Accident Analysis and Prevention 39 (2007) 1267–1278

ACCIDENT ANALYSIS & PREVENTION

www.elsevier.com/locate/aap

Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes

N.N. Sze, S.C. Wong *

Department of Civil Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong, China

Received 20 December 2006; received in revised form 3 March 2007; accepted 25 March 2007

Abstract

This study attempts to evaluate the injury risk of pedestrian casualties in traffic crashes and to explore the factors that contribute to mortality and severe injury, using the comprehensive historical crash record that is maintained by the Hong Kong Transport Department. The injury, demographic, crash, environmental, geometric, and traffic characteristics of 73,746 pedestrian casualties that were involved in traffic crashes from 1991 to 2004 are considered. Binary logistic regression is used to determine the associations between the probability of fatality and severe injury and all contributory factors. A consideration of the influence of implicit attributes on the trend of pedestrian injury risk, temporal confounding, and interaction effects is progressively incorporated into the predictive model. To verify the goodness-of-fit of the proposed model, the Hosmer–Lemeshow test and logistic regression diagnostics are conducted. It is revealed that there is a decreasing trend in pedestrian injury risk, controlling for the influences of demographic, road environment, and other risk factors. In addition, the influences of pedestrian behavior, traffic congestion, and junction type on pedestrian injury risk are subject to temporal variation.

Keywords: Binary logistic regression; Pedestrian casualty; Injury severity; Interaction effect; Hosmer–Lemeshow statistic; Logistic regression diagnostics

1. Introduction

In Hong Kong, pedestrian fatalities account for one half of total fatalities in traffic crashes, a number that is much higher than that in the United States, Japan, and most Western countries (Hayakawa et al., 2000; NHTSA, 2006). The high pedestrian injury risk in Hong Kong could be attributed to pedestrian–vehicle conflicts in a densely populated urban area with numerous roadside activities (Loo et al., 2007). Effective remedial measures and road safety strategies that reduce pedestrian fatalities and casualties are essential. Recently, in land use and transportation planning, a greater emphasis has been placed on the needs of pedestrians, and pedestrian zones are now widely established. This encourages more walking trips, which are non-polluting, and offers a safer pedestrian environment. More importantly, pedestrian–vehicle conflicts on crowded urban road networks can be eliminated and, consequently, the crash and injury risk of pedestrians can be reduced (Transport Bureau, 1999).

Many researchers have attempted to establish a predictive model that identifies possible explanatory factors, such as traffic characteristics, road environment, and human error, for the probability of pedestrian–vehicle crashes through spatial analysis with comprehensive geographical information systems (LaScala et al., 2000; Garder, 2004; Loo and Tsui, 2004; Schneider et al., 2004). In addition, crash consequence models have been established to determine the injury severity of pedestrian casualties by logistic or ordered probit regression (Davis, 2001; Demetriades et al., 2004; Zajac and Ivan, 2003; Lee and Abdel-Aty, 2005). The likelihood of mortality and severe injury has been found to be highly correlated to vehicular speed and the crash environment.

Specifically, Al-Ghamdi (2002) suggested that crash attributes, together with the injury and demographic characteristics of the victims, determined the likelihood of mortality in pedestrian–vehicle crashes. Wazana et al. (2000) and Graham et al. (2005) attempted to estimate the risk of severe injury and mortality for pedestrians of different socioeconomic status. They found that children were at a greater risk of mortality and injury in traffic crashes, and these crashes were strongly influenced by environment and driver characteristics. Ballesteros et al. (2004), Martinez and Porter (2004), and

* Corresponding author. Tel.: +852 2859 1964; fax: +852 2559 5337.
E-mail address: hhecwsc@hkucc.hku.hk (S.C. Wong).

0001-4575/$ – see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2007.03.017

Roudsari et al. (2004) established a logistic regression model to estimate the associations between pedestrian mortality risk and various human, vehicular, and environmental contributory factors. Pedestrian–vehicle crashes that involved light truck vehicles, vans, or sport utility cars led to a higher risk of mortality and serious injury, which is attributed to the vehicle masses, speeds, and front-end design. In addition, driver and pedestrian errors that were related to alcohol and drug impairment also significantly contributed to severe injury.

In Hong Kong, taking advantage of the well-compiled historical crash record in the Traffic Accident Database System (TRADS), Yau (2004) and Yau et al. (2006) attempted to identify the factors that contribute to severe single- and multiple-vehicle crashes, respectively, by binary logistic regression. Using information that was obtained from the same system, the current study emphasized instead the injury risk of pedestrian casualties. Logistic regression was again applied to estimate the likelihood of mortality and severe injury in pedestrian casualties by considering the associations of such factors as demographic characteristics, injury characteristics, crash time, location, road environment, traffic control, and traffic conditions.

Controlling for the temporal variation in pedestrian injury risk, a design variable was introduced to deduce the trend of mortality and injury risk that was associated with the change in implicit attributes over a number of years. This change is most probably related to improvements in road infrastructure, advances in vehicular performance, and the enhancement of safety through the various remedial measures and road safety strategies that have been undertaken. More importantly, to address concerns over the suitability and efficiency of the proposed model, the Hosmer–Lemeshow test and logistic regression diagnostics were applied to measure the overall goodness-of-fit and to identify any outlier or poorly fit observations, respectively.

2. Data

The information that is used in this study was obtained from the TRADS that is maintained by the Hong Kong Police Force and Transport Department. TRADS consists of three components: crash environment profile, casualty injury profile, and vehicle involvement profile. The crash environment profile illustrates precisely the crash date, time, location, number of vehicles and casualties that were involved, weather conditions, road type, traffic conditions, and status of traffic control. The casualty injury profile indicates the role (whether the casualty is the driver, a passenger, or a pedestrian) and demographic characteristics of every victim, the use of a seat belt or helmet, the injury characteristics, the location of the passenger and/or pedestrian that was involved, the actions of the pedestrian, and any other special circumstances. The vehicle involvement profile provides driver information, vehicle class, license status, and age, and crash information for each of the vehicles that were involved. These profiles were not compiled into a database system with consistent, comprehensive, and compatible information until 1991, when the new system was established through a strategic recoding and aggregation process.

