Analysis of Factors Contributing to "Walking Along Roadway" Crashes

PATRICK J. MCMAHON, CHANDLER DUNCAN, J. RICHARD STEWART, CHARLES V. ZEGEER, AND ASAD J. KHATTAK

There are a variety of factors widely acknowledged to have an impact on the risk of pedestrian motor-vehicle injuries. The factors that have been most extensively researched are geometric characteristics of the road, including the presence of sidewalks. In relevant epidemiological research, however, factors relating to demographics and neighborhood characteristics have been alluded to but not sufficiently researched. A case-control methodology has been used and conditional and binary logistic models have been applied to determine the effects of cross-sectional roadway design attributes and socioeconomic and other census block-group data on the likelihood that a site is a crash site. Analyzed were 47 crash sites and 94 comparison sites. Physical design factors found to be associated with a significantly higher likelihood of being a crash site are a higher speed limit; the lack of wide, grassy walkable areas; and the absence of sidewalks. When these roadway factors are controlled for, nongeometric factors associated with a significantly higher likelihood of being a crash site are high levels of unemployment, older housing stock, lower proportions of families within households, and more single parents. This information suggests that some neighborhoods may, due to increased exposure or specific types of exposure, be especially appropriate sites for pedestrian safety measures such as sidewalks, lower-speed roadway designs, and the addition of wide, grassy shoulders.

The goal of a transportation system is to provide safe and efficient mobility and access to a wide variety of travelers with diverse needs and for different modes of travel. Walking is the most basic form of transportation, and it is important for transportation officials to provide facilities that enhance safe movement for pedestrians along roads and streets. An individual's transportation needs, and his or her ability to achieve them, are likely to vary not only from the physical roadway environment but also from socioeconomic situations and the proximity of potential attractors. Neighborhoods have their own specific patterns of transportation, and travelers within those neighborhoods may be subject to different risks than are encountered in other areas.

The purpose of this paper is to identify the types of, and reasons for, risks to pedestrians who are walking along a roadway. Factors examined in this study include both roadway factors and neighborhood factors. The sampling methodology matches crash sites where pedestrians were struck walking along a roadway to comparison sites of similar zoning, parcel size, and development amount, much like the matching done in a case-control study. Such roadway factors as vehicle volume, pedestrian volume, presence of a sidewalk, shoulder width, and type of roadside are included in the analysis. Census data from the U.S. Census block group in which each site was located were attributed to that crash site in order to analyze the impact of socioeconomic and other neighborhood factors (e.g., unemployment level, age of housing, and number of parents in the household).

P. J. McMahon and A. J. Khattak, University of North Carolina–Department of City and Regional Planning, New East Building, Campus Box 3140, Chapel Hill, NC 27599-3140. C. Duncan, BRW Inc., Thresher Square, 700 Third Street South, Minneapolis, MN 55415. J. R. Stewart and C. V. Zegeer, University of North Carolina–Highway Safety Research Center, 730 Airport Road, Bolin Creek Center, Campus Box 3430, Chapel Hill, NC 27599-3430.

LITERATURE REVIEW

Hunter et al. found that "walking along roadway" crashes constituted 7.9 percent (400 of 5,073) of all pedestrian-motor vehicle crashes in a sample from six states (*1*). Of the 400 crashes, they found that 69 percent involved pedestrians walking with traffic and 21 percent involved pedestrians walking against traffic. Overrepresented variables of those 400 walking-along-roadway crashes included the following:

- Pedestrian ages 15 to 44 years;
- Alcohol involvement by the pedestrian and driver;
- Rural, two-lane roads;
- Dark conditions with no lights; and
- Interstate and county roads.

A number of studies have applied case-control methods to questions of pedestrian safety. For example, Carlin et al. used a case-control method to examine effects of the behavior and attributes of children in determining their propensity to be involved in bicycle crashes (*2*). One of the findings of this study was a disproportionate number of injuries among children in the lowest income category.

A limited number of studies have investigated roadway factors associated with walking-along-road pedestrian crashes. In a 1996 study by Knoblauch et al. involving an analysis of pedestrian collisions and exposure under various roadway situations, locations with no sidewalks were more than twice as likely to have pedestrian crashes than sites where sidewalks existed. The presence of a sidewalk was found to have a particularly large safety benefit in residential and mixed residential areas. Sidewalks, however, had no effect on pedestrian crash experiences in commercial areas (*3*).

