical and traffic environment in child pedestrian injuries, *Pediatrics*, **98**, 1096-1103.
2. Agresti, A. (1990) *Categorical Data Analysis*, Wiley-Interscience, New York.
3. Britt, J.W., Bergman, A.B. and Moffat, J. (1995) Law enforcement, pedestrian safety, and driver compliance with crosswalk laws: campaign in Seattle, *Transportation Research Record*, **1485**, 160-167.
4. de Ward, D. Jessurum, M. Steyvers, T.J. Raggatt, P.T., and Brookhuis, K.A. (1995) Effects of road layout and road environment on driving performance, driversÆ physiology and road appreciation, *Ergonomics*, **38**, 1395-1407.
5. Evans L. (1991) *Traffic Safety and the Driver*, Van Nostrand Reinhold, New York, 85.
6. Garder, P. (1989) Pedestrian safety at traffic signals carried out with the help of the traffic conflict technique, *Analysis and Prevention*, **21**, 435-444.
7. Herms, B.F. (1970) *Pedestrian Crosswalk Study: Accidents in {Painted and Unpainted Crosswalks*, San Diego, Police Department, Traffic Bureau.
8. Hosmer, D.W. and Lemeshow S. (1989) *Applied Logistic Regression*, Wiley-Interscience, New York.
9. Ivan, J. Pasupathy, R. and Ossenbruggen, P.J. (in press) Differences in causality factors for single and multi-vehicle crashes on two-lane roads, *Accident Analysis and Prevention*.

10. Kraus, J.F., Hooten, E.G., Brown, K.A., Peek-Asa, C., Heye C. and McArthur, D.L. (1996) Child pedestrian and bicyclist injuries: results of community surveillance and a case-control study. *Injury Prevention*, **2**, 212-218.
11. Ogden, K.W. (1997) The effects of paved shoulders on accidents on rural highways, *Analysis and Prevention*, **29**, 353-362.
12. Ossenbruggen, P.J. (in press) The impacts of a safety compliance in highway design, *Risk: Health, Safety & Environment*, **10**.
13. Shankar, V., Mannering F. and Barfield, W. (1994) Effect of roadway geometrics and environmental factors on rural freeway accident frequencies, *Accident Analysis and Prevention*, **27**, 371-389.
14. Tidwell, J.E. and Doyle, D.P. (1995) Driver and pedestrian comprehension of pedestrian law and traffic control devices, *Transportation Research Record*, **1502**, 119-128.

Table 1. Crash and Injury Counts at Selected Sites

Site	n_c	n_i
A	60	16
B	10	3
C	2	2

Table 2. ADT and the Five-year Risk Exposure

Site	ADT	e
A	17,253	31,486,725
B	13,371	24,402,075
C	4,506	8,223,450

Table 3. Individual Lifetime Risk of θ^* = 1 in 1,000.

Site	R^*	n_i
A	1.2	16
B	0.9	3
C	0.3	2

Table 4. Crash and Injury Crash Proportions at Sample Sites

Site	p_c	p_i
A	19.06×10^{-7}	5.09×10^{-7}
B	4.09×10^{-7}	1.23×10^{-7}
C	2.43×10^{-7}	2.43×10^{-7}

Table 5. Conditional Injury Probabilities at Sample Crash Sites

Site	$p_{i\|c}$
A	0.267
B	0.300
C	1.000

Table 6. Crash Counts by Land Use Activity

	Number	Five Year Crash Total	Crash Rate*	
Village	13	70	1.08	(0.79)
Residential	45	381	1.69	(2.00)
Shopping	24	441	3.68	(3.89)
Totals	82	892		

*crashes per year. Standard deviations are given in parentheses.

Table 7. Percentage of Sites with Pedestrian and Traffic Control Amenities

Amenities: Sites	Village	Residential	Shopping	All
Crosswalks	70	4	21	18
Sidewalks	61	7	29	22
Traffic control	61	22	54	38
On-street parking	77	16	33	30
Crosswalks and sidewalks	61	4	21	18
Crosswalks, sidewalks and traffic control	46	4	21	18

Table 8. Traffic Volumes on Main and Intersecting Roads in ADT (vpd)*

	Main Road		Intersecting Road	
Village	9,632	(7,011)	3,754	(5,396)
Residential	9,814	(6,043)	1,604	(1,979)
Shopping	13,033	(4,851)	4,291	(4,742)

*vpd = vehicle per day. Standard deviations are given in parentheses.

