Paper No. 01-2374

Estimating Pedestrian Exposure Prediction Model in Rural Areas

Xiao Qin and John N. Ivan

"Pedestrian exposure" is defined as the exposure risk of pedestrians to collisions with motor vehicles. It is one of the important factors influencing pedestrian crashes. Because pedestrian exposure or even pedestrian volume counts are not readily available, population density is usually used as a substitute in pedestrian crash prediction models. Unfortunately, population density is not a good replacement for pedestrian exposure because it does not account for the amount of walking people do. This study investigates the relationship between the weekly pedestrian exposure in rural areas of Connecticut and factors such as population density, presence of sidewalks, number of lanes, area type, traffic control type, and median household income. General linear modeling and Tukey and Duncan multiple comparison of means methods are used to identify the significant factors. Only the number of lanes, area type, and sidewalk system significantly explain the variation in the resulting pedestrian exposure prediction model. This study suggests extra improvement in pedestrian facilities for the areas with high pedestrian exposure. Ongoing research will take advantage of the model to estimate pedestrian crash models in rural areas of New England.

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

Since 1986, however, pedestrian fatalities have actually decreased. 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 the number of crashes per person in the population is considered; population increased from 237,626,036 in 1986 to 265,283,783 in 1996 (*3*). However, does that 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. If we want to learn what factors improve or worsen pedestrian safety, is more accurate to assess pedestrian safety by how many people actually walk on the streets.

Because the actual pedestrian time spent walking or pedestrian volumes are costly to observe and therefore 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, tourist sites frequently attract large numbers of visitors who are not counted in the population density. In other words, there is an unexpected high pedestrian volume that cannot be represented by the population density in these areas. In some areas the low pedestrian volume compared with the high population density may be attributed to the high vehicle-owner rate. Prediction models based on population density are, therefore, intrinsically 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 of the higher frequency of pedestrian crashes occurring there. However, there is little research studying the pedestrian activities 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 research findings on the effects that road features, neighborhood and land use, site characteristics, and demographic characteristics have on pedestrian activities in rural areas of Connecticut. It also sets forth the pedestrian exposure prediction model for determining or predicting the pedestrian travel patterns in rural areas.

LITERATURE REVIEW

In risk analysis "exposure" is a concept describing the opportunity for a random event to occur, that is, 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.

The choice of pedestrian exposure measure strongly influences the risk analysis results. Keall examined pedestrian crash data using the 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. In Keall's study, crashes per person overestimated the risk for people under 30 years old, underestimated the risk for people over 79 years old, 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*). They investigated site design characteristics, such as the mean block size, completeness and

University of Connecticut, Connecticut Transportation Institute, Civil and Environmental Engineering, Unit 2037, Storrs, CT 06269-2037.

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 being hazardous given socioeconomic and crime data (*9*). Roberts et al. [as reported by McMahon and colleagues (*10*)] noted a relationship between economic and ethnic differences with the pedestrian crash rate. Epperson recognized that the economic status of the neighborhood significantly influenced the predicted pedestrian crashes (*11*). McMahon and colleagues studied demographic variables such as the percentage of single parents with children, the percentage of housing stock built after 1980, whether 85 percent of households were composed of families, and whether the unemployment rate was less than 1.75 percent (*10*). The study showed that percentage of single parents with children and housing stock built after 1980 significantly influenced "walking along the road" crashes. The conclusion was that factors contributing to that type of crash included not only geometric characteristics of the sites but also demographics and neighborhood characteristics (*10*). 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 (*10*). However, there are several types of pedestrian activities, each bearing quite different risks of experiencing conflict with motor vehicles, for example, crossing a road, walking along a road, and walking on a sidewalk. Actually, 45 percent of all pedestrian crashes involve a pedestrian crossing a road, whereas only 14.1 percent involve a pedestrian walking along a road (*12*). Statistics suggest that crossing the street might be more dangerous than walking along the roadways or the crossing pedestrian exposure might be larger than the walking along the roadway pedestrian exposure.

