Random Parameter Model Used to Explain Effects of Built-Environment Characteristics on Pedestrian Crash Frequency

Satish Ukkusuri, Samiul Hasan, and H. M. Abdul Aziz

Pedestrian safety has been a major concern for megacities such as New York City. Although pedestrian fatalities show a downward trend, these fatalities constitute a high percentage of overall traffic fatalities in the city. Data from New York City were used to study the factors that influence the frequency of pedestrian crashes. Specifically, a random parameter, negative binomial model was developed for predicting pedestrian crash frequencies at the census tract level. This approach allows the incorporation of unobserved heterogeneity across the spatial zones in the modeling process. The influences of a comprehensive set of variables describing the sociodemographic and built-environment characteristics on pedestrian crashes are reported. Several parameters in the model were found to be random, which indicates their heterogeneous influence on the numbers of pedestrian crashes. Overall, these findings can help frame better policies to improve pedestrian safety.

Pedestrian crashes have been a major concern in urban areas. In New York City, pedestrian fatalities show a downward trend because of continuous efforts by public agencies. For instance, the city's overall pedestrian fatalities declined by 54%, from 366 to 168, between 1990 and 2006 (*1*). Although the number of pedestrian crashes has decreased, pedestrian fatalities in the city constitute a high percentage of overall traffic fatalities. According to the New York City Department of Transportation (DOT), the number of pedestrian fatalities as a percentage of total traffic fatalities has remained at roughly 50% since 1990 (*2*). In 2006, 52.7% of traffic crash fatalities in New York City were pedestrians, much higher than the rate for New York State or the United States (*3*). The pedestrian injury rate per 100,000 people in New York City (130.79) is 60% higher than that of New York State (79.61) and roughly 6.4 times as high as that of the United States (20.37) as a whole (*1*). These figures indicate the reasoning behind the persistent concerns within the public agencies towards pedestrian safety.

Understanding the factors that influence pedestrian crashes is an important step toward the improvement of pedestrian safety, because a thorough understanding can significantly help to develop more effective countermeasures. Previous studies therefore focused on determining the causes of the pedestrian crashes through various modeling techniques. Studies related to pedestrian crashes attempted primarily to model two different aspects: the frequency and the severity levels of crash occurrences. Most of these studies examined the effects of pedestrian and driver characteristics, vehicle characteristics and conditions, road infrastructure and traffic characteristics, community characteristics, and land use and physical environment on pedestrian crashes.

Despite significant research efforts related to pedestrian crashes (*4–13*), there remain unaddressed issues. One such issue is the presence of unobserved heterogeneity in the pedestrian crash data. Mostly, pedestrian crash data are collected with reference to a road or intersection and aggregated within a particular level of spatial aggregation (census tract level or zip code level) (*1*). A major problem with such data collection is that there might be unobserved variations from one spatial zone to the other (i.e., unobserved heterogeneity). Traditional count-data models, widely used to model crash frequency, restrict the same parameters for all the observations and thus ignore the issue of spatial unobserved heterogeneity across the observations. However, because of this unobserved heterogeneity, the estimated parameters from these models are expected to differ across the spatial zones (*14*). For example, the land use patterns vary across the census tracts and would affect the probability of crashes in a distinct way for a specific census tract. Correspondingly, if the parameters of a model are constrained to be fixed when they actually vary across observations, then the estimated parameters will be biased (*15*) and hence may lead to incorrect policy making. Therefore, the consideration of such heterogeneity by allowing the parameters to vary across observations has significant potential in pedestrian safety research. Models that allow the parameters to vary across the observations are called random parameters models. This approach is particularly suitable for modeling pedestrian crash data because it has a potential unobserved heterogeneity issue. This paper uses data from New York City to develop a random parameter model to determine the factors influencing the frequency of pedestrian crashes.

An econometric model for pedestrian crash frequency is developed with census tract level crash data. Specifically, a random parameter negative binomial model for predicting crash frequency is developed. This approach allows the incorporation of unobserved heterogeneity into the modeling process. An understanding of such unobserved heterogeneity across the spatial zones will provide better insight into the influences of various factors on pedestrian–vehicle crashes. Of these factors, the role of the built environment is highlighted. Studies have shown the evidence that countermeasures changing

School of Civil Engineering, Purdue University, 550 Stadium Mall Drive, West Lafayette, IN 47907. Corresponding author: S. Ukkusuri, sukkusur@purdue.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2237, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 98–106.
DOI: 10.3141/2237-11

the attributes of the built environment can significantly reduce the rate of pedestrian crashes (*16–25*). These studies mainly focus on before-and-after studies of specific countermeasures. However, to determine the relative influences of different built environment characteristics on pedestrian crash rates, a multivariate model is needed. Such a model would require a comprehensive data set that has detailed information on the characteristics of the study area. Thus, a data set that has been specifically developed by combining multiple data sources within New York City will be used.

This paper contributes to the pedestrian safety literature in two ways. First, it demonstrates the application of random parameter models for modeling pedestrian crash frequency, addressing the issue of unobserved heterogeneity present in the data. Second, it provides insights into, and a better understanding of, factors related to the built environment on pedestrian crash frequency by building a model for New York City. Such insights drawn from New York will be useful for other major cities having similar pedestrian activities.

LITERATURE REVIEW

Previous studies found a number of demographic factors that contribute to pedestrian crashes (*1*). For instance, pedestrians older than 65 years old are more likely to be struck by motor vehicles as a result of slower reaction times, and they are more likely to be injured in a crash as a result of their physical condition (*4*). In addition to slower walking speeds, the cognitive ability of older adults may slow their reaction time. Oxley et al. conducted experiments to examine how age affects road-crossing decisions when selecting appropriate gaps and found that age is associated with an increase in risky decisions to cross roads (*5*). Those in the oldest group (>75 years) most frequently made unsafe decisions (i.e., selected smaller gaps) when crossing streets, especially considering their slower walking speeds and slower reaction times to oncoming traffic. However, a study by Lobjois and Cavallo contradicts Oxley's study, finding no association between unsafe decisions and age (*6*). Instead, elderly participants made safe decisions by selecting larger gaps to compensate for their slower walking speed, thereby ensuring enough time to cross streets.