Table 1
Summary of the parameters in the pedestrian injury model

Factor	Attribute	Count (proportion)
Year	1991	6,326 (8.6%)
	1992	6,009 (8.1%)
	1993	6,005 (8.1%)
	1994	5,961 (8.1%)
	1995	5,435 (7.4%)
	1996	5,266 (7.1%)
	1997	5,320 (7.2%)
	1998	4,932 (6.7%)
	1999	4,830 (6.5%)
	2000	4,785 (6.5%)
	2001	4,978 (6.8%)
	2002	4,805 (6.5%)
	2003	4,517 (6.1%)
	2004	4,577 (6.2%)
Injury severity	Killed or severe injury	21,611 (29.3%)
	Slight injury	52,135 (70.7%)
Sex	Male	41,864 (56.8%)
	Female	31,882 (43.2%)
Age (years)	Under 15	15,304 (20.8%)
	15–65	45,214 (61.3%)
	Above 65	13,228 (17.9%)
Injury location	Head injury	24,898 (33.8%)
	Others	48,848 (66.2%)
Pedestrian location	On the crossing	10,422 (14.1%)
	Within 15 m of crossing	3,154 (85.9%)
	Others	60,170 (4.3%)
Pedestrian action	Crossing road or junction	35,831 (48.6%)
	Walking along footpath	21,503 (29.2%)
	Others	16,412 (22.2%)
Special circumstance	Overcrowded footpath	2,032 (2.8%)
	Obstructed footpath	927 (1.3%)
	Others	21,430 (29.1%)
	None	49,357 (66.8%)
Pedestrian contributory	Heedless crossing	23,747 (32.2%)
	Inattentive	2,208 (3.0%)
	Others	9,582 (13.0%)
	None	38,209 (51.8%)
Day of week	Monday–Friday	54,159 (73.4%)
	Weekend	19,587 (26.6%)
Time of day	7:00–9:59 a.m.	12,277 (16.6%)
	10:00 a.m.–3:59 p.m.	29,127 (39.5%)
	4:00–6:59 p.m.	16,810 (22.8%)
	7:00 p.m.–6:59 a.m.	15,532 (21.1%)
Speed limit	Above 50 km/h	1,245 (1.7%)
	50 km/h	70,934 (96.2%)
	Below 50 km/h	1,567 (2.1%)
Traffic aids	Poor aids	670 (0.9%)
	Normal	73,076 (99.1%)
Traffic congestion	Severe congestion	2,317 (3.1%)
	Moderate congestion	6,015 (8.2%)
	No congestion	65,414 (88.7%)
Obstruction	At or near obstruction	3,790 (5.1%)
	No obstruction nearby	69,956 (94.9%)
Junction control	Not at junction	38,379 (52.0%)
	Signalized intersection	14,997 (20.3%)
	Other control types	20,370 (27.6%)

Table 1 (*Continued*)

Factor	Attribute	Count (proportion)
Road type	Single-way carriageway	30,216 (40.9%)
	Two-way carriageway	31,825 (43.2%)
	Multi-/dual carriageway	11,705 (15.9%)
Lane number	More than two lanes	15,843 (21.5%)
	Two lanes	28,618 (38.8%)
	Single lane	29,285 (39.7%)
Environmental contributory	Pedestrian negligence	801 (1.1%)
	Other factors	1,946 (2.6%)
	None	70,999 (96.3%)

Number of observations = 73,746.

In Hong Kong, injury severity is divided into three levels: fatal, serious, and slight. Fatality refers to immediate death or subsequent death from injuries within 30 days of an accident, serious injury refers to injury and detention in a hospital for more than 12 h, and slight injury refers to an injury that does not require detention in a hospital for more than 12 h. This study measures the associations between the injury severity level of pedestrian casualties and all possible contributory factors. The dependent variable in the proposed model is injury outcome, which is dichotomous and in which the response of interest refers to killed and serious injury (KSI) and the response of contrast refers to slight injury.

By aggregating the crash environment profile and the casualty injury profile, this study attempted to establish a predictive model for pedestrian injury risk in which the predicting variables reflect the demographic characteristics of the pedestrian, including sex and age; the crash characteristics, including the injury location, crash location, crash time, special circumstances, and pedestrian contributory factors; and the traffic characteristics, including the road environment, speed limit, road type, traffic conditions, and junction controls. Table 1 presents a summary of the 73,746 pedestrian casualties from 1991 to 2004, which are used in the proposed predictive model.

3. Associations between injury severity and contributory factors

This study aims to evaluate the associations between pedestrian injury risk and possible contributory factors. The dichotomous nature of the dependent variable facilitates the application of binary logistic regression, for which the probability of killed or severe injury (KSI) against slight injury is estimated by the maximum likelihood method.

In the logistic regression model, a latent variable is formulated by the following expression:

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_j x_j + \cdots + \beta_p x_p, \quad (1)$$

where x_j is the value of the jth independent variable, with β_j as the corresponding coefficient, for $j = 1, 2, 3, \ldots, p$, and p is the number of independent variables.

With this latent variable, the conditional probability of a positive outcome is determined by

$$\pi(x) = \frac{\exp(g(x))}{1 + \exp(g(x))}. \quad (2)$$

The maximum likelihood method is then employed to measure the associations by constructing the likelihood function as follows:

$$l(\beta) = \prod_{i=1}^{n} \pi(x_i)^{y_i}(1 - \pi(x_i))^{1-y_i}, \quad (3)$$

where y_i denotes the ith observed outcome, with the value of either 0 or 1 only, and $i = 1, 2, 3, \ldots, n$, where n is the number of observations. By maximizing the log likelihood expression as

$$\mathrm{LL}(\beta) = \ln(l(\beta)) = \sum_{i=1}^{n} \{y_i \ln(\pi(x_i)) + (1 - y_i)\ln(1 - \pi(x_i))\}, \quad (4)$$

the best estimate of β could be obtained accordingly. The influence of attribute k on injury outcome could be revealed by the odds ratio

$$\mathrm{OR} = \exp(\beta_j), \quad (5)$$

with the 95% confidence intervals of $(\exp(\beta_j - 1.96 s_{\beta_j}), \exp(\beta_j + 1.96 s_{\beta_j}))$, where s_β is the standard error of the coefficient β. An odds ratio that is greater than 1 indicates that the concerned attribute leads to a higher injury risk, and vice versa.

The goodness-of-fit of the predictive model should be assessed. To evaluate the significance and predictive power of the logistic regression model, the change in deviance can be determined by comparing the log likelihood functions between the unrestricted model and the restricted model with the following expression:

$$G = -2(\mathrm{LL}(c) - \mathrm{LL}(\theta)), \quad (6)$$

where $\mathrm{LL}(c)$ is the log likelihood function of the restricted model and $\mathrm{LL}(\theta)$ is the log likelihood function of the unrestricted model. Under the null hypothesis that the coefficients for the predictive model are equal to zero, G is chi-square distributed with p degrees of freedom, where p is the number of variables that are considered. If G is significant at the 5% level, then the null hypothesis would be rejected, and one could conclude that the proposed model generally fits well with the observed outcome.

After the model building process with maximum likelihood, a series of analyses should be conducted to further assess the validity of the model. Its effectiveness in describing the associations between the dependent variable and possible contributory factors could be evaluated and revealed by numerical overall measures with the Hosmer–Lemeshow statistic (Hosmer and Lemeshow, 1980) and graphical microscopic illustrations with logistic regression diagnostics (Pregibon, 1981), respectively.

3.1. Hosmer–Lemeshow statistic

A summary measure verifies the fit of a model by first computing the difference between the observed and predicted scores throughout the entire collection of observations, then examining the individual contribution of the difference from each pair relative to the error structure of the whole model, and, finally, determining the suitability and efficiency of the model by a single statistic. Empirical consistency, which refers to the observed outcome behaving in accordance with the model prediction, is revealed by a chi-square statistic that is computed with a contingency table that collapses the observations into a fixed number of groups.