Case-control methods to predict pedestrian crashes or exposure have been widely applied (*7, 13, 14*). For example, Hess selected sites by controlling the variables that former research considered had an effect on pedestrian volume, such as gross population density, land use, and income. The hypothesis is that when control variables were held constant, other factors such as the mean block size and completeness and continuity of the public sidewalk system affected pedestrian volume (*6*).

In our study several factors that may be important contributors to the prediction of pedestrian exposure—site characteristics, traffic control types, demographic data, land use characteristic, and road site features—are investigated using general linear regression model. The weekly crossing pedestrian volume is used as the measure of pedestrian exposure, the dependent variable in the prediction model.

METHODOLOGY AND DATA PROCESSING

Study Design

Variable Selection

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

- Pedestrian amenities: Sidewalk is used as an independent variable to represent the site characteristic because sidewalk is an important pedestrian-friendly design. In general, sites with sidewalks can attract more pedestrians than sites without sidewalks.
- Traffic control: Pedestrian-friendly traffic control designs such as marked crosswalk, traffic signal, and so on, are assumed to encourage people to cross at the location.
- Demographic data: Median household income is an important demographic characteristic expected to be associated with pedestrian exposure. Previous studies show that neighborhoods with high income usually have less pedestrian activity than neighborhoods with low income.
- 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 expected to be quite different from others.
- Road site features: Some road geometric features can influence pedestrian exposure—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 control categorical variable is defined with five values:

1. No marked crosswalk and no traffic signal (3 sites),
2. Traffic signal without marked crosswalk (2 sites),
3. Marked crosswalk without traffic signal (18 sites),
4. Marked crosswalk with yellow caution signal (2 sites), and
5. Marked crosswalk with traffic signal (7 sites).

According to the sites available, we define area type with seven values (*15*):

1. *Downtown* areas are characterized by larger buildings abutting one another and abutting sidewalks (10 sites).
2. *Compact residential* areas predominantly have houses close together and generally visible from the road, and often have sidewalks (1 site).
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 (5 sites).
4. *Village* areas consist of smaller buildings and residences set back from the road. Sidewalks may or may not be present (7 sites).
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 (3 sites).
6. *Tourist* areas usually include crosswalk and sidewalk without signal. Higher pedestrian exposure is expected, and pedestrians' activities may be constant throughout the day with less pronounced peaks during commuting and lunch time than at other areas (5 sites).
7. *Campus* areas usually include crosswalks, sidewalks without signal, and narrow streets and speed limit. Much higher pedestrian exposure is expected, and pedestrians' activities are greatly changed throughout the day with pronounced peaks during class and lunch or dinnertime (1 site).

Data Collection and Processing

Collection of Field Data

The data on pedestrian activity were collected 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 noninclement weather, thereby avoiding undesirable conditions. 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 taken mostly on Saturday. Sunday counts were conducted at sites where, based on the characteristics of the site, they were expected to yield the more significant results; a town center with little retail activity, but with community amenities such as a church is an example.

Because each site featured different characteristics, such as pedestrian crossing facilities or crosswalks, not all pedestrian activities were applicable to each site. However, with respect to this study, which is investigating alternative measures of pedestrian exposure to crashes, only the total number of crossing pedestrian exposures was of interest. Walking on the sidewalk was not included because it is assumed that these pedestrians are unlikely to be hit by an automobile. Generally, observations took place from 8:00 a.m. to 5:30 p.m.

Processing of Field Data

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. Table 1 shows the weekday and weekend pedestrian volume and the total weekly pedestrian volume computed using the following relationship:

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

where

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

Collection and Processing of Demographic Data

The next step was to gather the demographic characteristics of the vicinity of each site. The type of demographic data used in this study was 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 walking distance around the study site. The walking distance to the study site is defined as 1097 m (3,600 ft), which corresponds to 20 min of walking at a speed of 0.9 m (3 ft) per second. Thus the neighborhood within walking distance was defined as a polygon encompassing all areas that are within 1097 m (3,600 ft) when walking on streets.