Alcohol and drug use by drivers or pedestrians increases the likelihood and severity of pedestrian crashes. Again, this can be attributed to the impaired response of individuals in traffic conflicts. Of concern, alcohol involvement in pedestrian–vehicle crashes appears to be related to certain population groups. Ryb et al. found that pedestrians with a blood alcohol content of more than 200 mg/dL, as well as socioeconomic characteristics such as "low income, low educational achievement, younger age," are linked to pedestrian–vehicle crash injuries (*7*). Several studies have found that pedestrian–vehicle crashes involving alcohol or drugs are especially high among Hispanics and Native Americans (*8*).

Pedestrian volume and traffic volumes are also associated with the frequency of pedestrian crashes. For example, a study of King County, Washington, that used crash data from between 1999 and 2004 found that an increase in annual average daily traffic (AADT) correlates with an increase in crashes (*9*). An increase in AADT leads to higher exposure, which increases the risk of a crash. Traffic speeds also increase the frequency and severity of pedestrian crashes (*10*). Higher speeds reduce the time available for drivers and pedestrians to react while increasing the distance and time needed for drivers to successfully brake, greatly increasing the likelihood of a collision. Higher speeds at impact increase the kinetic energy absorbed by the pedestrian, thus increasing the likelihood of serious injury and death. In addition to speeds and volume, the number of traffic lanes (three or more versus two) is a statistically significant predictor of collisions because the wider roadway may be related to higher vehicle speeds (*11*). Narrower streets or longer signal timing can prove safer for pedestrians, especially for the elderly, many of whom have difficulty crossing streets within the times allowed by signals (*12*). For population groups requiring special attention (e.g., elderly, children, or people with disabilities), a redesign or reconfiguration of pedestrian safety-related facilities needs to be considered. According to a study by Lee and Abdel-Aty, pedestrian infrastructure—including signalized crossings, median barriers, and islands—and wide sidewalks contribute to the safety of pedestrians (*13*).

Theoretically, congestion has two-way effects on pedestrian crashes. On one hand, congested roads tend to have more motorist–pedestrian conflicts, leading to a higher frequency of crashes. On the other hand, congested roads also slow traffic, resulting in lower levels of injury severity when collisions occur (*18*). This is particularly relevant for New York City, where the dense urban environment and chronic traffic congestion have tempered the overall severity of pedestrian–vehicle crashes. Additional infrastructure characteristics contributing to this effect include tight intersection geometry with no offset from intersection to crosswalk, the nearly universal presence of sidewalks, the relative absence of curb cuts, the nearly universal use of a 15-ft curb radius, a longstanding ban on right turns on red, and the use of pedestrian signal heads and crosswalks at all signalized intersections.

Several studies have examined the relationship between pedestrian–vehicle collisions and the social and physical characteristics of neighborhoods (*1*). Researchers have found a relationship between an area's socioeconomic composition and pedestrian crash frequency. For example, LaScala et al. examined environmental factors related to child pedestrian crashes in four communities in California, finding that higher rates of crashes are associated with higher youth population densities, higher unemployment, lower household income, and higher traffic flow (*25*). Although some studies have pointed to ethnicity and race as factors in pedestrian safety, socioeconomic conditions appear to be more relevant to a pedestrian's higher exposure to crashes with vehicles. In Arizona, for example, Hispanics were 60% more likely to be involved in pedestrian–vehicle crashes, and higher fatalities were observed in areas with high densities and low income (*26*). According to the study, Arizona Hispanics tend to have lower incomes, show lower vehicle ownership, live in high-density areas with high-traffic volumes, and walk more (*27*). The combination of these factors leads to Arizona Hispanics' much higher rate of pedestrian exposure to vehicles (*28*). The National Household Survey supports this reasoning: in 2001, walking was a primary means of transportation for 11.8% of Hispanics, compared with 8.6% of whites (*29*). It has been suggested that the issue may be less related to Hispanics as a group and more correlated with low socioeconomic status (*30*). However, the identification of the relationship between higher pedestrian–vehicle crashes and race or ethnicity may not be simple and differs by locale. In New York City, the 2000 census revealed that although about 10.5% of Hispanics walked to work (which is almost the same as the city average of 10.4%), about 12% of white and 13.2% of Asian residents walked to work.

Land use has been found to be associated with the frequency of pedestrian crashes; however, the findings are often contradictory (*1, 24*). The presence of retail and entertainment uses (especially liquor stores, bars, restaurants that serve alcohol, and retail stores) tends to predict higher levels of vehicle–pedestrian collision rates

(*25*). Dense urban areas with high levels of pedestrian activity also tend to have higher frequencies of pedestrian–vehicle collisions (*31*). However, high-density pedestrian activity areas can also congest traffic flow, reduce traffic speeds, and therefore reduce the severity of the pedestrian–vehicle crashes (*18*). Schools, commercial facilities, and multifamily housing are frequently found to be linked to high collision frequency (*32*). Together with demographic data, land use and physical environment information are proxies that can control for the level of pedestrian activity in a neighborhood. Many of these conditions are relevant in New York City and may interact with infrastructure characteristics. Retail streets in the city are likely to be truck routes and bus routes with metered parking, resulting in large vehicles, high pedestrian volumes, and many parking maneuvers, which may increase either the severity or frequency of pedestrian crashes. Public housing and large, private multifamily buildings are likely to be located near large arterial streets, in part because of zoning rules. Hence, a modeling strategy based on isolating and then regrouping these often interactive factors is needed to identify the primary determinants of pedestrian crash frequency in New York City.