Hosmer and Lemeshow (1980) proposed a grouping strategy that is based on the value of estimated probabilities. Supposing that there are n observations in the probability model, Group 1 would consist of the n/g observations with the lowest predicted probabilities, Group 2 would consist of the n/g observations with the next lowest predicted probabilities, and so on. Once all of the groups are created, a Pearson chi-square statistic is estimated based on the observed and expected number of observations in the interest category of every group. When the Pearson chi-square statistic is significant, the null hypothesis that the proposed model sufficiently describes the empirical association is rejected.

Let n_k be the number of observations in the kth group and y_i be the response of the ith observation. Then, the number of observed responses of interest for the kth group would be

$$o_k = \sum_{i=1}^{n_k} y_i. \tag{7}$$

In contrast, if $\hat{\pi}_i$ denotes the predicted probability for the ith observation, then the average estimated probability for this group would be

$$\bar{\pi}_k = \sum_{i=1}^{n_k} \frac{\hat{\pi}_i}{n_k}. \tag{8}$$

Eventually, the Pearson chi-square statistic is determined by

$$\hat{C} = \sum_{k=1}^{g} \frac{(o_k - n_k\bar{\pi}_k)^2}{n_k\bar{\pi}_k(1-\bar{\pi}_k)}, \tag{9}$$

with a $(g-2)$ degree of freedom.

The Hosmer–Lemeshow statistic is usually computed using $g=10$ groups. However, because the restriction of $g>p+1$, where p is the number of covariates in the proposed model, arose in the simulation process that was demonstrated by Lemeshow and Hosmer (1982), $g=50$ is used in the current study, as more than 40 covariates are considered in the proposed logistic regression model.

3.2. Logistic regression diagnostics

The Hosmer–Lemeshow test evaluates the goodness-of-fit with a single number that summarizes the discrepancy between the observed and the predicted probability. It may be considered that the use of a single number to summarize a considerable amount of information is not satisfactory; therefore, a more precise approach may be sought. Regression diagnostics, which scan through the entire collection of observations and illustrate the agreement between the observed and predicted scores for all of the individual observations, should sufficiently satisfy doubts.

Let $\boldsymbol{X}$ denote the design matrix for the entire collection of observations, and the quantities that are central to the formation of regression diagnostics are the leverage values that are derived (Pregibon, 1981) by approximating the hat matrix

$$\boldsymbol{H} = \boldsymbol{V}^{1/2}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}^{1/2}, \tag{10}$$

where $\boldsymbol{V}$ is a diagonal matrix that is associated with the residuals of the regression model, for which $\boldsymbol{V}^{1/2}$ should be a diagonal matrix with a general element equal to $\sqrt{v}$.

After the estimation of leverage values, one can compute several useful diagnostic statistics that examine the influential power of the deficiency of every observation; in other words, the impacts on such attributes as likelihood ratio, deviance, and coefficient estimates. For regression diagnostics, one relies primarily on visual assessment. Generally, the model fits well when the influence diagnostic is not large over the entire collection of observations. In the current study, we examine by logistic regression diagnostics the influential power throughout the entire collection of observations on coefficient estimates with

$$\Delta\hat{\beta}_i = (\hat{\beta} - \hat{\beta}_{(-i)})^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{X})(\hat{\beta} - \hat{\beta}_{(-i)}), \tag{11}$$

where $\hat{\beta}$ denotes the estimated coefficients with a full sample and $\beta_{(-i)}$ denotes the estimated coefficients with a full sample that excludes the ith observation.

3.3. Associations measure

We first attempted to measure the associations between the KSI likelihood of pedestrian casualty and primary risk factors, including only demographics, crash, environment, and traffic characteristics, by logistic regression. Table 2 shows the results of odds ratio estimation. The sex and age of the casualty, injury location, pedestrian location and action, special circumstances of the crash, time of day, speed limit, congestion, junction controls, road type and geometry, and environmental contributory factors all significantly determined the probability of mortality and severe injury at the 5% level.

The following factors led to a significantly lower probability of KSI: the involvement of youths under the age of 15 years (odds ratio = 0.73) or pedestrians walking along a footpath (0.90); crashes that occurred on overcrowded (0.76) or obstructed (0.83) footpaths during the daytime (either morning peak [0.81], off-peak [0.70], or afternoon peak [0.75] hours), and on road sections with a speed limit under 50 km/h (0.88); moderate (0.82) and severe (0.87) congestion; and crashes that occurred at junctions with traffic controls other than traffic signals (0.92) and on roadways with two lanes (0.94) or a single lane only (0.79).

In contrast, the following led to a higher probability of KSI: the involvement of elderly people above the age of 65 years (odds ratio = 2.39), casualties with head injuries (3.82), pedestrians at a crossing (1.16) or within 15 m of a crossing (1.17), and pedestrians crossing a road or a junction (1.14); crashes that occurred on road sections with a speed limit over 50 km/h (1.97) and on

Table 2
Results of logistic regression on the base model

Factor	Attribute	Control	Odds ratio (95% CI)
Sex	Male	Female	0.96 (0.93–1.00)*
Age (years)	Under 15	15–65	0.73 (0.69–0.76)**
	Above 65		2.39 (2.28–2.49)**
Injury location	Head injury	Others	3.82 (3.69–3.96)**
Pedestrian location	On the crossing	Others	1.16 (1.10–1.22)**
	Within 15 m of crossing		1.17 (1.08–1.28)**
Pedestrian action	Crossing road or junction	Others	1.14 (1.09–1.20)**
	Walking along footpath		0.90 (0.85–0.94)**
Special circumstance	Overcrowded footpath	None	0.76 (0.68–0.86)**
	Obstructed footpath		0.83 (0.69–0.98)*
	Others		1.36 (1.30–1.42)**
Pedestrian contributory	Heedless crossing	None	1.04 (0.99–1.08)
	Inattentive		1.03 (0.92–1.15)
	Others		0.97 (0.92–1.03)
Day of week	Monday–Friday	Weekend	1.03 (0.99–1.07)
Time of day	7:00–9:59 a.m.	7:00 p.m.–6:59 a.m.	0.81 (0.77–0.86)**
	10:00 a.m.–3:59 p.m.		0.70 (0.67–0.74)**
	4:00–6:59 p.m.		0.75 (0.71–0.79)**
Speed limit	Above 50 km/h	50 km/h	1.97 (1.74–2.24)**
	Below 50 km/h		0.88 (0.77–1.00)*
Traffic aids	Poor aids	Normal	0.97 (0.83–1.15)
Traffic congestion	Severe congestion	None	0.82 (0.74–0.91)**
	Moderate congestion		0.87 (0.81–0.93)**
Obstruction	At or near obstruction	None	1.09 (1.01–1.18)*
Junction control	Not at junction	Signal	1.05 (1.00–1.10)
	Other control types		0.92 (0.87–0.98)**
Road type	Two-way carriageway	One way	1.23 (1.18–1.28)**
	Multi-/dual carriageway		1.43 (1.35–1.51)**
Lane number	Two lanes	>2	0.94 (0.90–0.99)*
	Single lane		0.79 (0.75–0.83)**
Environmental contributory	Pedestrian negligence	None	1.29 (1.10–1.51)**
	Other factors		1.08 (0.97–1.21)
Number of observations			73,746
Restricted log likelihood			−44605.78
Unrestricted log likelihood			−39034.41
Likelihood ratio statistic			11142.74**
Hosmer–Lemeshow statistic			92.38** (d.f. = 48)

* Statistically significant at the 5% level.
** Statistically significant at the 1% level.

single two-way (1.23) and segregated (1.43) carriageways; and pedestrian negligence (1.29).