A wide variety of studies have acknowledged the increased risk of some socioeconomic groups to pedestrian injuries, with research focused on children and youths. These studies generally have shown that children of minorities and low-income families tend to be disproportionately represented in groups especially prone to pedestrian-motor vehicle crashes and injuries. For example, Roberts recognized a connection in economic and ethnic differences, concentrating on the increased "exposure to risk" of the children of single parents, often in lower income brackets, with regards to supervision of playing near roads and walking to school (*4,5*). King et al. found that neighborhoods of high pedestrian injuries in the West Midlands, United Kingdom, had high proportions of immigrant heads of households (*6*). Epperson noted that the economic status of neighborhood residents plays an important role in the prediction of areas with high bicycle-crash rates (*7*). He theorized that this was due to an increased dependence on bicycles as a mode of transportation. It may be concluded from these studies that the overrepresentation of crash risk among some lower-income people and minorities is likely the result of different travel patterns, proximity to dangerous streets, less supervision of children at play near streets, and the lower percentage of large, fenced-in yards, compared with higher-income neighborhoods, etc.

Some of these researchers have used methods similar to those in this paper. Christie used logistic regression to determine the socioeconomic and environmental factors that increased the likelihood of individual children to have been involved in a crash (*8*). Bagley, on the other hand, analyzed the likelihood of neighborhoods to be sites of crashes given socioeconomic and crime data (*9*). He found "general" crime, the percentage of housing with subsidies, population density, percentage unemployed, and low birth weights to be correlated with pedestrian crashes, and the percentage of park space to be negatively correlated with crashes (i.e., more park space correlates with lower risk for pedestrian crashes) (*9*). Finally, Braddock et al. used the geographic information system (GIS) to map out the pedestrian-motor vehicle collisions involving youths and found a nonuniform distribution but did not appear to have looked at the socioeconomic characteristics of the neighborhoods (*10*).

OBJECTIVE AND SCOPE

The first objective of this paper was to identify roadway design factors and neighborhood socioeconomic factors that distinguish walking-along-roadway crash sites from other matched sites in the same neighborhood and also in faraway neighborhoods. The second objective was to suggest measures that are likely to reduce the occurrence of such crashes.

Wake County, North Carolina, was selected as the study area because it contains a mix of urban, suburban, and rural conditions, and four years of crash data were easily available for research purposes. Sites could be visited and geometric data obtained for the crash sites and faraway comparison sites. Finally, the number of sites—141 including the crash sites and both nearby and faraway comparison sites—was adequate to conduct multivariate analysis with minimal difficulties.

The 47 walking-along-roadway crashes in this four-year sample (1993–1996) constituted 6.61 percent of the 711 pedestrian-motor vehicle crashes in Wake County during the same period. Based on the crash reports, it was found that roughly 77 percent of the collisions involved pedestrians walking with traffic and 23 percent involved pedestrians walking against traffic. Small proportions of the pedestrians walking with traffic were hit by oncoming traffic and small proportions of the pedestrians walking against traffic were hit from behind. These trends compare closely with the 1996 study by Hunter et al. (*1*), as discussed earlier.

METHODOLOGY

This paper uses a case-control method often used in epidemiology. The study analyzes variation between sites, not individuals. For this reason, the case sites were matched with nearby (same neighborhood) and faraway (other side of town) comparison sites. The comparison sites were chosen on the basis of current land use. A tax assessors' map from the Wake County Planning Department was used to match crash sites to nearby and more distant comparison sites.

Crash sites were pinpointed on the color-coded zoning map. Then, for each case site, comparison areas of the county with the same type of zoning covering about the same amount of area with the same-sized parcels as the crash site—but with no crashes—were identified. From this suitable comparative area of the county, a segment of roadway with the same functional class and number of lanes as the crash segment was chosen without regard to vehicle volume, sidewalk presence, or any other variables for which data were collected. The entire process was done prior to visiting the sites. Comparison site selection was done in this manner to avoid bias in the selection.

The issue of exposure was addressed in the analysis by including vehicle and pedestrian counts as independent variables in the models. Hence, the analysis controls for exposure without matching sites by traffic counts. Vehicle volume amounts listed in this paper refer to the number of vehicles counted in the outside lanes at the sites (i.e., the lanes closest to pedestrians walking along the roadway) over the course of one hour. These volume measurements, like the other site measurements, were taken at approximately the same time of day that the crash occurred. However, no counts were taken after midnight or before 6:00 a.m.