METHODOLOGY

Traditionally, crash frequency analyses are done by using count-data modeling techniques because the number of crashes specific to a zone, intersection, or roadway segment per unit time is a nonnegative integer. A comprehensive recent review on the count-data modeling methodologies and the related issues on crash frequency data is available elsewhere (*14*). For frequency analysis, because the dependent variables are nonnegative integers, researchers typically use a Poisson regression model or its derivatives, including the negative binomial and zero-inflated models. According to the Poisson regression model, the probability $P(y_i)$ of having y_i number of pedestrian–vehicle crashes per five years at census tract i can be written as

$$P(y_i) = \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i!} \tag{1}$$

where λ_i is the Poisson parameter for census tract i, which is equal to census tract's expected number of pedestrian crashes per five years, $E[y_i]$. The Poisson regression model in Equation 1 can be estimated by specifying the Poisson parameter λ_i (i.e., the expected number of crashes) as a function of explanatory variables. The most common functional form is

$$\lambda_i = \exp(\boldsymbol{\beta X_i}) \tag{2}$$

where $\boldsymbol{X_i}$ is a vector of explanatory variables and $\boldsymbol{\beta}$ is a vector of estimable parameters.

Nevertheless, Poisson models are limited by their underlying assumptions. The Poisson model assumes the mean and variance of the number of crashes to be equal (i.e., $E[y_i] = \mathrm{Var}[y_i]$). If this equality does not hold, the data are said to be underdispersed ($E[y_i] > \mathrm{Var}[y_i]$) or overdispersed ($E[y_i] < \mathrm{Var}[y_i]$), and the standard errors of the parameter estimates will be incorrect (*15*), and consequently incorrect inferences will be drawn. Thus the Poisson model may not be appropriate for data sets violating this assumption. To address this underdispersion or overdispersion issue of the crash frequency data, the negative binomial model is derived by rewriting Equation 2 as follows:

$$\lambda_i = \exp(\boldsymbol{\beta X_i} + \epsilon_i) \tag{3}$$

where ϵ_i reflects random error term. For mathematical convenience, $\exp(\epsilon_i)$ is usually assumed gamma-distributed with a mean of 1 and a variance of α. The addition of this term allows the variance to differ from the mean as $\mathrm{Var}[y_i] = E[y_i][1 + \alpha E[y_i]] = E[y_i] + \alpha E[y_i]^2$. Then the probability density function for y_i is

$$P(y_i) = \left(\frac{1/\alpha}{(1/\alpha)+\lambda_i}\right)^{1/\alpha} \frac{\Gamma[(1/\alpha)+\lambda_i]}{\Gamma(1/\alpha)y_i!} \left(\frac{\lambda_i}{(1/\alpha)+\lambda_i}\right)^{y_i} \tag{4}$$

where $\Gamma(\cdot)$ is a gamma function. Note that the Poisson regression is actually a limiting model of the negative binomial regression as α approaches 0. Thus, if α is significantly (statistically) different from 0, the negative binomial is appropriate; otherwise, the Poisson model is an appropriate model (*15*).

To address the issue of unobserved heterogeneity, the concept of random parameters can be introduced in count-data models. Using simulated maximum likelihood, Greene developed procedures for estimating random parameter Poisson and negative binomial models (*33*). Anastasopoulos and Mannering applied these random parameter models for modeling the frequency of vehicle crashes occurring at interstate highways (*34*). In this study, a similar random parameter modeling technique is used that allows spatial unobserved heterogeneities for modeling pedestrian crash frequency. Thus the study has a novel contribution to the existing pedestrian crash frequency modeling.

Typically, random parameters are included in count-data models by writing the parameters as (*35*)

$$\beta_i = \beta + \omega_i \tag{5}$$

where ω_i is a random term. With this equation, the Poisson parameter of the negative binomial model in Equation 3 can be written as

$$\lambda_i | \omega_i = \exp(\beta_i X_i + \epsilon_i) \tag{6}$$

with the corresponding probability $P(y_i | \boldsymbol{\omega_i})$. The log likelihood function for the negative binomial model can now be written as

$$\mathrm{LL} = \sum_{\forall i} \ln \int_{\omega_i} g(\boldsymbol{\omega}_i) P(y_i | \boldsymbol{\omega}_i) d\boldsymbol{\omega}_i \tag{7}$$

where $g(\cdot)$ is the probability density function of the $\boldsymbol{\omega_i}$. Since the integration in Equation 7 has no closed form, a numerical integration of the negative binomial function over the distribution of the random parameters is required. This makes the analytical approach of the maximum likelihood estimation of the model computationally cumbersome. A simulation-based maximum likelihood method is therefore used. For simulation, Halton draws—which are more efficient than Monte Carlo random draws—are used (*36, 37*).

DATA

To estimate the model, the data of pedestrian–vehicle crashes that occurred in New York City are used. The data set consists of the records of crashes that occurred between 2002 and 2006 (*1*). The data set was built from various sources, including the New York City DOT, the New York State DOT, the Office of Vital Statistics of the New York City Department of Health and Mental Hygiene, the New York City Police Department, and NHTSA's Fatal Accident Reporting System (FARS), which are compiled and cross referenced (*1*). In addition, the geo-coded fatality analysis reporting system of the National Center for Statistical Analysis and the Multiple Cause of Death Files maintained by the Centers for Disease Control and Prevention are also considered for a clearer picture of the data. The Center for Transportation Injury Research compiled this comprehensive data set for the study (*1*).