Pedestrian injury risk was found to be strongly associated with various factors, including demographic, traffic, and environmental characteristics. With regard to demographics, the KSI risk of the elderly was double that of younger adults. This is expected, as the elderly are usually weaker in terms of their physiological condition and perception of safety, and slower in reacting to hazardous situations. Regarding injury consequence, no one is likely to disagree that casualties with head injuries are more likely to suffer severe injury or even die in traffic crashes.

With regard to spatial distribution, pedestrians who cross a road either at or outside of a marked crosswalk are more likely to have severe injury or mortality compared with those who walk along a footpath or in any other areas. This could be due to direct interaction with vehicles, especially in the absence of protective rails (Martinez and Porter, 2004). In addition, pedestrians who were heedless of crossings had a higher KSI likelihood. Their injury risk may have been modified by pedestrian impairments, such as alcohol intoxication, carelessness, or misjudgment, and also by the availability and accessibility of marked crosswalks (Al-Ghamdi, 2002; Loo and Tsui, 2005).

For crashes that occurred at junctions with traffic controls other than traffic signals, the risk of mortality and severe injury was lower, which could be because of conscious attentiveness on the part of both drivers and pedestrians at non-signalized intersections. Noland and Quddus (2005) studied the relationship between junction characteristics and crash risk and suggested that junctions were relatively hazardous for pedestrian–vehicle crashes in which junction density was found to be positively associated with slightly injured pedestrian casualties. However, some studies have suggested that crash risk is positively associated with non-signalized intersections and that the associations are different between urban and rural areas (Garder, 2004; Lee and Abdel-Aty, 2005). We suspect that certain other factors, such as land use, collision type, point of impact, and road user faults, which are not available in the current crash database, could enhance the accountability of the model.

With regard to environmental characteristics, pedestrian–vehicle crashes that occur during the daytime obviously have a lower injury risk. This finding is associated with the poor visual conditions, higher vehicular speed under lighter traffic flows, and possible negligence or inattentiveness that occur during nighttime.

With regard to geometry and traffic characteristics, crashes that occur near overcrowded or obstructed footpaths, on road sections with severe or moderate congestion, and on those of only a single lane or two lanes have a lower injury risk, whereas crashes that occur on road sections with speed limits higher than 50 km/h and on dual or multiple carriageways have a significantly higher injury risk. These phenomena are correlated with driver perception and vehicular speed under the corresponding circumstances, which, in turn, determine the collision speed and the crash consequence. An association between traffic safety and congestion level was revealed in a spatial analysis of the London metropolitan area, which found that traffic congestion tended to mitigate more severe traffic crashes and that the road infrastructure condition significantly interacted with pedestrian casualties. Higher population density, road density, and pedestrian activity near public transportation stations were found to be positively associated with mortality and severe injury during congested periods, especially in dense urban areas (Noland and Quddus, 2005).

4. Temporal change in pedestrian injury risk

Considering the results of the summary measure (in Table 2), our first attempt was unsatisfactory in approximating the KSI probability, with a Hosmer–Lemeshow statistic of 92.38 (d.f. = 48, $p < 0.001$). This suggested that the underlying associations between the KSI likelihood and primary risk factors should be revised, as some significant attributes had probably not yet been accounted for.

As revealed by the figures on pedestrian casualties by injury severity between 1991 and 2004 (as shown in Fig. 1), the proportion of mortality or severe injury decreased. This may be because of improvements in the road environment, advances in vehicular performance, and greater awareness of safety among drivers and pedestrians due to the various road safety campaigns and strategies that have been implemented in the past decade. The information on pedestrian flow or pedestrian–vehicle conflict could reveal the influence of pedestrian exposure to traffic crash on injury risk; this information, however, is not presented in the current study.

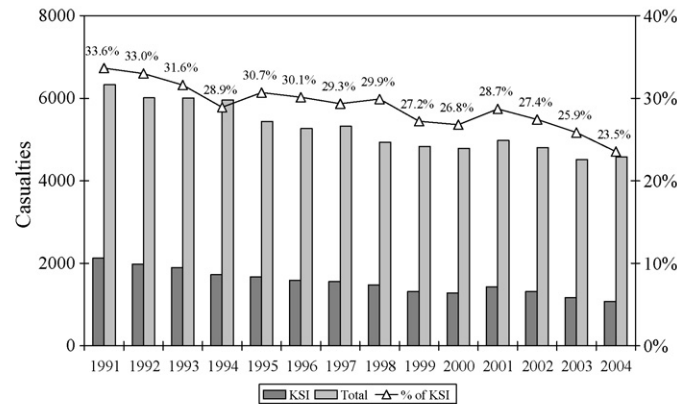

Fig. 1. Pedestrian casualties by injury severity from 1991 to 2004.

Controlling for the temporal confounding effect on KSI probability, the pedestrian injury model is revised by introducing a design variable, t, in the following expression:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \alpha t, \tag{12}$$

where t denotes the crash year ($t = 0$ for 1991, $t = 1$ for 1992, $t = 2$ for 1993, and so on).

Table 3 shows the results of logistic regression on the revised model, incorporating the temporal confounding effect. As expected, the temporal trend significantly altered the pedestrian injury risk at the 1% level. The KSI likelihood has decreased over the years (odds ratio = 0.97). In addition, the goodness-of-fit was improved (Hosmer–Lemeshow statistic = 70.45).

In addition to the impact of primary risk factors, such as demographic, environmental, and traffic characteristic, we must also take into consideration the temporal trend in pedestrian injury risk. The appreciable enhancement in pedestrian safety has been brought about by road safety projects and safety programs that have targeted vulnerable road-user or pedestrian groups. Previous research (Wong et al., 2004) has considered the causal effect of various road safety strategies on the reduction of the mortality and injury rate in Hong Kong. In general, the road safety strategies implemented in 1990s were effective. They included publicity through printed materials and electronic media, safety campaigns by the police, and publicity programs by non-governmental organizations, all of which played a crucial role in reducing the pedestrian fatality rate. In addition, the promotion of safe driving attitudes, police enforcement against speeding, traffic light violations, and drink driving, and driver improvement schemes also partially contributed to the improvement in safety. Moreover, the establishment of pedestrianized areas helped to create a safe pedestrian environment and to reduce potential pedestrian–vehicle conflicts in crowded urban areas.