Because of the low frequency of walking-along-roadway crashes, it would take many years of studying a very large number of links before enough crashes could be observed to use in some types of experimental designs (e.g., before/after with a cohort analysis). For this reason, this study begins with known crash sites, then identifies matched nearby and faraway comparison sites. While a segment was defined as an unintersected stretch of roadway, specific measurements of segment lengths were not taken; however, because crash and comparison segments were in areas with similar levels of development and on the same functional class of roadway, they were, in all cases, of comparable length. These sampling strategies are appropriate given the conceptual framework of the study and the policy questions being addressed.

This sampling method yields a set of three sites for each of the 47 crash sites (all of the walking-along-roadway crashes in Wake County from the year 1993 through 1996), consisting of (*a*) the crash site, (*b*) a nearby comparison site, and (*c*) a distant comparison site. Hence, observed was a total of 141 sites. Each crash site was the location of a single pedestrian-vehicle crash during the period studied (although two crashes involved two pedestrians each). Data collectors visited these matched sites generally during the same hour the crash occurred. For example, if a crash occurred at 6:30 a.m., then data collectors visited the crash and faraway-comparison sites from 6:00 a.m. to 7:00 a.m. on a given morning, collecting pedestrian and vehicular volumes and making detailed measurements of cross-sectional design attributes.

Two different types of modeling analyses were conducted to quantify roadway and neighborhood effects on walking-along-roadway pedestrian collisions. The first of these was a matched case-control analysis carried out using Statistical Analysis System (SAS) PROC PHREG and examining the geometric roadway data at all three types of sites (i.e., crash sites, nearby comparison sites, and faraway comparison sites), while the other used Statistical Package for the Social Sciences (SPSS) and a binary logistic regression and addressed both geometric and neighborhood demographic factors at the crash site and the distant comparison site.

SAS PROC PHREG is a procedure for survival analysis based on the Cox proportional hazards model. It turns out that an *m:n* matched conditional logistic model has a likelihood function that is a special case of a proportional hazards model. Details can be found in Stokes et al. (*11*). Another reference is SAS/STAT Software Changes and Enhancements through Release 6.11 (*12*). SPSS for Windows, Release 7.5.1 was used to estimate the parameters of the binary logistic model.

Several of these variables differ considerably between the crash sites and faraway comparison sites. On the other hand, both the demographic variables and the roadway characteristics would have been expected to differ only slightly between those at the crash sites and the noncrash nearby comparison sites—since the nearby comparison sites typically were selected on a nearby street crossing the street of the crash site.

It was expected that the nearby locations would have been exposed to much of the same population of drivers and pedestrians as the crash sites and, hence, to some extent control for these factors. Yet, there was concern that street design within the same neighborhood might be so uniform as to tend to control out the factors of interest (e.g., sidewalk present or absent). The more distant locations would not be expected to be exposed to the same distributions of motorists and pedestrians, but the matching might be expected to yield some degree of similarity. These sites also should add variability to the observed roadway designs. Therefore, the selection of both a nearby and faraway comparison site was considered desirable for each crash site.

ANALYSIS OF THE ROADWAY FACTORS USING CONDITIONAL LOGISTIC ANALYSIS

Among the data elements collected at each location, the following were the key variables used in the statistical analysis:

- Speed limit,
- Sidewalk (present or absent),
- Paved shoulder width,
- Gutter pan width,
- Pedestrian volume,
- Traffic volume in the outside lanes, and
- Unpaved walkable space.

A four-way tabulation is given in Table 1 with speed limit, site type, sidewalk presence, and presence or absence of a paved shoulder .7625 m (2.5 ft) in width or wider. The crash sites differed from the comparisons on a few of the roadway variables, namely, speed limit, sidewalk presence, and traffic volume. Table 1 shows that these differences primarily are between the crash sites and the more distant comparison sites. The distribution of observed traffic volumes for the three site types is presented in Table 2. Observed pedestrian volumes appeared to differ very little across site types with at least one pedestrian observed about 50 percent of the time at each of the three site types. The 90th percentiles for observed hourly pedestrian volumes were 10, 8, and 8 for crash, far, and near sites, respectively.

With respect to other characteristics of the data, paved shoulders (of any width) were present at 61.7 percent of the crash sites, 29.8 percent of the far comparison sites, and 57.4 percent of the near comparison sites. There were no sidewalks on either side of the street at 80.9 percent of the sites visited, with no sidewalks at 91.5 percent of the crash sites and 75.5 percent of the noncrash comparison sites. In general, the crash locations tended to be more rural, have higher speed limits, have higher traffic volumes, be more likely to have paved shoulders, and be less likely to have sidewalks when compared with the noncrash comparison sites.