Initially, the data were aggregated at two levels: zip code and census tract. In this paper, the focus is on the census tract level because the data at this level provide a greater number of explanatory variables and hence can give better insight into the effects of various factors on the number of pedestrian crashes. Because the number of fatal crashes is relatively low at the census tract level, the total number of crashes are used for modeling purpose instead of considering fatal and severe crashes separately. Figures 1 and 2 show the pedestrian–vehicle fatal

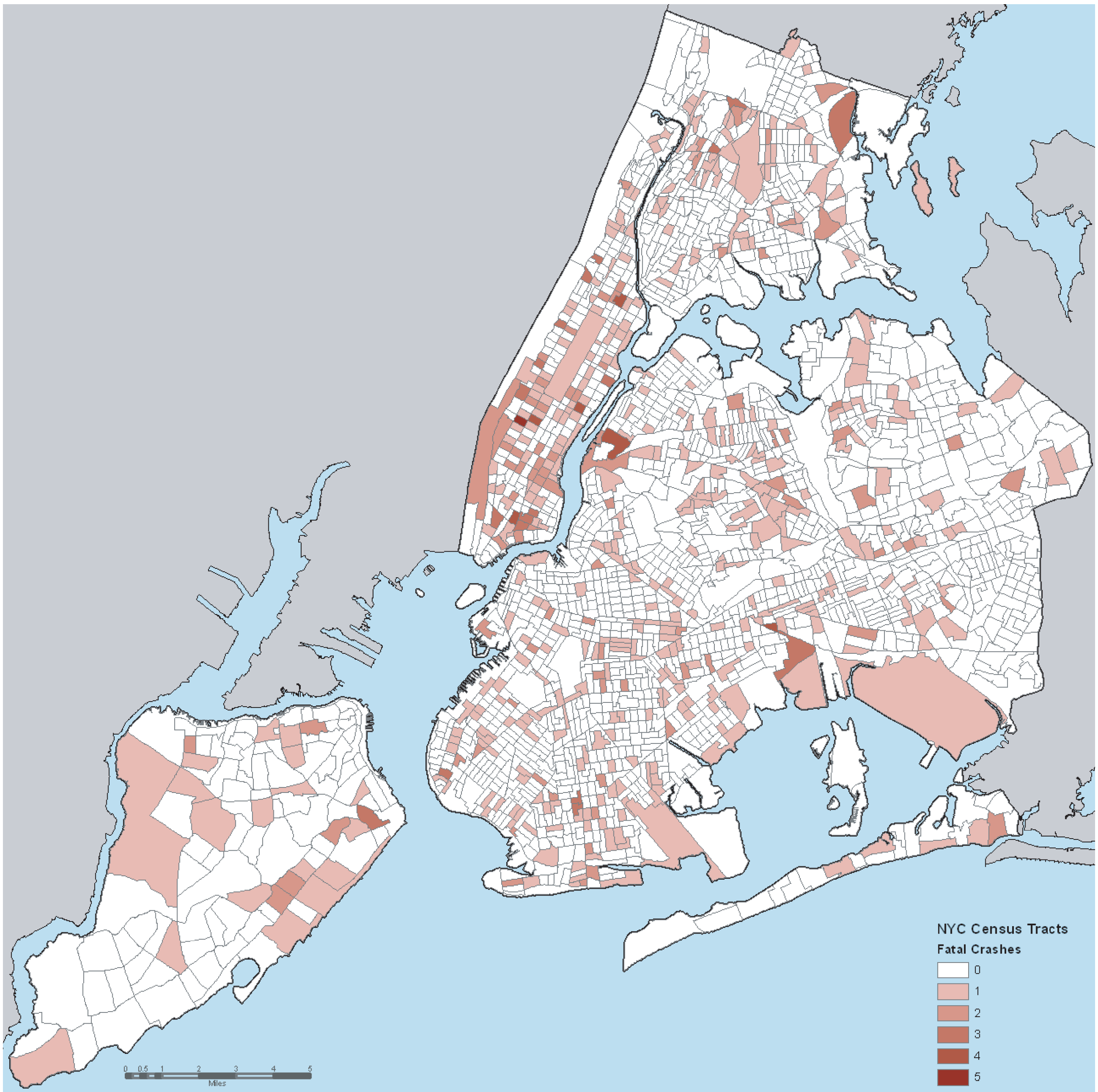

FIGURE 1 Geographical distribution of aggregated fatal pedestrian crashes in New York City at census tract level, 2002 to 2006 (*1*).

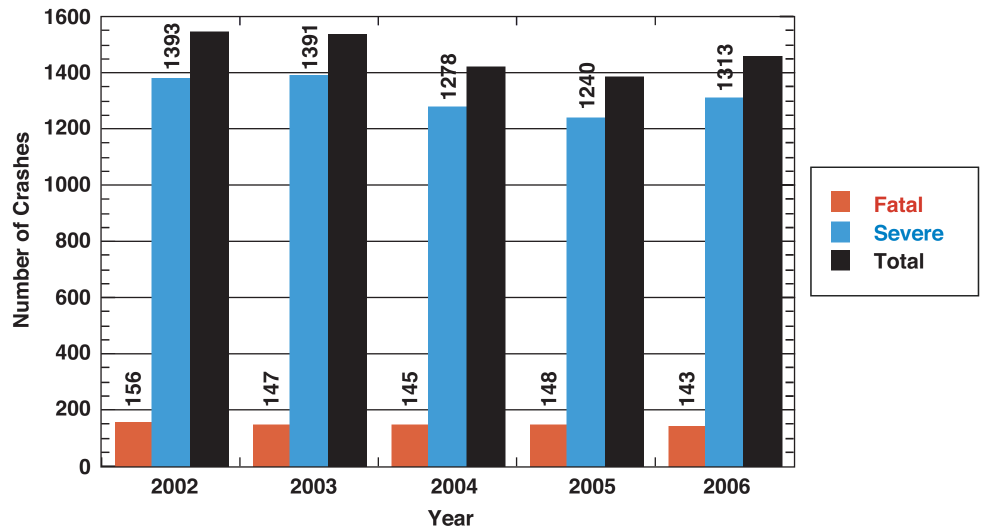

FIGURE 2 Aggregated pedestrian crashes in New York City at census tract level, 2002 to 2006 (*1*).

crashes at census tract level in New York City. In the data set, for each of 2,216 census tracts of New York City, the total number of fatal and severe pedestrian crashes are aggregated and the corresponding demographic information, land use patterns, and traffic system characteristics are gathered. However, because of missing information, a few observations from the model estimation are excluded (Table 1).