Research has been conducted to identify the factors that are associated with temporal changes in crash and injury rates.

Table 3
Results of revised model, controlling for the temporal confounding effect

Factor	Attribute	Control	Odds ratio (95% CI)
Sex	Male	Female	0.96 (0.93–1.00)*
Age (years)	Under 15	15–65	0.72 (0.69–0.75)**
	Above 65		2.39 (2.29–2.50)**
Injury location	Head injury	Others	3.79 (3.66–3.93)**
Pedestrian location	On the crossing	Others	1.17 (1.11–1.23)**
	Within 15 m of crossing		1.21 (1.11–1.32)**
Pedestrian action	Crossing road or junction	Others	1.19 (1.14–1.25)**
	Walking along footpath		0.96 (0.91–1.01)
Special circumstance	Overcrowded footpath	None	0.81 (0.71–0.91)**
	Obstructed footpath		0.82 (0.69–0.98)*
	Others		1.36 (1.31–1.42)**
Pedestrian contributory	Heedless crossing	None	1.05 (1.00–1.09)*
	Inattentive		1.08 (0.97–1.19)
	Others		0.97 (0.91–1.03)
Day of week	Monday–Friday	Weekend	1.03 (0.99–1.07)
Time of day	7:00 am–9:59 am	7:00 p.m.–6:59 a.m.	0.81 (0.76–0.85)**
	10:00 am–3:59 pm		0.70 (0.67–0.73)**
	4:00 pm–6:59 pm		0.75 (0.71–0.79)**
Speed limit	Above 50 km/h	50 km/h	1.94 (1.71–2.20)**
	Below 50 km/h		0.90 (0.79–1.03)
Traffic aids	Poor aids	Normal	0.98 (0.81–1.20)
Traffic congestion	Severe congestion	None	0.88 (0.79–0.99)*
	Moderate congestion		0.87 (0.82–0.93)**
Obstruction	At or near obstruction	None	1.07 (0.99–1.16)
Junction control	Not at junction	Signal	1.07 (1.02–1.13)**
	Other control types		0.92 (0.87–0.97)**
Road type	Two-way carriageway	One way	1.20 (1.15–1.25)**
	Multi-/dual carriageway		1.39 (1.32–1.47)**
Lane number	Two lanes	>2	0.95 (0.90–0.99)*
	Single lane		0.78 (0.74–0.82)**
Environmental contributory	Pedestrian negligence	None	1.39 (1.18–1.63)**
	Others		1.03 (0.94–1.13)
Temporal confounding effect			0.974 (0.97–0.98)**
Number of observations			73746
Restricted log likelihood			−44605.78
Unrestricted log likelihood			−38974.38
Likelihood ratio statistic			11262.79**
Hosmer–Lemeshow statistic			70.45* (d.f. = 48)

* Statistically significant at the 5% level.
** Statistically significant at the 1% level.

Noland and Quddus (2004a) measured the time-series trend of pedestrian and bicycle casualties in the United Kingdom with cross-sectional time-series data. Proxy variables that represented the level of medical care and technology were applied to determine their impact on the associations between road infrastructure, demographics, and the casualty rate, and a year dummy was incorporated to deduce the time trend of the casualty rate. In addition, Beenstock and Gafni (2000) analyzed road safety performance in Israel. They noted not only that there is a global downward trend in the rate of road accidents, but also that vehicle quality, road infrastructure, police enforcement, and economic growth are significantly associated with the decline in the crash rate. The vehicle age distribution was an important intervening attribute. The findings of the current study support the position of Noland and Oh (2004), who argue that the time trend of the crash, injury, and fatality rate cannot be neglected in an effective and efficient injury risk model.

5. Interaction effects

Although the goodness-of-fit of the pedestrian injury risk model was substantially improved after a consideration of the temporal confounding effect on the association measures, the model was still not satisfactory (Hosmer–Lemeshow statistic = 70.45, p = 0.02). To further enhance the model, a careful investigation of the temporal variation in the associations between pedestrian injury risk and individual primary contributory factors (which are discussed in Section 3) was conducted.

In epidemiology, researchers apply the concepts of confounding and interaction to describe the change in associations between the outcome variable and the primary risk factors by any additional covariate adherence (Clayton and Hills, 1993; Sahai and Khurshid, 1995; Selvin, 1996). Confounding refers to the presence of partial or complete changes of the associations between the dependent and the independent variables of primary interest by a third variable; this is discussed in Section 4. However, in interaction, the associations between the dependent and independent variables are significantly differentiated at different levels of the third variable.

The confounding and interaction effects are normally assessed by logistic regression, adjusting the association measure with the potential confounders. As revealed in the previous section, the presence of a temporal confounding effect on pedestrian injury risk was significant. KSI likelihood noticeably decreased between 1991 and 2004. Now, by interaction modeling, we further examine the temporal variations over the underlying associations between KSI likelihood and every primary risk factor. This is to identify the significance of a change in the influential power of any primary risk factor on injury severity over the years.

The logistic regression modeling technique (Hosmer and Lemeshow, 2000) can be applied to assess the confounding and interaction effects. Denote Models (I) and (II) as the profiles with primary risk factor (x) only and with a potential confounder (t, the design variable for the crash year), respectively, as follows.

Model (I): $y = \beta_0 + \beta_1 x$.
Model (II): $y = \beta_0 + \beta_1 x + \beta_2 t$, where $t = 0, 1, 2, \ldots, T$, and T is the number of years, which is taken to be 13 in this study. To determine the significance of a confounding factor, the odds ratio estimates for the primary risk factor between Models (I) and (II) should be compared. For the interaction effect, Model (III), in which an interaction term ($x \times t$, the product of the primary risk factor and time variate) is induced as follows, should also be considered.
Model III: $y = \beta_0 + \beta_1 x + \beta_2 t + \beta_3 x \times t$, where $t = 0, 1, 2, \ldots, T$.

Now, the change in deviance (G) and the fit for the logit models with (III) and without (II) the interaction term should be determined. If the change in deviance (chi-square distributed with 1 degree of freedom) is significant, then the interaction effect should not be neglected, and, consequently, the odds ratio estimate of the concerned primary risk factor should be modified accordingly.

Table 4 illustrates the results of interaction modeling. The significance of temporal variation over the underlying associations between every potential risk factor and pedestrian injury risk was determined. Out of the 24 primary attributes under investigation, 11 experienced significant temporal change, for which the interactions by t on head injury (G = 4.31) and crashes at crossings (5.97), on overcrowded footpaths (6.31), at junctions controlled by means other than traffic signals (4.90), and on two-way carriageways (4.28) were significant at the 5% level, whereas casualties under the age of 15 (11.33), pedestrian crossings on roads or at junctions (30.04), heedless crossing (14.84), road sections with severe (18.38) or moderate (15.20) congestion, and junctions without any control devices (26.74) were significant at the 1% level.