Statistical analyses were carried out by including the variables listed above in a series of conditional logistic models. Conditional logistic analysis is a standard method of analysis for matched case-control studies (*12*). In this setting, each observation consists of a response variable indicating a case (crash location) or comparison and values of each of the independent variables. The procedure maximizes

TABLE 1 Frequencies of Speed Limit, Paved Sidewalk, and Wide Paved Shoulder by Site Type

Site Type	Paved Sidewalk	Paved Shoulder*	Speed Limit in kph (mph)						Total
			32 (20)	40 (25)	48 (30)	56 (35)	72 (45)	89 (55)	
Crash Site	No	No	1	2	0	11	14	8	36
	No	Yes	0	0	0	2	4	1	7
	Yes	No	0	0	0	4	0	0	4
	Yes	Yes	0	0	0	0	0	0	0
Far Comparison	No	No	2	13	0	5	8	1	29
	No	Yes	0	0	0	0	2	1	3
	Yes	No	0	5	1	6	3	0	15
	Yes	Yes	0	0	0	0	0	0	0
Near Comparison	No	No	1	6	0	11	15	3	36
	No	Yes	0	0	0	0	2	1	3
	Yes	No	0	2	0	4	0	0	6
	Yes	Yes	0	0	0	1	1	0	2
Total			4	28	1	44	49	15	141

*Paved shoulder with width ≥ 2.5 feet (.7625 meters).

TABLE 2 Motor-Vehicle Traffic Volume Distributions by Site Type (Vehicle volume hourly for the outside lanes at the hour the crash took place)

	Site Type		
Percentiles	Crash Site	Far Comparison	Near Comparison
10th	32	13	14
25th	85	22	26
50th	174	66	103
75th	502	241	285
90th	942	625	644

the likelihood that a case is correctly classified as a function of the independent variables (e.g., roadway variables), given the condition that exactly one of each matched triplet is a case. Table 3 contains results from PROC PHREG obtained by using a best-subset selection criterion for selecting best subsets of seven independent variables input to the procedure.

The results in Table 3 show speed limit clearly to be the dominant variable for discriminating between crash and comparison sites. Beyond this, the models are not very unique in the sense that different combinations of variables yield similar values of the score statistic, a goodness-of-fit measure having an approximate χ^2 distribution with degrees of freedom equal to the number of included variables. This is not surprising, given the correlated nature of the variables. Table 4 shows the parameter estimates and associated statistics for the best three-variable model, those being speed limit, presence of sidewalk, and traffic volume.

This model shows speed limit to be highly significant while the presence of sidewalks and traffic volume are significant at levels just below and just above the .05 level. While speed limits are not perfect measures of roadway speeds, it was believed that this adequately approximates the speeds of vehicles on these roadways for the purpose of this analysis. When more variables are added to the model, significance levels of some variables increase to about .10 or higher. Table 4 also contains a column headed "Risk Ratio."

For the variable "presence of sidewalk," the risk ratio is 0.118. This means that given the data at hand, when speed limit and traffic volume are taken into account, the likelihood of a site with a paved sidewalk being a crash site is 88.2 percent lower than a site without a sidewalk. It should not be interpreted to mean that installing sidewalks would necessarily reduce the likelihood of pedestrian crashes by 88.2 percent in all situations. The presence of a sidewalk, however, clearly has a strong beneficial effect of reducing the risk of a walking-along-roadway pedestrian collision with a motor vehicle.

Risk ratios for speed limit and traffic volume also are shown in Table 4. As expected, increases in traffic volume and speed limit are associated with the greater likelihood of being a crash site. The results of Table 4 also show that by increasing the traffic volume by 1 unit (e.g., from 300 to 301 vehicles per hour), the risk ratio is 1.002. This means that the probability that the site is a crash site increases by 0.2 percent. For an increase in vehicle volume of 100 (e.g., from 300 to 400), there would be an increase in the probability the site is a crash site by a factor of $(1.002)^{100} = 1.22$ (a 22 percent increase). The speed-limit risk ratio was 1.116. This means that each 1.6 km/h (1.0 mph) increase in speed limit increases the likelihood of a crash site by the factor 1.116. An 8 km/h (5 mph) increase in speed limit increases the likelihood by $1.116^5 = 1.73$, while a 16.1 km/h (10 mph) increase yields the factor $(1.116)^{10} = 3.0$.

With regards to other potential roadway variables, data were collected for on-street parking and for street lighting. None of the sites sampled contained bicycle lanes and data were not collected for sidewalk buffer strips. Given the sample size, the diversity of urban and rural sites, and the minimal number of nighttime crashes, on-street parking and street lighting did not prove to be significant in any of the final specifications of these models. This is not to suggest that these variables are not important factors in the incidence of walking-along-roadway crashes but that the sample here was not adequate to effectively analyze these factors.