EMPIRICAL RESULTS AND DISCUSSION

The estimation results of the random parameters negative binomial model of the number of pedestrian crashes (severe and fatal) for a 5-year period are shown in Table 1. The explanatory variables used in the model are presented for three categories: demographic characteristics, land use patterns, and traffic system attributes (geometry and design related). Of these, the land use pattern and traffic system characteristics are classified as features of the built environment, and their influences are discussed separately. To estimate the model, the probability density function (distribution) of the parameters is specified and a simulation-based maximum likelihood approach with 200 Halton draws is applied (*33*). The random parameters are defined to have normal distribution, although other distributions (e.g., lognormal, uniform, or triangular) can also be used.

As shown in Table 1, the negative binomial dispersion parameter is significantly different from zero, which implies that the Poisson model would not have been appropriate for this data set. Among all the explanatory variables, 11 are found to be randomly distributed (normally distributed, in this case). When the standard deviation of a random parameter is not statistically different from zero (determined by *t*-stat), the parameter is considered to be fixed. The results indicate a reasonable increase in the log likelihood function value at convergence compared with the value at zero (restricted likelihood value), which is a sign of the good overall fit of the model.

Results show the effects of the demographic characteristics of New York City at the census tract level on the total number of pedestrian–vehicle crashes. The coefficient for total population (in millions) of a tract is found to be normally distributed with a mean of 0.073 and standard deviation of 0.001. Although this indicates that for all the zones, larger populations result in more pedestrian crashes, the influence of the population toward crash frequencies is not uniform across the tracts. LaScala et al. also found a similar positive relationship between population density and pedestrian crash rates for San Francisco, California (*38*). This finding suggests that greater regulation to minimize pedestrian–vehicle conflicts in places with higher populations may result in a reduction of pedestrian crashes.

The proportions of African-American and Hispanic populations also have significant correlations with crash occurrences. That is, the zones with a higher proportion of African-American and Hispanic populations are the most likely to have higher numbers of crashes. This is supported by other studies. For instance, a study on pedestrian crashes of Los Angeles, California, also found that pedestrian crashes are more likely to occur in low-income, minority neighborhoods (e.g., a neighborhood with an Hispanic population) (*31*). A similar trend is seen in zones with a greater proportion of the population group composed of individuals with a high school education. These population groups may be associated with a greater level of walking activities because of their low income, lack of personal vehicles, and types of occupation, and hence they have a greater exposure to traffic risks. Further, the coefficient for the proportion of the uneducated (without any schooling) population group is found to be normally distributed with a mean of 4.632 and standard deviation of 5.467. This implies that for a majority of the cases (80.16%), tracts with a greater proportion of an uneducated population are more prone to pedestrian crashes, for the similar reason of having a greater level of walking activities. However, 19.84% of the distribution is less than zero, which indicates that a few of the zones with a large uneducated population group are less prone to pedestrian–vehicle crashes. For these zones, some other unobserved factors, such as the level of traffic regulation and pedestrian safety measures, are compensating for the influences of this variable. These findings have implications for enhanced pedestrian safety measures in places with greater proportions of these population groups.

In addition, more median-age individuals within a zone results in more crashes. The zones with a higher proportion of the population that has reached the median age (as defined by the U.S. Census Bureau) are more likely to experience a greater number of pedestrian crashes. A greater proportion of median-age people may result from capturing older people in the statistic, which would explain the increased crashes.

TABLE 1 Estimation Results of Random Parameter Negative Binomial Model

Variable Description	Estimated Coefficients	*t*-Statistic
Demographic Characteristics		
Census tract population of 2000 (in millions)	0.073	16.797
(Standard deviation of the parameter)	(0.001)	(0.656)
Proportion of African-American population	0.295	8.623
Proportion of Hispanic population	0.181	3.702
Median-age population proportion	–0.031	–1.254
Proportion of the population who are high school graduates	0.507	4.195
Proportion of uneducated population	4.632	8.421
(Standard deviation of the parameter)	(5.467)	(13.864)
Land Use Attributes		
Industrial land use proportion of total land use	2.153	13.579
Open land use proportion of total land use	0.511	2.755
Commercial land use proportion of total land use	1.389	7.238
(Standard deviation of the parameter)	(0.692)	(5.065)
Total park area (in thousand acres)	–1.024	–6.175
(Standard deviation of the parameter)	(0.788)	(4.977)
Total number of schools in the area	0.026	3.681
Road Network and Intersection Operation Characteristics		
Total number of all-way stop intersections	0.007	1.691
Total number of signalized intersections	0.077	28.792
(Standard deviation of the parameter)	(0.018)	(14.367)
Number of three-approach intersections	–0.003	–3.188
(Standard deviation of the parameter)	(0.006)	(10.547)
Number of five-approach intersections	0.033	2.383
(Standard deviation of the parameter)	(0.065)	(5.042)
Primary roadway (with limited access) proportion of total roadway length	–1.050	–7.504
Primary roadway (without access restriction) proportion of total roadway length	0.530	5.842
Local rural road proportion of total roadway length	–0.207	–11.976
Other thoroughfare roadway proportion of total roadway length	–0.704	–4.728
Four-lane roadway proportion of total roadway length	1.243	6.321
Five-lane roadway proportion of total roadway length	2.896	4.049
Proportion of length of one-way streets to total roadway length	0.214	7.877
(Standard deviation of the parameter)	(0.083)	(6.541)
Proportion of length of roads with widths less than 30 ft to total roadway length	–0.418	–7.122
(Standard deviation of the parameter)	(0.564)	(14.321)
Number of subway stations in tract	0.114	6.337
(Standard deviation of the parameter)	(0.047)	(3.077)
Number of bus stops in tract	0.012	4.943
(Standard deviation of the parameter)	(0.010)	(9.010)
Dispersion parameter	6.425	15.283

NOTE: Constant: Estimated coefficient = –0.433, *t*-statistic = –8.277; number of observations = 2,193; log likelihood at zero = –11,313.61; log likelihood at convergence = –4,234.611; $\rho^2 = 0.626$.

