Random parameter probit models to analyze pedestrian red-light violations and injury severity in pedestrian–motor vehicle crashes at signalized crossings

Jie WANG[1], Helai HUANG[2*], Pengpeng XU[3], Siqi XIE[3], S.C. WONG[3]

[1] School of Traffic and Transportation Engineering, Changsha University of Science & Technology, Changsha, Hunan, China
[2] School of Traffic and Transportation Engineering, Central South University, Changsha, Hunan, China
[3] Department of Civil Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong, China

***** Correspondence
E-mail address: jie_wang@csu.edu.cn (J. Wang), huanghelai@csu.edu.cn (H. Huang), xh150723@163.com (P. Xu), seakay@connect.hku.hk (S. Xie), hhecwsc@hku.hk (S.C. Wong)

Abstract: Pedestrian red-light violations at signalized crossings are an important traffic safety concern, particularly in densely populated cities. In this study, we quantitatively investigated the factors associated with pedestrian red-light violations and injury severity resulting from pedestrian–motor vehicle crashes at signalized crossings. Random parameter probit models are used to account for individual-specific heterogeneity that arises from a set of unmeasured factors related to traffic conditions and the pedestrians' physical and mental status. Data for the analysis are based on the historical crash record maintained by the Hong Kong Transport Department. Children younger than 11 years are not only associated with a higher likelihood of pedestrian red-light violations but also tend to have a higher probability of severe injuries. Factors including summer, dual carriageways with a central traffic island, and pedestrian age of 12 to 25 years are solely related to a higher likelihood of pedestrian red-light violations; meanwhile, variables solely associated with a higher probability of severe injuries include crashes that occur between 22:00 and 06:59, crashes occurring in rainy weather, crashes involving pedestrians older than 46 years, and bus crashes. Appropriate countermeasures are recommended to curb pedestrian red-light violations and to reduce the injury severity of pedestrian crashes.

Keywords: Pedestrian crash, Signalized crossing, Red-light violation, Injury severity, Random parameter probit model

1. Introduction

Motor vehicle–pedestrian crashes are of particular concern in densely populated urban areas. For example, in Hong Kong, pedestrian fatalities accounted for 62% of the total fatalities in traffic crashes in 2017 (Transport Department, 2018). Signalized pedestrian crossings are commonly designed for pedestrians to safely cross a road by separating pedestrians from vehicles in time. They are typically located at intersections where two or more roads meet or in the middle section of a road (usually known as mid-block crossings). If both motor vehicles and pedestrians strictly obey the rules of the traffic signals, the risk of conflicts or crashes can be minimized. However, widespread non-compliance with traffic rules has been observed, particularly by pedestrians. Pedestrian red-light violations at signalized crossings are an important risk factor of pedestrian crashes. Several early studies have found that approximately 25% of pedestrians cross illegally at intersections (Mullen et al., 1990). Similarly, in Hong Kong, pedestrian red-light violations contributed to a quarter of pedestrian crashes at signalized crossings (Transport Department, 2018). A comprehensive investigation of pedestrian red-light violation-related crashes would be very helpful for identifying key risk factors and proposing safety countermeasures, particularly for densely populated cities.

Both observational studies and experimental research approaches have been widely adopted to examine the pedestrian characteristics, road/traffic characteristics, and environmental factors that may increase the probability of pedestrian red-light violations when crossing a road. However, important gaps still exist. Researchers of observational studies conducted in real-life situations find it difficult to examine the role of precise traffic- or pedestrian-related characteristics. Experimental research may be performed to control factors such as traffic and pedestrian age but cannot provide a perfect assessment of the frequency of unsafe behavior (Holland and Hill, 2010). Furthermore, neither of the abovementioned approaches is capable of capturing the direct factors and/or behaviors that contribute to pedestrian crashes because observational and/or experimental data are always collected over a relatively short period (e.g., a few hours). Pedestrian red-light violations are regarded as one of the predominant factors associated with pedestrian crashes; however, the effects of red-light violations on crash occurrences may be inconsistent in different populations (e.g., young people vs. old people), and in different traffic situations (e.g., one-way vs. two-way streets). For example, previous studies (Guo et al., 2011; Ren et al., 2011) have shown that young pedestrians are more likely than older people to cross a road against traffic signals, yet the crash risk of red-light violations for young people is not higher because of their faster walking speed and better ability to perceive risk.