Analysis Methodology

Many previous studies have found a nonlinear relationship between pedestrian exposure and the independent variables, and the value of pedestrian cannot be negative. Scatter plots of raw data showed that there might be a positive linear relationship between ln$V$ and ln$P$. Therefore, the general form for our prediction model was as follows:

TABLE 1 Number of Pedestrian Exposures for Each Site

Site Number	Town	Weekday Pedestrian Exposure (V_{wd})	Weekend Pedestrian Exposure (V_{we})	Total Weekly Pedestrian Exposure (V)
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

$$V = P^{\alpha} e^{(\beta_0 + X_S\beta_S + X_D\beta_D + X_L\beta_L + X_R\beta_R + \epsilon)} \quad (2)$$

where

P = computed population density in the walk area,
α = exponent on population density to be estimated,
X_S = site characteristics,
X_D = demographic characteristics,
X_L = land use characteristics,
X_R = road characteristics,
β_0, β_S, β_D, β_L, and β_R = parameters to be estimated, and
ϵ = error term.

After the natural log transformation, Equation 2 is turned into a simpler linear form as follows:

$$\ln V = \alpha \ln P + \beta_0 + X_S\beta_S + X_D\beta_D + X_L\beta_L + X_R\beta_R + \epsilon \quad (3)$$

Generalized linear regression, the statistical analysis system procedure, 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. That is, the interaction effects between different predictor variables or explanatory variables are neglected. If the null hypothesis that our assumptions about the variables are significant is 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

To identify the patterns between pairs of continuous variables in a model, we examined the relationship between the variables using a matrix plot. There are three continuous variables in our model: natural log of weekly pedestrian exposure ($\ln V$), natural log of population density in the walking area ($\ln P$), and median household income (M). The relationships between $\ln V$ versus $\ln P$ and $\ln V$ versus M are identified. If the plots are randomly distributed and lack linear patterns, the linear relationship between two variables is not obvious, such as the relationship between $\ln V$ and M. Otherwise the linear relationship between two variables is obvious, such as the relationship between $\ln V$ and $\ln P$. Despite the fact that the selected sites cover a wide income range from $16,822 to $60,953 per year, after the test of between-subject effects, the income factor shows a nonsignificant role in the model prediction and should be discarded. Because the income was 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 pedestrian exposure. Figure 1*a* shows the lack of a clear relationship between variable $\ln V$ and M and Figure 1*b* shows the possible linear relationship between $\ln V$ and $\ln P$.

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

$$\ln V = \alpha \ln P + \beta_0 + \beta_a X_a + \beta_s X_s + \beta_l X_l + \beta_w X_w \quad (4)$$

where

a = categorical variable area type,
s = categorical variable traffic control type,
l = categorical variable number of lanes, and
w = categorical variable sidewalk.

Table 2 lists all variables with brief definitions.

To test the significance of variables regression coefficients were tested for each variable using Type III sum of squares (SSIII). Also referred to as partial sum of squares, SSIII is considered by many to be the most accurate coefficient because the hypothesis for an effect does not involve parameters of other variables (*16*, *17*). Estimation results (shown in Table 3) as Model 1 reveal that natural log of population density in the walking area ($\ln P$) is nonsignificant; factors such as with or without sidewalk (X_w), number of lanes (X_l), area type (X_a), and traffic control type (X_s) are significant in predicting the response variable $\ln V$ at the 90 percent significance level. On the basis of previous studies that show population density has a significant effect on pedestrian exposure, population density is temporarily kept in the model to be tested further.

However, whether or not specific levels within each categorical variable are significantly different from each other 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 4 gives the results of main effect of levels in each categorical variable.