EFFECTS OF BUILT ENVIRONMENT CHARACTERISTICS

Land use is an important built-environment factor for explaining crash frequency. Table 1 points to various land use attributes that have influence on the number of pedestrian crashes in a census tract. For example, the zones with the greater industrial and open land use are found to have more crashes. Although zones with greater industrial activity are assumed to have less pedestrian activity, these zones may be associated with greater crash risks because of a higher proportion of truck traffic. Open land use can increase pedestrian activities in a zone and thus increase the exposure to crash risks. This finding suggests, for improved patient safety, the implementation of countermeasures in places with a greater proportion of industrial and open land uses. Similarly, zones with higher commercial land use are more prone to pedestrian crashes. The coefficient for the proportion of commercial land use is found to be normally distributed with a mean of 1.389 and standard deviation of 0.692. This implies that for the majority of the tracts, pedestrian crashes increase with the increase of commercial land uses. Such land use can give rise to more pedestrian activities in a zone and can result in a relatively greater number of crashes compared with zones having low pedestrian activity. However, the random parameter associated with this variable indicates that the influence of the commercial land use pattern is not uniform across the zones. This reflects the presence of heterogeneous influences associated with commercial land use patterns and warrants appropriate methodology to allow such heterogeneities. A greater number of schools in a census tract is likely to increase the chances of crashes. These effects may be linked directly to higher exposure to crash risks because the presence of schools. A study of Los Angeles pedestrian crashes found more pedestrian crashes in areas with educational facilities and a higher percentage of commercial land use (*31*).

The coefficient for park coverage is found to be normally distributed with a mean of −1.024 and standard deviation of 0.788. This implies that for 90.31% of cases, the crash likelihood reduces with more park coverage, whereas for 9.69% of cases, the crash likelihood increases with more park coverage. The reduction of crashes could be related to the lack of risk exposure because of a lower density of roads in parks. However, unobserved factors such as the number of users of the parks (especially elderly people and children) may have attributed to the increase of crashes in some cases.

The variables related to road network and intersection operation characteristics are found to have significant influences on pedestrian crash occurrences at census tracts. Results (Table 1) indicate that areas with greater numbers of all-way-stop intersections are associated with more pedestrian–vehicle crashes. However, the parameter for the number of signalized intersections is found to be normally distributed, with a mean of 0.077 and standard deviation of 0.018, implying that for almost all zones, a greater number of signalized intersections results in more pedestrian crashes. Furthermore, the influence of each additional signalized intersection on pedestrian crashes will vary across the zones, reflecting the heterogeneous influences of unobserved factors such as the presence of marked crosswalks, pedestrian phases, the availability of raised medians, intersection traffic volume, and turning traffic.

Similarly, the coefficient for the number of three-approach intersections is found to be random, with normal distribution, with a mean −0.003 and standard deviation of 0.006. These values imply that for 69.15% of the tracts, a greater number of three-approach intersections would have fewer pedestrian crashes, whereas for 30.85% of tracts, more three-approach intersections would result in more pedestrian crashes. These varied influences of the three-approach intersections may be attributed to the previously mentioned unobserved factors related to the intersections. The coefficient for the number of five-approach intersections is also found to be random, with normal distribution, with a mean 0.033 and standard deviation of 0.065. With these values, 30.58% of the distribution is less than zero and 69.42% of the distribution is more than zero. This implies that for 30.58% of the tracts, more five-approach intersections would have fewer pedestrian crashes, whereas for 69.42% of tracts, more five-approach intersections result in more pedestrian crashes. That five-approach intersections are more positively related to pedestrian crashes than are three-approach intersections may be attributed to the risk of exposure, which is greater in five-way intersections than in three-way intersections. Nonetheless, these findings suggest that priority should be given to implementing pedestrian safety countermeasures at intersections to reduce the number of pedestrian crashes.

Additionally, zones with a greater proportion of four-lane roadways, five-lane roadways, and primary roadways without access restrictions are more prone to pedestrian–vehicle crashes. This outlines the importance of traffic calming on high-speed, high-volume roadways. Zones with a greater proportion of primary roadways with limited access, rural roadways, and thoroughfare roadways are likely to experience fewer pedestrian–vehicle crashes. This may reflect less pedestrian activity on these types of roads. The parameters for the proportion of one-way streets and narrow roads (width less than 30 ft) have been found to be normally distributed. For most of the cases (99.24%), pedestrian crashes increase with a greater proportion of one-way streets. This may be attributed to the absence of raised medians that provide refuge to pedestrians crossing the road. A majority of the zones (86.88%) with a greater proportion of narrow roads have a reduced crash likelihood. This finding suggests reducing the length of crosswalks to reduce the distances that pedestrians must cover when crossing a road.

In general, higher levels of transit infrastructure represent higher levels of pedestrian activities and thus might possess greater risks for pedestrian crashes. However, the model results indicate that the parameters for the number of subway stations and the number of bus stops follow normal distributions instead of fixed values, indicating their heterogeneous influences across different zones. Such heterogeneous influences may arise because of unobserved factors, such as the presence of marked crosswalks near subway stations, overpasses, the distance between the bus stop and crosswalk, and the lighting conditions of the roadways. Studies also revealed the influence of bus stops on pedestrian crashes. For example, Walgren found that in Seattle, 89% of high crash locations were within 150 ft of a bus stop, and 90% of those locations were within 70 ft of a crosswalk (*39*). A strong correlation between bus stops and more pedestrian crashes may be a result of visual impairments created by stopped buses for pedestrians trying to cross the roads. This finding suggests relocating or redesigning bus stops to avoid pedestrian visual impairments due to stopped buses.

SUMMARY AND CONCLUSIONS

This paper addressed the issue of unobserved heterogeneity for modeling pedestrian crash frequencies. A random parameter negative binomial model of pedestrian crash frequencies was presented for New York City at the census tract level. The study included a comprehensive set of variables describing the sociodemographic and built-environment characteristics of the tracts. The model found several parameters to be random, indicating their heterogeneous influences on the numbers of pedestrian crashes at a census tract level. These random parameters include the total population, the proportion of uneducated population, commercial land use, park coverage, the number of signalized intersections, the number of three-approach intersections, the number of three-approach intersections, the proportion of one-way streets and narrow roads, the number of subway stations, and the number of bus stops. Although these random parameters may be attributed to many unobserved factors involved with the pedestrian crashes, they provide better insights for understanding the heterogeneous influences on pedestrian crashes.