The pedestrian injury risk model was subsequently enhanced, controlling for the influence of significant temporal interaction on underlying attributes (at the 0.01 level) by incorporating corresponding temporal interaction terms. Table 5 shows the results by logistic regression of the enhanced model. The model prediction now fit well, with the observed outcome at the 95% confidence level (Hosmer–Lemeshow statistic = 61.55, d.f. = 48, p = 0.09). The odds ratio estimation in the enhanced model revealed that the crossing road/junction-temporal interaction (odds ratio = 1.02), the severe congestion-temporal interaction (1.05), and the not at junction-temporal interaction (0.98) were significant at the 1% level, whereas the heedless crossing-temporal interaction (1.01) was marginal at the 5% level.

In addition to the summary measure with the Hosmer–Lemeshow statistic, we also examined the agreement between the observed and predicted probability throughout the entire set of observations, with a graphical illustration by logistic regression diagnostics. As shown in Fig. 2a, the leverage values were negligible (all were less than 0.006), which indicated that there was no significant outlying observation that lay far away from the rest of the data. However, as is shown by the influence diagnostic in Fig. 2b, the changes in the coefficient estimate were very minor (all were less than 0.02). There was no significant extremity throughout the entire collection of observations, and their influential powers on the model estimate were negligible. These factors collectively verified the goodness-of-fit of the enhanced model.

The primary risk factors of the demographic, traffic, and environmental characteristics that contribute to mortality and severe injury in pedestrian casualties were identified in the first attempt with logistic regression. Through a step-by-step investigation, the temporal confounding and interaction effects on the associations between KSI likelihood and primary risk factors were then explored. By incorporating these influences, the efficiency and predictive power of the proposed injury risk model were progressively enhanced.

A consideration of temporal variation in predictive models is not rare in road safety research. Kopits and Cropper (2005) applied time-series data to identify that factors such as income level, demographics, motorization, road infrastructure, population, and the growth of vehicle fleets and urban areas had led to a decline in passenger and pedestrian fatalities in industrialized countries. Road–time and vehicle–time interactions were incorporated into the predictive model to approximate the effects of motorway building and vehicle fleet expansion on pedestrian fatalities over a number of years. The impact of vehicle fleet growth magnified over the years, whereas that of road construction diminished. Other researchers have applied panel data to

Table 4
Results of the hypothesis test for the temporal interaction effect

Factor	Attribute	Control	*G*
Sex	Male	Female	1.58
Age (years)	Under 15	15–65	11.33**
	Above 65		0.09
Injury location	Head injury	Others	4.31*
Pedestrian location	On the crossing	Others	5.97*
	Within 15 m of crossing		0.11
Pedestrian action	Crossing road/junction	Others	30.04**
Special circumstance	Overcrowded footpath	None	6.31*
	Obstructed footpath		0.17
	Others		1.33
Pedestrian contributory	Heedless crossing	None	14.84**
Time of day	7:00–9:59 a.m.	7:00 p.m.–6:59 a.m.	0.09
	10:00 a.m.–3:59 p.m.		0.94
	4:00–6:59 p.m.		2.26
Speed limit	Above 50 km/h	50 km/h	0.07
Congestion	Severe congestion	None	18.38**
	Moderate congestion		15.20**
Junction control	Not at junction	Signal	26.74**
	Other control types		4.90*
Road type	Two-way carriageway	Single way	4.28*
	Multi-/dual carriageway		3.15
Lane number	Two lanes	More than two lanes	2.91
	Single lane		0.54
Environmental contributory	Pedestrian negligence	None	0.66

* Statistically significant at the 5% level.
** Statistically significant at the 1% level.

measure the change in fatality and injury rate that is associated with improvements in road infrastructure and medical technology. Proxy variables, including road length and width, number of lanes, length of hospital detention, hospital staff, and hospital density, were applied to approximate the influence of implicit attributes on the casualty rate trend. It was revealed that medical technology enhanced road safety performance over a time-series course in terms of total fatalities, severe injury, and slight injury (Noland, 2003; Noland and Quddus, 2004b).

From the results of the enhanced model, temporal effect was found to modify significantly the underlying associations between KSI likelihood and certain primary risk factors. First, pedestrians crossing at roads or junctions had a higher KSI likelihood than did those standing or walking on footpaths. The

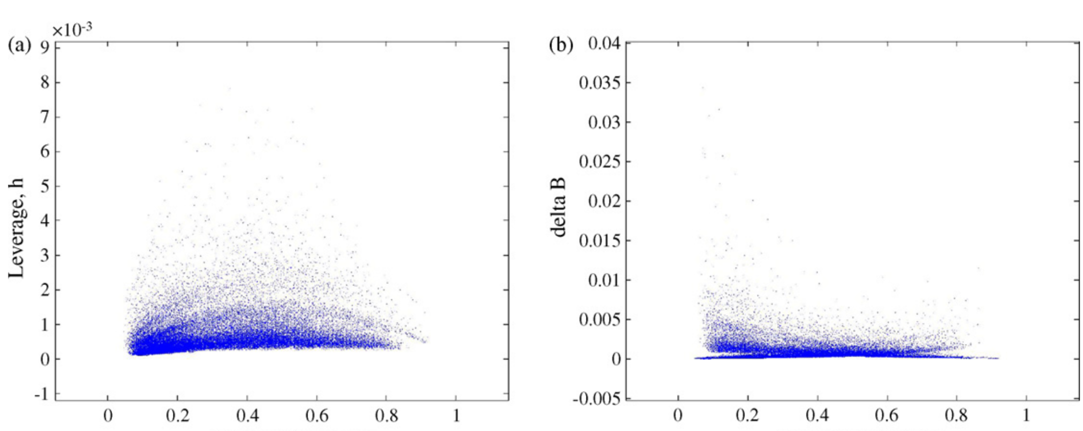

Fig. 2. Logistic regression diagnostics of the enhanced model.

Table 5
Results of the enhanced model, controlling for the temporal interaction effect