ANALYSIS OF THE ROADWAY AND DEMOGRAPHIC FACTORS USING THE BINARY LOGISTIC MODEL

Following the collection of roadway attributes, information from each of these crash sites was manually entered into ArcView for GIS plotting purposes. The sites then were linked with the roadway characteristics mentioned above and then with the information from the 1990 U.S. Census block group within which they were located. As the nearby comparison sites were likely to have the same census data, and thus no variation in the neighborhood variables, this analysis used only the crash sites and the faraway comparison sites. This data

TABLE 3 PHREG Procedure: Best Subsets Selected by Score Criterion

No. of Var.	Score Value	Variables Included
1	15.39	Speed limit
1	11.11	Gutter pan width
2	19.18	Speed limit, gutter pan width
2	19.17	Speed limit, paved sidewalk
3	22.25	Speed limit, paved sidewalk, traffic volume
3	21.51	Speed limit, paved sidewalk, gutter pan width
4	23.59	Speed limit, traffic volume, gutter pan width
4	22.61	Speed limit, unpaved walkable space, gutter pan width

TABLE 4 Results for Three Variable Models

Variable	Coefficient (Estimate)	s.e.	χ^2	P-value	Risk Ratio	95% Confidence Intervals
Speed limit (m.p.h.)	0.1094	0.0381	8.22	0.0041	1.116	(1.035, 1.202)
Paved sidewalk	-2.1346	1.077	3.93	0.0474	0.118	(0.014, 0.976)
Traffic volume	0.0019	0.0010	3.69	0.0549	1.002	(1.000, 1.004)

s.e. = standard error

file then was loaded into SPSS, in which a binary logistic regression was run. The locations of crash sites, the noncrash faraway comparison locations, and the percentage of single-parent households by census block group are illustrated in Figure 1.

The methods used to compare the attributes of the sites include GIS maps, a binary logistic regression model, and a table showing the mean of each variable for the crash sites and the corresponding mean for the comparison sites. Since this analysis only looked at the crash and faraway comparison sites, the binary logistic model was chosen because it accounts for the alternate possibilities that a site will be either a crash (1) or a comparison (0) site. This model identifies the statistical significance and degree to which certain attributes distinguish a crash site from a comparison site and gives the marginal effects of each relationship.

The model specification included a variety of geometric factors that were statistically significant in the previously discussed analysis. Added to these factors were a number of variables based on census data that were believed to be potentially correlated with crash sites and also that approximate some ways in which specific neighborhoods may contain greater, or different, risks of walking-along-roadway crashes for pedestrians.

The average median household income in the block groups of crash sites is $31,653, while it is $41,279 at noncrash faraway comparison sites. Nearly 2.7 percent of the residents around crash sites take the bus to work and 2.7 percent walk. At the noncrash faraway comparison sites, less than 0.25 percent take the bus and 1.1 percent walk. Minorities constituted 39 percent of the block groups around crash sites but only 15 percent of noncrash faraway comparison sites, and just over 63 percent of the homes in crash neighborhoods were owner-occupied while noncrash faraway comparison sites were nearly 76 percent owner-occupied.

Through the binary logistic regression (see Table 5), several roadway design factors and neighborhood characteristics were found to be statistically significant. These results were very similar to the results of the conditional logistic modeling above and confirmed that reduced crash risk was associated with lower traffic volume, lower