In this study, several sociodemographic characteristics that influence pedestrian crash likelihood were reported. For instance, a significant positive correlation between pedestrian crash frequencies and African-American or Hispanic neighborhoods, and between areas with a higher proportion of median-age population and uneducated population, was found. These findings are important for prioritizing neighborhoods for the development of effective pedestrian safety countermeasures in a megacity like New York, which has great population diversity.

Significant influences of built environment characteristics on pedestrian crashes are also found. For example, areas with a greater number of schools or a greater number of commercial and industrial land uses are more prone to pedestrian crashes. For majority of the cases, park coverage reduces the number of crashes. Road infrastructure such as intersection operation characteristics, the type of access control for the roads, the number of lanes, and the prevalence of one-way and narrow roads, significantly influences the number of crashes. Transit infrastructure such as the number of subway stations and bus stops also influences the number of pedestrian crashes.

Allowing for the unobserved heterogeneities across the tracts, the results show significant potential for framing policy for improved pedestrian safety countermeasures. New policy directions include

- Enhanced pedestrian safety measures through education and outreach efforts for areas with greater proportions of specific population groups, such as African-Americans, Hispanics, and the uneducated;
- More focus on engineering and education efforts in areas with significant commercial and industrial land use and a greater number of schools; and
- Specific regions having the scope of improving pedestrian safety include areas with a greater number of intersections, roads without access restrictions, and roads with more lanes.

The findings also imply prioritizing locations such as one-way streets, bus stops, and subway stations when implementing a safety program.

Although this study focused on unobserved heterogeneity issues, it had a few limitations. The model used an aggregate number of crashes during the 5-year period. However, in New York City, 5 years is a long time, because built-environment characteristics could change. Despite these shortcomings, the study presented a rigorous approach to modeling pedestrian crash frequency and guidance to New York City agencies to mitigate pedestrian crashes.

ACKNOWLEDGMENT

This work was funded by the New York City Department of Transportation.

REFERENCES

1. Shin, H.-S., G. Chen, K. Majka, A. Shorris, and S. Ukkusuri. *Pedestrian Fatality and Severe Injury Accidents in New York City.* New York City Department of Transportation, New York, 2010.
2. *Safe Streets NYC: Traffic Safety Improvements in New York City.* New York City Department of Transportation, New York, June 2007. http://www.nyc.gov/html/dot/pdf/safetyrpt07_1.pdf. Accessed Aug. 1, 2007.
3. *Traffic Safety Data Reports in 2006 by County.* Governor's Traffic Safety Committee. New York State, Albany, http://www.nysqtsc.state.ny.us/06data/datapack06.htm. Accessed Aug. 31, 2008.
4. Campbell, B. J., C. V. Zegeer, H. H. Huang, and M. J. Cynecki. *A Review of Pedestrian Safety Research in the United States and Abroad.* FHWA-RD-03-042. FHWA, U.S. Department of Transportation, 2004.
5. Oxley, J. A., E. Ihsen, B. N. Fildes, J. L. Charlton, and R. H. Day. Crossing Roads Safely: An Experimental Study of Age Differences in Gap Selection by Pedestrians. *Accident Analysis and Prevention,* Vol. 37, No. 5, 2005, pp. 962–971.
6. Lobjois, R., and V. Cavallo. Age-Related Differences in Street-Crossing Decisions: The Effects of Vehicle Speed and Time Constraints on Gap Selection in an Estimation Task. *Accident Analysis and Prevention,* Vol. 39, No. 5, 2007, pp. 934–943.
7. Ryb, G. E., P. C. Dischinger, J. A. Kufera, and C. A. Soderstrom. Social, Behavioral, and Driving Characteristics of Injured Pedestrians: A Comparison with Other Unintentional Trauma Patients. *Accident Analysis and Prevention,* Vol. 39, No. 2, 2007, pp. 313–318.
8. Leaf, W. A., and D. F. Preusser. *Identification of Alcohol-Pedestrian Crash Problems among Selected Racial-Ethnicity Groups.* DOT HS 808 641. U.S. Department of Transportation, 1997.
9. Moudon, A. V., L. Lin, P. Hurvitz, and P. Reeves. Risk of Pedestrian Collision Occurrence: Case Control Study of Collision Locations on State Routes in King County and Seattle, Washington. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2073,* Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 25–38.
10. Gårder, P. E. The Impact of Speed and Other Variables on Pedestrian Safety in Maine. *Accident Analysis and Prevention,* Vol. 36, No. 4, 2004, pp. 533–542.
11. Zegeer, C. V., R. Stewart, H. H. Huang, P. A. Lagerway, K. Todd, and P. Jacobsen. Safety Effects of Marked vs. Unmarked Crosswalks at Uncontrolled Locations: Analysis of Pedestrian Crashes in 30 Cities. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1773,* TRB, National Research Council, Washington, D.C., 2001, pp. 56–58.
12. Langlois, J., A. Penelope, M. Keyl, J. Guralnik, D. J. Foley, R. A. Marottoli, and R. B. Wallace. Characteristics of Older Pedestrians Who Have Difficulties Crossing the Street. *American Journal of Public Health,* Vol. 87, No. 3, 1997, pp. 393–397.
13. Lee, C., and M. Abdel-Aty. Comprehensive Analysis of Vehicle-Pedestrian Crashes at Intersections in Florida. *Accident Analysis and Prevention,* Vol. 37, No. 2, 2005, pp. 775–786.
14. Lord, D., and F. L. Mannering. The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives. *Transportation Research Part A,* Vol. 44, No. 5, 2010, pp. 291–305.
15. Washington, S. P., M. G. Karlaftis, and F. L. Mannering. *Statistical and Econometric Methods for Transportation Data Analysis.* Chapman and Hall, CRC, Boca Raton, Fla., 2003.
16. Stevenson, M. Building Safer Environments: Injury, Safety, and Our Surroundings. *Injury Prevention,* Vol. 12, No. 1, 2006, pp. 1–3.
17. Retting, R. A., S. A. Ferguson, and A. T. McCartt. A Review of Evidence-Based Traffic Engineering Measures Designed to Reduce Pedestrian-Motor Vehicle Crashes. *American Journal of Public Health,* Vol. 93, No. 9, 2003, pp. 1456–1463.
18. Graham, D. J., and S. Glaister. Spatial Variation in Road Pedestrian Casualties: The Role of Urban Scale, Density, and Land-Use Mix. *Urban Studies,* Vol. 40, No. 8, 2003, pp. 1591–1607.
19. Kim, K., and E. Yamashita. Motor Vehicle Crashes and Land Use: Empirical Analysis from Hawaii. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1784,* Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 73–79.
20. Kim, K., I. M. Brunner, and E. Y. Yamashita. Influence of Land Use, Population, Employment, and Economic Activity on Accidents. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1953,* Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 56–64.
21. Clifton, K. J., and K. Kreamer-Fults. An Examination of the Environmental Attributes Associated with Pedestrian–Vehicular Crashes Near Public Schools. *Accident Analysis and Prevention,* Vol. 39, No. 4, 2007, pp. 708–715.
22. McGinn, A. P., K. R. Evenson, A. H. Herring, S. L. Huston, and D. A. Rodriguez. Exploring Associations Between Physical Activity and Perceived and Objective Measures of the Built Environment. *Journal of Urban Health: Bulletin of the New York Academy of Medicine,* Vol. 84, No. 2, 2007, pp. 162–184.
23. Schuurman, N., J. Cinnamon, V. A. Crooks, and S. M. Hameed. Pedestrian Injury and the Built Environment: An Environmental Scan of Hotspots. *BMC Public Health,* Vol. 9, 2009, p. 233.