In results of the Tukey and Duncan method, levels indicated by the same letter code are not significantly different from each other. Table 4 shows that these differences for the sidewalk, number of lanes, and traffic control 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. There are no overlaps using the Duncan method, though. Because the Duncan method is more sensitive in identifying the difference between the levels, we use only the Duncan method to regroup the categorical variables. As shown in Table 4, the coefficients on tourist area

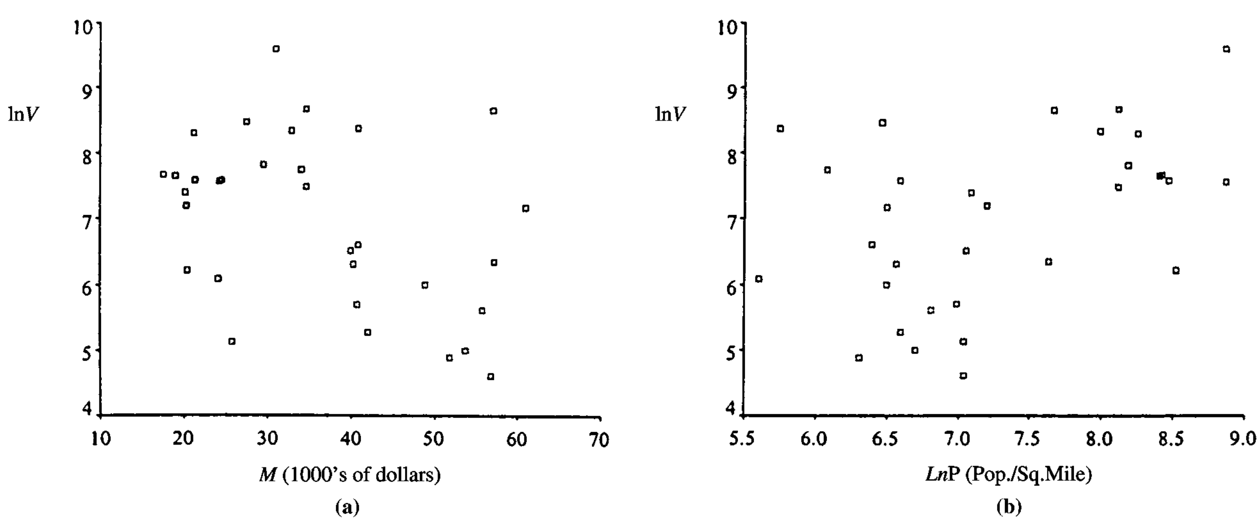

FIGURE 1 Relationship between continuous variables: (*a*) natural log of weekly pedestrian exposure ($\ln V$) and median income (M); (*b*) $\ln V$ and natural log of population density in the walking area ($\ln P$).

TABLE 2 Variable Information for 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
Xl	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

and downtown are not significantly different from each other in the 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 traffic control type from five to two. Table 5 shows the new categorical variables in Model 2.

Model 3 is a reduced model derived from Model 2 without the traffic control type variable because this effect is confounded with the others; that is, it is not independent, but somewhat correlated with other factors. As can be seen in Table 3, in Model 2 and Model 3, the negative coefficient for the population density (ln$P$) means high population density causes lower pedestrian exposure when other factors are controlled, which is opposite to the positive relationship between ln$V$ and ln$P$ shown in scatter plots of raw data in figure 1*b*. Therefore, population density in the walking area (ln$P$) 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.

Based on F-extra test between full Model 1 and Model 2 (as depicted in Table 6), the null hypothesis that the surplus variables in full model are nonsignificant at 95 percent level of significance cannot be rejected. That is, the evidence does not prove that the additional information in Model 1 helps to better predict pedestrian exposure. Compared with Model 1, the alternative Model 2 is much simpler,