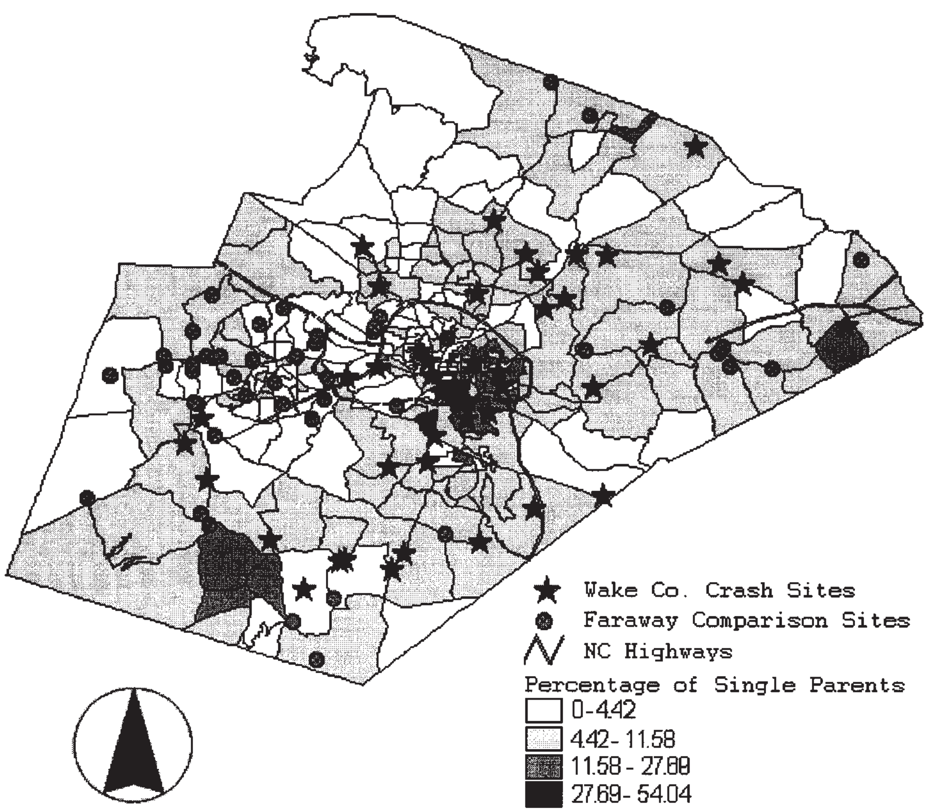

FIGURE 1 Wake County walking-along-roadway pedestrian crash locations (1993–1996), noncrash faraway comparison sites, and single-parent household percentages.

TABLE 5 Effects of Marginal Changes in Independent Variables in Binary Logistic Model

Variable	Mode or Mean	Coefficient (Estimate)	P-Value (Significance)	Probability Site is a Crash Site at Mean/Mode	Probability Site is a Crash Site with Marginal Change
(constant)		-8.8161	0.0002		
Paved Sidewalk Absent	1	1.9903	0.0572	0.4771	0.1109
Speed Limit (mph) Divided by 5	8	0.8135	0.0572	0.4771	0.6730
Paved Shoulder Greater than 2.5 feet	0	31.1738	0.0210	0.4771	1.0000
Paved Shoulder Dummy Variable Interacted with Speed Limits	0	-3.3395	0.0171	1.0000	0.9875
Grassy and Other Unpaved Shoulder Space	1	-2.3327	0.0062	0.4771	0.9039
Percentage of Single Parents with Children	7	0.2646	0.0040	0.4771	0.5431
Less than 30 Percent of Housing Stock Built After 1980	0	2.4171	0.0129	0.4771	0.9110
More than 85 Percent of Households Composed of Families	0	-2.1301	0.0416	0.4771	0.0978
Unemployment Less than 1.75 Percent	0	-1.9415	0.0322	0.4771	0.1158
Vehicle Volume	280	0.0025	0.0336	0.4771	0.4833
Pedestrian Volume	3	0.0022	0.9821	0.4771	0.4776

vehicle speeds, and having a sidewalk present. The binary logistic analysis also showed wide, unpaved shoulders to reduce the risk of walking-along-roadway crashes. These geometric factors have been controlled for, however. Census variables that were significantly correlated with crash sites were (*a*) the percentage of single parents with children and (*b*) less than 30 percent of housing stock being built post-1980. On the other hand, variables significantly correlated with noncrash faraway comparison sites were (*a*) a high percentage of families within households and (*b*) extremely low levels of unemployment (less than 1.75 percent).

When reading through the results of this binary logistic regression, and specifically when looking at the coefficients and marginal effects, it is important to keep in mind the fact that the data set used to produce this model consisted of sites in only one North Carolina county. Therefore, the numeric values associated with each variable should be understood to be relative and indicative of a clear pattern but not a definitive finding that necessarily can be generalized to all other jurisdictions. A brief discussion of some of the neighborhood factors follows.

Percentage of Single Parents

As mentioned in Roberts, single parents are less likely to have the ability, given the other demands on their time, to extensively monitor their children (*4,5*). Children in neighborhoods characterized by large numbers of this household type may likely take part in riskier behavior without the supervision of an adult. Thus, it was expected that the percentage of single parents might be positively correlated with the likelihood of a site to be a crash site.

The model supports this hypothesis, showing that an increase from 7 to 8 percent in the number of single-parent households results in a 13 percent increase in the likelihood of a site being a crash site. Complicating this finding, however, is a lack of a clear relationship between the age of pedestrians struck walking along the roadway and the percentage of single parents in the block group.