24. Ukkusuri, S. V., L. F. Miranda, G. Ramadurai, and J. Isa. *A Frequency Model of Pedestrian Accidents Accounting for Built Environment Characteristics.* Purdue University, West Lafayette, Ind., 2010.
25. LaScala, E. A., P. J. Gruenewald, and F. W. Johnson. An Ecological Study of the Locations of Schools and Child Pedestrians Injury Collisions. *Accident Analysis and Prevention,* Vol. 36, No. 4, 2004, pp. 569–576.
26. Campos-Outcalt, D., C. Bay, A. Dellapenna, and M. K. Cota. Pedestrian Fatalities by Race/Ethnicity in Arizona, 1990–1996. *American Journal of Preventive Medicine*, Vol. 23, No. 2, 2002, pp. 129–135.
27. Voas, R. B. A., S. Tippts, and D. A. Fisher. *Ethnicity and Alcohol-related Fatalities: 1990 to 1994.* NHTSA. U.S. Department of Transportation, 2000. http://www.nhtsa.dot.gov/people/injury/alcohol/Archive/ethnicity/ethnicity.html#Execsum. Accessed Aug. 10, 2008.
28. Gantz, T., E. J. D. L. Garza, D. R. Ragland, and L. Cohen. *Traffic Safety in Communities of Color.* Berkeley Traffic Safety Center, Calif., 2003. http://repositories.cdlib.org/cgi/viewcontent.cgi?article=1003&context=its/tsc. Accessed Aug. 10, 2008.
29. Pucher, J., and J. L. Renne. *Socioeconomics of Urban Travel: Evidence from the 2001 NHTS.* U.S. Department of Transportation. http://nhts.ornl.gov/2001/articles/socioeconomicsOfUrbanTravel.pdf. Accessed Feb. 11, 2008.
30. *Lives Still at Risk: A Look at the Groups that Traffic Safety Has Left Behind.* Berkeley Traffic Safety Center, Calif., http://www.tsc.berkeley.edu/html/newsletter/FAll03/pfleft_behind.html. Accessed Feb. 11, 2008.
31. Loukaitou-Sideris, A. Death on the Crosswalk: A Study of Pedestrian-Automobile Collisions in Los Angeles. *Journal of Planning Education and Research,* Vol. 26, No. 3, 2007, pp. 338–351.
32. Levine, N., K. E. Kim, and L. H. Nitz. Spatial Analysis of Honolulu Motor Vehicle Crashes: I. Spatial patterns. *Accident Analysis and Prevention,* Vol. 27, No. 5, 1995, pp. 663–674.
33. Greene, W. *Limdep, Version 9.0.* Econometric Software, Plainview, N.Y., 2007.
34. Anastasopoulos, P. C., and F. L. Mannering. A Note on Modeling Vehicle Accident Frequencies with Random-Parameters Count Models. *Accident Analysis and Prevention,* Vol. 41, No. 1, 2009, pp. 153–159.
35. Washington, S. P., M. G. Karlaftis, and F. L. Mannering. *Statistical and Econometric Methods for Transportation Data Analysis,* 2nd ed. Chapman and Hall, CRC, Boca Raton, Fla., 2010.
36. Bhat, C. Simulation Estimation of Mixed Discrete Choice Models Using Randomized and Scrambled Halton Sequences. *Transportation Research Part B,* Vol. 37, No. 1, 2003, pp. 837–855.
37. Train, K. *Halton Sequences for Mixed Logit.* PhD dissertation, Department of Economics, University of California, Berkeley, 1999.
38. LaScala, E. A., D. Gerber, and P. J. Gruenewald. Demographic and Environmental Correlates of Pedestrian Injury Collisions: A Spatial Analysis. *Accident Analysis and Prevention,* Vol. 32, No. 5, 2000, pp. 651–658.
39. Walgren, S. Using Geographic Information System (GIS) to Analyze Pedestrian Accidents. Presented at 68th Annual Meeting of the Institute of Transportation Engineers, Toronto, Ontario, Canada, 1998.

The authors are responsible for the findings in this work; the findings do not reflect the opinions of the New York City Department of Transportation.

The Safety Data, Analysis, and Evaluation Committee peer-reviewed this paper.