TABLE 3 Temporal Factor Models for Weekly Pedestrian Exposure in Walk Area*

Variable	Model 1	Model 2	Model 3	Model 4
Intercept	3.88/3.09	5.59/4.93	5.37/4.76	3.78/6.89
ln$P$	0.005/0.03	-0.23/-1.57	-0.24/-1.60	—
With sidewalk	1.19/2.81	1.13/2.85	1.39/4.19	1.33/3.91
Without sidewalk	Base	Base	Base	Base
2-lane highway	2.06/3.27	1.86/3.78	1.86/3.75	1.87/3.67
4-lane highway	Base	Base	Base	Base
Campus	2.42/2.93	3.09/4.35	3.1/4.34	2.62/3.92
Tourist area	2.36/3.40	1.09/3.30	1.1/3.28	0.88/2.80
Downtown	1.48/2.22			
Village area	1.24/1.72	Base	Base	Base
Medium and low density commercial area	0.86/1.02			
Compact residential area	-0.83/-1.10			
Low density residential area	Base			
No crosswalk and no signal	-1.33/-2.24	-0.47/-1.19	—	—
Signal without crosswalk	-0.85/-1.12			
Crosswalk without signal	-1.28/-3.11	Base		
Crosswalk with yellow caution signal	-0.95/-1.73			
Crosswalk with signal	Base			
R square	0.909	0.832	0.823	0.805
RMSE	0.511	0.588	0.593	0.609

Coefficient (Estimate)/t-Value (Significance)
*** Significance at 90 percent**

TABLE 4 Multiple Means Comparison*

Categorical Variable	Value	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
X_a	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
X_s	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

*Levels indicated by the same letter code are not significantly different from each other.

and the predicted accuracy is good because of the close R^2 and root-mean-square error value of the two models. The result of *F*-partial test or *t*-test between Model 2 and 3 (see Table 6) suggests that the variable traffic control type is not an important contributor to predict pedestrian exposure. Finally, the small increases in residual deviance of Model 4 compared with Model 3 (see Table 3) is offset by the reduction in model complexity. Generally, the more parameters in a model, the more unstable it is; therefore, Model 4 is preferred in this case. Figure 2 is a plot of 90 percent confidence interval of the predicted pedestrian exposure versus actual pedestrian exposure for Model 4.

Model 4 with the parameter values inserted may be written as follows:

$$\ln V = 3.78 + 1.33X_W + 1.87X_L + 2.62X_C + 0.88X_{TD} \tag{5}$$

where

X_W = sidewalk ($X_W = 1$, with sidewalk; $X_W = 0$, without sidewalk),
X_L = number of lanes ($X_L = 1$, two-lane highway; $X_L = 0$, four-lane highway),
X_C = campus factor ($X_C = 1$, campus; $X_C = 0$, others), and
X_{TD} = tourist and downtown factor ($X_{TD} = 1$, tourist and downtown areas; $X_{TD} = 0$, others).

After the transformation, the equation is as follows:

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

TABLE 5 Variable Information for Model 2

Categorical Variable	Levels Number	Description
X_W	1	With sidewalk
	0	Without sidewalk
X_l	1	2-lane highway
	0	4-lane highway
X_a	1	Campus
	2	Tourist area and Downtown area
	3	Others
X_s	1	No crosswalk
	2	Crosswalk

Table 7 shows the predicted values, based on the three variables in Model 4, for each combination of predictor variables. Because of the limited number of sites and actual characteristics, some cell combinations are not observed, such as sites in tourist or downtown areas without sidewalks. We predicted values for these site types anyway for comparison purposes. It can be seen that the predicted values are very close to the observed values.

CONCLUSIONS

Pedestrian exposure is an important variable in the prediction of pedestrian crashes because it represents pedestrians' risk of being struck by vehicles (*18*). There are many factors that might influence pedestrian travel patterns and pedestrian volumes such as population density, demographic characteristics, site characteristics, land use, and highway geometric characteristics (*19*).

In this study, on the one hand, the effects of several factors do not conform to our expectation. For example, population density is not significant. It shows that pedestrian safety analyses based on population density may distort the true risk values. Second, control is still nonsignificant despite the fact that according to our observation when counting, most pedestrians appear to use a crosswalk or wait for the signal when crossing the street. The reason for the nonsignificance of the control variable may be that the control type effect is confounded with the others; that is, it is not independent, but somewhat correlated with other factors. Actually, it can be regarded as a response to high pedestrian volumes, rather than a contributor to high pedestrian volumes.