Percentage of Housing Stock Built After 1980

In this model, newer housing is believed to be a proxy for a number of possible characteristics of neighborhoods that may influence the likelihood of a walking-along-roadway crash. It is expected that newer neighborhoods are more likely to contain amenities such as better designed roads, large yards, and nearby parks. These factors might influence the driving patterns and use of the roadway by residents. Additionally, it is likely that the percentage of newer housing is indicative of higher property values and more recent development, both indicators of the socioeconomic status of a neighborhood. For these reasons, it is expected that areas with a small percentage of houses built after 1980 might be more highly correlated with crash sites.

This model supports the hypothesis that older neighborhoods are more likely to contain pedestrian crashes compared with newer neighborhoods. The marginal effects indicate that sites in neighborhoods with less than 30 percent of their housing built since 1980 were 90 percent more likely to be crash sites than areas with more than 30 percent new homes.

Percentage of Families

The U.S. Census defines a family as “a group of two or more people, one of whom is the householder, living together, who are related by birth, marriage, or adoption.” Families indicate the presence of groups of people who can rely on each other for a variety of different resources. The possibility that another family member will have a vehicle and provide transportation reduces the need for family members to walk as a form of transportation. Families also are likely to participate in activities together and provide supervision of other family members and children. Because of these types of behavior, it is expected that areas with a large proportion of families will be more strongly correlated with the noncrash faraway comparison sites than with crash sites.

The model supports the hypothesis that having more family households reduces the likelihood of a site being a crash site. The marginal effects show that areas with more than 85 percent of households composed of families were 79 percent less likely to be a crash site than areas with less than 85 percent families.

Unemployment

Areas with extremely low levels of unemployment also are those in which individuals could be assumed to place a high value on their time and personal safety. Additionally, employed individuals will likely have less free time to walk in the street and are financially more able to own a car and thus take a greater percent of their trips by car. For all of these and other possible reasons, it is expected that areas with very high employment would be less likely to be crash sites than areas with higher levels of unemployment.

The model supports this hypothesis and had a negative coefficient for the dummy variable testing for areas with less than 1.75 percent unemployment. The analysis found these sites to be 75 percent less likely to be a crash site when compared with neighborhoods with a greater level of unemployment.

Pedestrian and Vehicle Volumes

The socioeconomic variables discussed above measure exposure indirectly, whereas pedestrian and vehicle volumes directly measure exposure. It is expected that greater numbers of either pedestrians or vehicles would increase the likelihood of a site being a crash site, because greater numbers of pedestrians and vehicles are present for possible conflict. The noncrash faraway comparison sites were picked without considering pedestrian or vehicle volumes, however, so it is expected that the variations in pedestrian and vehicle volumes should not be substantial.

This model found that pedestrian volume, while higher at the crash sites, was not statistically significant in this particular sample of sites. This result does not mean that pedestrian exposure is not important. Instead, it is clearly the result of pedestrian volumes being relatively low (generally fewer than five pedestrians per hour) at most of the sites and fairly similar between crash sites and control sites. Vehicle volume, however, was significant and positively correlated with crash sites. The marginal effects show that the increase from 280 to 290 vehicles per hour, for example, increased the likelihood that a site was a crash site by 1.3 percent.

Grassy and Other Unpaved Shoulder Spaces

It was expected that very wide grassy areas and other unpaved shoulder spaces might be less likely to be crash sites, compared to sites with little or no shoulder. This is because even where sidewalks are absent, a wide, unpaved space on the shoulder provides a safe environment for people to walk off of the asphalt. Further, it may be less likely for a vehicle to strike a pedestrian on an unpaved shoulder than a paved shoulder, because the vehicle must run off the road causing noise and other disturbance that could alert the pedestrian or the driver to the problem prior to a collision.

Four feet is comparable with a narrow sidewalk and is a wide enough space that it is expected to provide the average pedestrian a place to walk off of the pavement. Segments with four feet or more of walkable unpaved space were expected to be less likely to be a crash site than those without it.

The model found that an unpaved shoulder space of four feet or more makes a site 89 percent less likely to be a crash site. This is consistent with the conceptual reasoning on this variable and suggests that such sites have adequate walking space and are less in need of sidewalks than sites with less than four feet of walkable unpaved space.