Table 9. Paved Shoulder Width (in feet)

	Average	Standard. Deviation.
Village	2.1	3.4
Residential	3.2	3.0
Shopping	5.6	3.3

Table 10. Description of Logistic Regression Model Variables

Symbol	Description	
	Land Use Activity Variables:	
R	Residential	$\{R = 1$ when a residential site, $R = S = 0\}$
V	Village	$\{V = 1$ when a village site, $R = S = 0\}$
S	Shopping	$\{S = 1$ when a shopping site, $R = V = 0\}$
	Roadside Design Variables:	
C	Crosswalk	$\{C = 1$ when present, otherwise $C = 0\}$
W	Sidewalk	$\{W = 1$ when present, otherwise $W = 0\}$
	Traffic Control Variable:	
M	Traffic Control Device	$\{M = 1$ when present, otherwise $M = 0\}$
	Surrogate Speed Variables:	
P	On-street Parking	$\{P = 1$ when observed, otherwise $P = 0\}$
S_w	Shoulder width, measured in feet.	
	Crash Type Variable:	
T	Crash Type	$\{T = 0$ if single-vehicle crash, $T = 1$ if multi-vehicle crash$\}$
	Merging and Crossing Traffic Variables:	
F_m	Traffic flow on main road measured in ADT.	
F_i	Traffic flow on intersecting road in ADT.	

Table 11. Logistic Regression Modeling Results

Description	Symbol	Coefficient	Estimate	Standard. Deviation.	p level
Crash Model:					
Constant		a_0	-15.79	0.1982	_ 0.0001
Crash Type	T	a_1	0.7122	0.2056	0.0007
Traffic Device	M	a_2	0.6480	0.1833	0.0006
Sidewalk	W	a_3	-0.7080	0.2461	0.0047
Interaction Term	R_F_i	a_4	0.0000897	0.0000428	0.0380
Interaction Term	$S\_T\_F_i$	a_5	0.0000837	0.0000201	_ 0.0001
Injury Crash Model:					
Crash Type	T	b_1	-0.5632	0.1066	0.0007
Traffic Flow	F_i	b_2	-0.000043	0.0000184	0.0208

Table 12. Linear Predictor Functions by Typical Site Class

Crash Models:	M	W	u_c
Village	1	1	$-15.85 + 0.7122\ \_T$
Residential	0	0	$-15.79 + 0.7122\ \_T + 0.0000897\ \_F_i$
Shopping	1	0	$-15.14 + 0.7122\ \_T + 0.0000837\ \_F_i$
Injury Crash Model:			$u_{i\vert c}$
$-0.5632\ \_T - 0.000043\ \_F_i$			

Table 13. Roadway Risk by Crash Type

	F_i	T	$p(T)$	$\varpi_c(T,F_i)$	$\varpi_{i\vert c}(T,F_i)$
Village	3,754	0	0.23	$1.31\text{x}10^{-7}$	0.460
		1	0.77	$2.66\text{x}10^{-7}$	0.326
Residential	1,604	0	0.31	$1.60\text{x}10^{-7}$	0.483
		1	0.69	$3.27\text{x}10^{-7}$	0.347
Shopping	4,291	0	0.17	$2.65\text{x}10^{-7}$	0.454
		1	0.83	$7.75\text{x}10^{-7}$	0.321

Table 14. Roadway Risks by Typical Site Type

		Crash:		Injury Cases:		
	F_m	ϖ_c R^*	N_c	ϖ_i	N_i	
Village	9,632	$2.35\text{x}10^{-7}$	4.13	$0.81\text{x}10^{-7}$	1.41	0.69
Residential	9,814	$2.75\text{x}10^{-7}$	4.93	$1.02\text{x}10^{-7}$	1.83	0.70
Shopping	13,033	$6.88\text{x}10^{-7}$	16.37	$2.26\text{x}10^{-7}$	5.39	0.93

Table 15. Adjusted Odds Ratios

<u>Factor</u>	Ω
Crash Model:	
T	2.04
M	1.91
W	0.49
R_F_i with $R = 1$	1.15 for F_i = 1,604 vpd, Residential
$S\_T\_F_i$ with $S = T = 1$	1.43 for F_i = 4,291 vpd, Shopping
Injury Crash Model:	
T	0.57
F_i	0.93 for F_i = 1,604 vpd, Residential
	0.85 for F_i = 3.754 vpd, Village
	0.83 for F_i = 4,291 vpd, Shopping