Factor	Attribute	Control	Odds ratio (95% CI)
Sex	Male	Female	0.96 (0.93–1.00)*
Age (years)	Under 15	15–65	0.69 (0.64–0.75)**
	Above 65		2.39 (2.29–2.50)**
Injury location	Head injury	Others	3.79 (3.66–3.92)**
Pedestrian location	On the crossing	Others	1.17 (1.11–1.23)**
	Within 15 m of crossing		1.21 (1.11–1.31)**
Pedestrian action	Crossing road/junction	Others	1.10 (1.03–1.18)**
	Walking along footpath		0.97 (0.92–1.03)
Special circumstance	Overcrowded footpath	None	0.81 (0.72–0.92)**
	Obstructed footpath		0.82 (0.69–0.97)*
	Others		1.36 (1.31–1.42)**
Pedestrian contributory	Heedless crossing	None	0.98 (0.94–1.03)
	Inattentive		1.08 (0.97–1.20)
	Others		0.97 (0.92–1.03)
Day of week	Monday–Friday	Weekend	1.03 (0.99–1.07)
Time of day	7:00–9:59 a.m.	7:00 p.m.–6:59 a.m.	0.81 (0.76–0.85)**
	10:00 a.m.–3:59 p.m.		0.70 (0.67–0.73)**
	4:00–6:59 p.m.		0.75 (0.71–0.79)**
Speed limit	Above 50 km/h	50 km/h	1.92 (1.69–2.18)**
	Below 50 km/h		0.92 (0.80–1.05)
Traffic aids	Poor aids	Normal	0.97 (0.79–1.18)
Traffic congestion	Severe congestion	None	0.55 (0.40–0.76)**
	Moderate congestion		0.85 (0.75–0.97)*
Obstruction	At or near	None	1.08 (1.00–1.16)
Junction control	Not at junction	Signal	1.20 (1.12–1.29)**
	Other control types		0.92 (0.87–0.98)**
Road type	Two-way carriageway	One way	1.20 (1.15–1.25)**
	Multi-/dual carriageway		1.39 (1.32–1.47)**
Lane number	Two lanes	>2	0.95 (0.90–0.99)*
	Single lane		0.78 (0.74–0.83)**
Environmental contributory	Pedestrian negligence	None	1.37 (1.17–1.61)**
	Others		1.04 (0.94–1.15)
Temporal confounding effect			0.97 (0.96–0.98)**
Temporal interaction effect	Age under 15 years		1.01 (0.99–1.02)
	Crossing road/junction		1.015 (1.01–1.02)**
	Heedless crossing		1.01 (1.00–1.02)*
	Severe congestion		1.05 (1.02–1.09)**
	Moderate congestion		1.00 (0.99–1.02)
	Not at junction		0.98 (0.97–0.99)**
Number of observations			73746
Restricted log likelihood			−44605.78
Unrestricted log likelihood			−38950.57
Likelihood ratio statistic			11310.42**
Hosmer–Lemeshow statistic			61.55 (d.f. = 48)

* Statistically significant at the 5% level.
** Statistically significant at the 1% level.

positive coefficient of the crossing road-temporal interaction suggested that the adverse impact magnified unfavorably over the years. This could be attributed to the increase in pedestrian activity that accompanied the extensive urban development and rapid economic growth in Hong Kong and to more frequent jay-walking or other hazardous pedestrian behavior, especially in the crowded CBD during peak hours.

Second, severe and moderate traffic congestion reduced the pedestrian KSI risk. The positive coefficient of the severe congestion-temporal interaction implied that the favorable effect diminished over time. The association between traffic congestion and pedestrian safety was sensitive to population density, road density, and pedestrian activity. The expansion of the railway system and the development of public transportation stations generated more pedestrian activity in urban areas, increased pedestrian flow and density, and, subsequently, modified the sensitivity of pedestrian fatality and injury risk to traffic congestion. This was also revealed in an analysis of the influence of land-use features on injury severity in vehicle–pedestrian crashes (Zajac and Ivan, 2003). However, the temporal interaction could also be attributed to changes in the police assessment and coding of traffic congestion in crash records. Road sections that in the past were regarded as “not congested,” and thus had a higher pedestrian KSI likelihood, may be regarded as “congested” today, as road users desire a higher traffic speed and flow to match advanced vehicular performance. Therefore, the pedestrian injury risk on “congested” road sections has increased.

Third, crashes that are not at intersections experienced a higher KSI likelihood relative to those at signalized intersections. This adverse impact favorably diminished over time, which is affirmed by the negative coefficient of temporal interaction. This may have been brought about by increased numbers of protective railings on footpaths and the frictional resistance of road surfaces, which reduce the impact force in vehicle–pedestrian collisions and thus decrease the sensitivity of junction type to KSI risk. In addition, remedial measures, such as converting yield-controlled intersections into signal-controlled intersections or providing grade-separated crossings, have been implemented at hot spots of pedestrian injury risk. Hence, hazardous crosswalks have been eliminated and pedestrian injury risk reduced. Moreover, changes in junction density due to substantial urban development could also contribute to the change in the relationship between pedestrian injury risk and junction type (Noland and Quddus, 2005).

6. Conclusions

This study established a predictive model for pedestrian injury risk in Hong Kong based on information from a comprehensive traffic crash record. A logistic regression model was used to identify possible contributory factors, such as the demographic characteristics of the victims, circumstances related to human error and crash characteristics, traffic flow conditions, and crash environment, to mortality and severe injury in pedestrian casualties in traffic crashes. The factors that led to a noticeably lower risk of pedestrian mortality and severe injury were being male and aged below 15, being on an overcrowded or obstructed footpath, and being involved in a daytime crash on a road section with severe or moderate congestion. A casualty age above 65 years, head injury, a crash at a crossing or within 15 m of a crosswalk, and a crash on a road section with a speed limit above 50 km/h, a signalized intersection, or two or more lanes obviously led to a higher risk of mortality and severe injury. The pedestrian injury risk underwent a noticeably decreasing trend from 1991 to 2004, which could be due to the remedial measures, road safety campaigns, pedestrianization, and traffic-calming strategies that were undertaken during the same period.

This study has also demonstrated a systematic data mining approach to an examination of significant contributory attributes from a huge database that consists of numerous covariates. The suitability, efficiency, and effectiveness of the predictive model were improved by step-by-step revision. The primary risk factors of demographic, traffic, environmental, and crash characteristics on pedestrian injury were first identified, followed by an exploration of the temporal confounding and interaction effects. Eventually, an enhanced predictive model was established by incorporating all of these. More importantly, the goodness-of-fit of the logistic model was verified with macroscopic evaluation by the Hosmer–Lemeshow statistic and microscopic illustration by logistic regression diagnostics.

Nevertheless, temporal variations over the underlying associations between injury risk and individual contributory factors are worth discussing. This study should serve as a first attempt at pedestrian safety assessment through the systematic data mining approach, controlling for the temporal interaction effect. However, the temporal variation on association measures is controversial. It is anticipated that the information on pedestrian flow and pedestrian–vehicle conflicts would help to reveal the influence of temporal variation on pedestrian exposure to traffic crashes. A more extensive data collection, such as from travel surveys (Allsop, 2005; To et al., 2005) and comprehensive analysis of the interaction effects between pedestrian flow and all risk factors could be an interesting topic in future research.

Acknowledgements

We gratefully acknowledge the assistance of Mr. T.F. Leung and Mr. Dominic Y.W. Wong of the Road Safety and Standards Division of the Transport Department of the Government of the Hong Kong Special Administrative Region (HKSAR) for providing traffic accident data for the research work that is reported in this paper. The views that are expressed in this paper are those of the authors and do not represent those of the Government of the HKSAR. Special thanks must also be given to Professor David W. Hosmer and Professor Stanley Lemeshow for their generous advice on the logistic diagnostics that are applied in this study. We would also like to acknowledge Professor Richard Allsop for his valuable recommendations on the analysis. The work that is described in this paper was supported by grants from the University Research Committee of the University of Hong Kong (10206125, 10207394), the Research Grants Council of the Hong Kong Special Administrative Region of China (Project No. HKU7183/06E), the National Natural Science Foundation of China (70629001), and the National Basic Research Program of China (2006CB705503).