CONCLUSION

Roadway characteristics such as the absence of sidewalks, higher traffic volume, higher vehicle speed, and smaller width of unpaved shoulder increase the likelihood that a walking-along-roadway pedestrian crash will occur. By controlling for these factors, this study also found that neighborhood factors influence the likelihood of a site being a crash site. Specifically, neighborhood characteristics that increase the likelihood of a walking-along-roadway crash include high percentages of single parents, large amounts of older housing (i.e., housing built before 1980), few households composed of families, and high levels of unemployment. It is believed that neighborhood factors capture the extent and type of exposure (i.e., safe or unsafe walking behavior) that takes place in conjunction with these factors.

These findings suggest that certain types of neighborhoods contain factors that increase the risk of pedestrians being involved in walking-along-roadway crashes. Specifically, sidewalks appear to be the most appropriate treatment on neighborhood streets, while wide, unpaved shoulders may be more suitable in more rural areas. Further, neighborhoods with larger numbers of single parents, older housing stock, greater dependency on public transit, fewer families, and higher unemployment might deserve added consideration for improvements in pedestrian facilities. These would provide a safer place for pedestrians to walk than in the travel lane or on a paved shoulder by physically separating the individual further from the traffic while walking adjacent to the street or roadway.

As part of a larger study for the Federal Highway Administration, the results of this study are being used to develop improved national priorities and guidelines for the installation of sidewalks, walkways, and shoulders. Providing such facilities for pedestrians will not only reduce pedestrian crashes but also improve pedestrian access. Such facilities also encourage more walking, which improves health and longevity of those who walk regularly. The Institute of Transportation Engineers gives design guidelines for sidewalks in its 1998 report titled *Design and Safety of Pedestrian Facilities* (*13*).

It also should be mentioned that the analyses in this study were limited to all walking-along roadway crashes in Wake County, North Carolina, in a four-year period (1992–1996). While such a study certainly could be conducted in other areas of the country, the results here are consistent with findings of previous research on roadway and neighborhood issues in terms of factors that affect pedestrian crash experience.

Beyond the variables tested here, it is expected that other variables, such as alcohol and drug usage, crime rate, and other indications of risk taking, could be incorporated successfully into this type of model to further explain behaviors that increase the risk of exposure to this type of crash. Even more appropriate would be more qualitative research involving the interviewing of neighborhood residents and those familiar with a variety of neighborhoods to determine some of the root causes for the current levels of risk and what they believe to be locally appropriate countermeasures.

ACKNOWLEDGMENT

The research that was used to develop this paper was funded by the Federal Highway Administration in the study "Evaluation of Pedestrian Facilities." The study was conducted by the University of North Carolina at Chapel Hill's Highway Safety Research Center. The results and opinions contained in this study are those of the authors, who are responsible for the facts and accuracy of the information in this paper.

REFERENCES

1. Hunter, W., J. Stutts, W. Pein, and C. Cox. *Pedestrian and Bicycle Crash Types of the Early 1990's.* Publication FHWA-RD-95-163. FHWA, U.S. Department of Transportation, June 1996.
2. 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.
3. Knoblauch, R. L., B. H. Justin, S. A. Smith, and M. T. Pretrucho. *Investigation of Exposure-Based Pedestrian Accident Areas: Crosswalks, Sidewalks, Local Streets, and Major Arterials.* Report FHWA/RD-87-038. FHWA, U.S. Department of Transportation, Feb. 1987.
4. Roberts, I. Adult Accompaniment and the Risk of Pedestrian Injury on the School-Home Journey. *Injury Prevention,* Vol. 1, No. 4, 1995, pp. 242–244.
5. 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.
6. King, D. M., S. D. Lawson, S. Proctor, and K. Johal. Child Pedestrian Accidents in Inner Areas: Patterns and Treatment. *Proc., Traffic Management and Road Safety,* V. P291, PTRC Summer Annual Meeting, University of Bath, England, Sept. 1987, pp. 159–177.
7. 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.
8. Christie, N. The High Risk Child Pedestrian: Socio-Economic and Environmental Factors in Their Accidents. *TRL Project Report,* PR 117, 1995, 33 pp.
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. Braddock, M., et al. Using a Geographic Information System To Understand Child Pedestrian Injury. *American Journal of Public Health,* Vol. 84, No. 7, 1994, pp. 1158–1161.
11. Stokes, M. E., C. S. Davis, and G. G. Koch. *Categorical Data Analysis Using the SAS System.* SAS Institute, Inc., Cary, N.C., 1995.
12. *SAS/STAT Software Changes and Enhancements through Release 6.11.* SAS Institute, Inc., Cary, N.C., 1996.
13. Institute of Transportation Engineers. *Design and Safety of Pedestrian Facilities.* Committee TENC-5A-5, March 1998.

Publication of this paper sponsored by Committee on Pedestrians.