Furthermore, it is interesting to find that the median household income is not significant either. The result is different from many previous studies (*9–11*), which are usually done under an urban setting or urban, suburb, and rural mixture conditions. The discrepancy might be due to the limited number of sites or the homogenous data resource because all of our data are from rural areas of Connecticut. It is obvious that the vehicle-owner ratio is higher in rural areas, and variations for the demographic factors such as neighborhood environment, household median income, and unemployment are not as significant as they are in urban areas. Thus, the study suggested the necessity of considering an urban setting and rural setting separately.

TABLE 6 Model Selection

Model Selection	F-Value	F Critical Value	Conclusion
Model 1~Model 2	2.14	F(7,18 | 95%)=2.58	F-value<F-critical value
			Model 2 is better
Model 2~Model 3	1.42	F(1,25 | 95%)=4.24	F-value<F-critical value and *Xs* is discarded
			Model 3 is better
Model 3~Model 4	2.56	F(1,26 | 95%)=4.23	F-value<F-critical value and ln*P* is discarded
			Model 4 is better

95 percent significance level is used in the model selection.

On the other hand, some factors have expected effects. For example, sidewalk has a positive effect. The provision of a sidewalk system apparently encourages people to walk for trips in the area. Two-lane highway attracts more people to cross than four-lane highway, probably because it is less threatening and poses less risk for pedestrians to crash with vehicles. This finding also shows that area type is a significant contributor to predicting pedestrian volume. Campus area type has the greatest positive effect on pedestrian exposure. Next to that are tourist and downtown areas. These areas, therefore, deserve additional consideration for improvements in pedestrian facilities, such as warning devices, speed limit, stop sign, marked crosswalk, and so on. Furthermore, the model presents an effective way to categorize area type into three categories rather than the original seven types, which may be difficult to differentiate.

The model provided us with a simple way to predict the pedestrian exposure in rural areas using variables that are readily available, such as sidewalk system, number of lanes on the highway, and area type. Therefore, time-consuming and expensive manual pedestrian counts and unreliable substitute population density can be replaced by a predicted value. In addition, it is expected that a greater number of either pedestrians or vehicles would increase the likelihood of a site being dangerous, so this prediction model can be used to identify potentially hazardous locations for pedestrians. It is helpful for further research and analysis on the pedestrian crash rate (*20, 21*).

Because our study is limited to pedestrian crossing volumes in rural areas of Connecticut, new models should be estimated before applying the results to other regions. The limited number of sites did not cover all site feature combinations indicated in Table 7. Hence, future research could include collecting counts at sites with those combinations to test the predicted pedestrian exposure, using the same process and testing the significance of all variables if applied

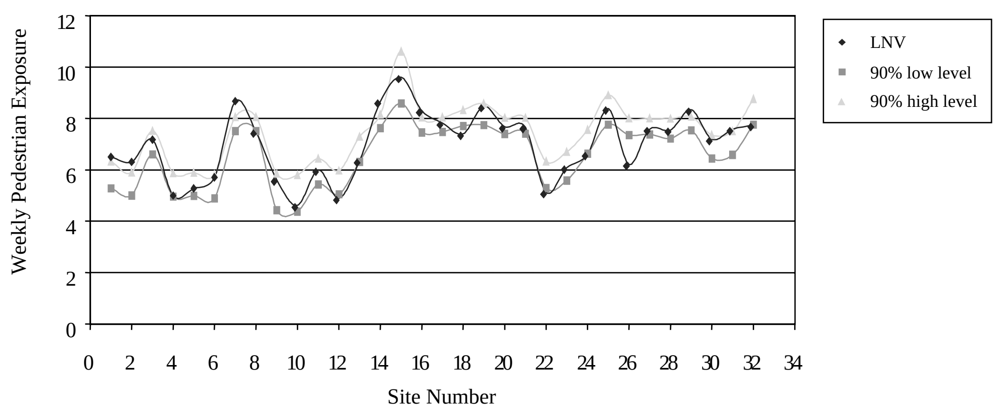

FIGURE 2 Actual ln*V* versus the range of predicted ln*V* (90 percent) in Model 4.