References

Al-Ghamdi, A.S., 2002. Pedestrian–vehicle crashes and analytical techniques for stratified contingency tables. Accid. Anal. Prev. 34, 205–214.

Allsop, R.E., 2005. Travel survey data required to inform transport safety policy and practice. Transportmetrica 1, 241–245.

Ballesteros, M.F., Dischinger, P.C., Langenberg, P., 2004. Pedestrian injuries and vehicle type in Maryland. Accid. Anal. Prev. 36, 73–81.

Beenstock, M., Gafni, D., 2000. Globalization in road safety: explaining the downward trend in road accident rates in a single country. Accid. Anal. Prev. 32, 71–84.

Clayton, D., Hills, M., 1993. Statistical Models in Epidemiology. Oxford University Press, New York, USA.

Davis, G.A., 2001. Relating severity of pedestrian injury to impact speed in vehicle–pedestrian crashes. Transport. Res. Rec. 1773, 108–113.

Demetriades, D., Murray, J., Martin, M., Velmahos, G., Salim, A., Alo, K., Rhee, P., 2004. Pedestrians injured by automobiles: relationship of age to injury type and severity. J. Trauma 199, 382–387.

Garder, P.E., 2004. The impact of speed and other variables on pedestrian safety in Maine. Accid. Anal. Prev. 36, 533–542.

Graham, D., Glaister, S., Anderson, R., 2005. The effects of area deprivation on the incidence of child and adult pedestrian casualties in England. Accid. Anal. Prev. 37, 125–135.

Hayakawa, H., Fischbeck, P.S., Fischhoff, B., 2000. Traffic accident statistics and risk perceptions in Japan and the United States. Accid. Anal. Prev. 32, 827–835.

Hosmer, D.W., Lemeshow, S., 1980. A goodness-of-fit test for the multiple logistic regression model. Commun. Stat. A 10, 1043–1069.

Hosmer, D.W., Lemeshow, S., 2000. Applied Logistic Regression. Wiley-Interscience, New York, USA.

LaScala, E.A., Gerber, D., Gruenewald, P.J., 2000. Demographic and environmental correlates of pedestrian injury collisions: a spatial analysis. Accid. Anal. Prev. 32, 651–658.

Lee, C., Abdel-Aty, M., 2005. Comprehensive analysis of vehicle–pedestrian crashes at intersections in Florida. Accid. Anal. Prev. 37, 775–786.

Lemeshow, S., Hosmer, D.W., 1982. A review of goodness of fit statistics for use in the development of logistic regression models. Am. J. Epidemiol. 115, 92–106.

Loo, B.P.Y., Tsui, M.K., 2004. Pedestrian safety in Hong Kong: trends and spatial patterns. In: Proceedings of the Ninth International Conference of the Hong Kong Society for Transportation Studies, December 11, Hong Kong, p. 128.

Loo, B.P.Y., Tsui, M.K., 2005. Pedestrians' crossing behaviour in Hong Kong. In: Proceedings of the Tenth International Conference of the Hong Kong Society for Transportation Studies, 10 December, Hong Kong, pp. 471–480.

Loo, B.P.Y., Wong, S.C., Hung, W.T., Lo, H.K., 2007. A review of the road safety strategy in Hong Kong. J. Adv. Transport. 41, 3–37.

Kopits E., Cropper, M., 2005. Why have traffic fatalities declined in industrialized countries? Implications for pedestrian and vehicle occupants. World Bank Policy Research, Working Paper No. 3678.

Martinez, K.L.H., Porter, B.E., 2004. The likelihood of becoming a pedestrian fatality and drivers' knowledge of pedestrian rights and responsibilities in the Commonwealth of Virginia. Transport. Res. F 7, 43–58.

National Highway Traffic Safety Administration (NHTSA), 2006. Traffic Safety Facts, 2005 ed. National Highway Traffic Safety Administration, Washington, DC, USA.

Noland, R.B., 2003. Traffic fatalities and injuries: the effect of changes in infrastructure and other trends. Accid. Anal. Prev. 35, 599–611.

Noland, R.B., Oh, L., 2004. The effect of infrastructure and demographic change on traffic-related fatalities and crashes: a case study of Illinois county-level data. Accid. Anal. Prev. 36, 525–532.

Noland, R.B., Quddus, M.A., 2004a. Analysis of pedestrian and bicycle casualties with regional panel data. Transport. Res. Rec. 1897, 28–33.

Noland, R.B., Quddus, M.A., 2004b. Improvements in medical care and technology and reductions in traffic-related fatalities in Great Britain. Accid. Anal. Prev. 36, 103–113.

Noland, R.B., Quddus, M.A., 2005. Congestion and safety: a spatial analysis of London. Transport. Res. A 39, 737–754.

Pregibon, D., 1981. Logistic regression diagnostic. Ann. Stat. 9, 705–724.

Roudsari, B.S., Mock, C.N., Kaufman, R., Grossan, D., Henary, B.Y., Crandall, J., 2004. Pedestrian crashes: higher injury severity and mortality rate for light truck vehicles compared with passenger vehicles. Inj. Prev. 10, 154–158.

Sahai, H., Khurshid, A., 1995. Statistics and epidemiology: methods, techniques and applications. CRC Press, New York, USA.

Schneider, R.J., Ryznar, R.M., Khattak, A.J., 2004. An accident waiting to happen: a spatial approach to proactive pedestrian planning. Accid. Anal. Prev. 36, 193–211.

Selvin, S., 1996. Statistical Analysis of Epidemiologic Data, second ed. Oxford University Press, New York, USA.

To, D.K.B., Yau, K.T., Lam, A., 2005. Travel characteristics survey—method of expanding household interview survey data. Transportmetrica 1, 247–260.

Transport Bureau, 1999. Hong Kong moving ahead: a transport strategy for the future, October 1999, Hong Kong.

Wazana, A., Rynard, V.L., Raina, P., Krueger, P., 2000. Are child pedestrians at increased risk of injury on one-way compared to two-way streets? Can. J. Public Health 91, 201–206.

Wong, S.C., Leung, B.S.Y., Loo, B.P.Y., Hung, W.T., Lo, H.K., 2004. A qualitative assessment methodology for road safety policy strategies. Accid. Anal. Prev. 36, 281–293.

Yau, K.K.W., 2004. Risk factors affecting the severity of single vehicle traffic accidents in Hong Kong. Accid. Anal. Prev. 36, 333–340.

Yau, K.K.W., Lo, H.P., Fung, S.H.H., 2006. Multiple-vehicle traffic accidents in Hong Kong. Accid. Anal. Prev. 38, 1157–1161.

Zajac, S.S., Ivan, J.N., 2003. Factors influencing injury severity of motor vehicle-crossing pedestrian crashes in rural Connecticut. Accid. Anal. Prev. 35, 369–379.