TABLE 7 Predicted Weekly Pedestrian Exposure Values

Factor 1	Factor 2	Value	Campus	Tourist/Downtown	Other
Sidewalk	Two-lane highway	Predicted Value	14,765	2,592	1,075
		Observed Average	14,724	3,047	1,178
		Sample Size	1	15	5
	Four-lane highway	Predicted Value	2,276	399	166
		Observed Average	—	—	187
		Sample Size	0	0	2
Without sidewalk	Two-lane highway	Predicted Value	3,905	685	285
		Observed Average	—	—	255
		Sample Size	0	0	9
	Four-lane highway	Predicted Value	602	106	44
		Observed Average	—	—	—
		Sample Size	0	0	0

in an urban setting, and collecting pedestrian information from other states to learn how these effects differ geographically. The predicted pedestrian exposure to walking trips is prepared for the facilities studied to be used for analyzing pedestrian fatality and injury rates in our future research.

ACKNOWLEDGMENTS

This research was funded by the U.S. Department of Transportation through the New England University Transportation Center and was performed at the Connecticut Transportation Institute. The authors also gratefully acknowledge contributions to this project made by Chunyan Wang, Tobias Kuhnimhof, and Sylvia Zajac.

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. Shapiro, R. J., J. F. Holmes, N. M. Gordon, J. F. Long, and D. H. Weinberg. *1997 Population Profile of the United States.* U.S. Department of Commerce, Economic and Statistics Administration and Bureau of the Census. Sept. 1998. http://www.census.gov/prod/3/98pubs/p23-194.pdf. Accessed May 2000.
4. Hauer, E. *Traffic Conflicts and Exposure.* Presented at the International Symposium on Risk-Exposure Measurement, Aarhus, Denmark, June 1980.
5. Keall, M. 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, P. M., A. V. Moudon, M. C. Snyder, and K. Stanilov. Site Design and Pedestrian Travel. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1674,* TRB, National Research Council, Washington, D.C., 1999, pp. 9–19.
7. Shriver, K. 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, 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. McMahon, P. J., C. Duncan, J. R. Stewart, C. V. Zegeer, and A. J. Khattak. Analysis of Factors Contributing to "Walking Along Roadway" Crashes. In *Transportation Record: Journal of the Transportation Research Board, No. 1674,* TRB, National Research Council, Washington, D.C., 1999, pp. 41–48.
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. North Carolina Department of Transportation. *Bicycle and Pedestrian Safety Study.* December 1997.
13. Moudon, A. V., P. M. Hess, M. C. Snyder, and K. Stanilov. Effect 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.
14. Carlin, J. B., P. 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.
15. Zajac, S. S. *Factors Influencing Injury Severity of Motor Vehicle-Crossing Pedestrian Crashes in Rural Connecticut.* M.S. thesis. Civil and Environmental Engineering, University of Connecticut, 2000.
16. *SAS/STAT User's Guide Volume 2, GLM-VARCOMP,* vers. 6, 4th ed. SAS Institute, Inc., Cary, N.C., 1990.
17. Draper, N. R., and H. Smith. *Applied Regression Analysis,* 3rd ed. Wiley-Interscience Publication. John Wiley & Sons, Inc., 1998.
18. Hauer, E. On Exposure and Accident Rate. *Traffic Engineering and Control,* Vol. 36, No. 3, 1995, pp. 134–138.
19. Garber, N. J., and T. K. Lienau. *Traffic and Highway Geometric Characteristics Associated with Pedestrian Crashes in Virginia.* Report No. VTRC 96-R29. Virginia Transportation Research Council, 1996.
20. Jensen, S. U. Pedestrian Safety in Denmark. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1674,* TRB, National Research Council, Washington, D.C., 1999, pp. 61–69.
21. Weatherall, R. What We Need to Know About Walking. *Traffic Engineering and Control,* Vol. 38, No. 7/8, 1997, pp. 385–387.

Publication of this paper sponsored by the Committee on Pedestrians.