Accident study is another potential method to investigate irregular maneuvers. Pedestrian–motor crashes usually occur as a result of irregular maneuvers of pedestrians (such as failing to yield the right of way, crossing during a red-light phase, and walking while intoxicated) and/or illegal behaviors of drivers (such as excessive speed, running a red light, distraction, and reckless driving). Relying on long-term police-report crash data, we can model the likelihood of pre-crash pedestrian red-light violations as a function of the individual casualty, weather/temporal, and traffic/road characteristics. To the best of our knowledge, few studies have examined factors that affect pedestrian red-light violations at signal crossings by using an accident data analysis.

Pedestrian–vehicle crashes tend to have a higher proportion of fatal or serious injuries than vehicle–vehicle crashes. Moreover, pedestrian crashes at crossing locations have a higher proportion of fatal or serious injuries than crashes that occur on footpaths and carriageways (no crossing control) (Transport Department, 2018). Several previous studies have identified risk factors that influence the injury severity of pedestrian crashes at different locations, i.e., rural locations (Islam and Jones, 2014), urban locations (Abdul Aziz et al. 2013) and intersections (Haleem et al., 2015; Xu et al., 2016a; 2016b). However, the findings from a traffic safety study are not necessarily transferable between distant geographic locations (Ulfarsson et al., 2010). Therefore, a greater understanding of the location-specific factors associated with the injury severity of pedestrian crashes at signalized crossings is required to develop effective safety countermeasures.

Prevention of crash occurrence and reduction of crash injury severity are two basic strategies to improve traffic safety. In the case of pedestrian crashes at signalized crossings, pedestrian red-light violations are regarded as the predominant contributing factor to crash occurrence. Meanwhile, pedestrian crashes at crosswalks or signalized intersections tend to have a high risk of fatal or serious injury. This study applies a simultaneous analysis (two random parameter probit models: one for pedestrian red-light violations and the other for pedestrian injury severity) to investigate pedestrian crashes at signalized crossings. A simultaneous analysis can help build more comprehensive safety portraits because it simultaneously identifies significant factors related to the risk of pedestrian red-light violations and injury severity. If pedestrian characteristics increase the likelihood of red-light violations and fatal or serious injuries, for example, a safety enhancement program for targeted pedestrian groups could be considered a priority. Furthermore, a simultaneous analysis may ease the task of selecting circumstances for safety improvement interventions and potentially provide a more effective solution than a separated analysis and interventions for red-light violations and injury severity. Moreover, because details about several potential factors that affect pedestrian behavior and injury severity (such as the physical health of the pedestrians and the speed of the approaching vehicles) are unavailable in police-based crash records, the random parameter probit models are applied to account for the effects of the unobserved/uncollected factors.

2. Review

2.1 Pedestrian violation behaviors at crossings

In the past two decades, considerable observational and experimental studies have examined a variety of aspects related to the violation behaviors of pedestrians when crossing a road. These factors can be mainly divided into three groups: individual characteristics, roadway/traffic conditions, and weather/temporal conditions.

Individual characteristics such as pedestrian age and gender have been shown to be important contributing factors. Most of the previous studies have shown that male pedestrians appear to violate traffic rules more frequently than female pedestrians (Rosenbloom, 2009; Guo et al., 2011; Poó et al., 2018). Ren et al. (2011) showed a contradictory finding that women (particularly middle-aged women) are more likely to disregard traffic signals once they find a gap to cross. In contrast, in several other studies, gender failed to yield significant differences in offending crossing behavior (Dommes et al., 2015). The over-representation of older pedestrians in crash statistics was often explained by their significant frailty, slow walking speeds, and poor cognitive function (Holland and Hill, 2010; Dommes et al, 2014). Some studies also showed that older pedestrians wait for a longer time than younger ones at signalized crossings and are more inclined to obey traffic laws (Guo et al., 2011; Ren et al., 2011, Zhang et al., 2018).

Pedestrian red-light violations are also frequently associated with road and traffic characteristics, including the number of lanes (Cambon de Lavalette et al., 2009; Diependaele, 2018), the presence of a central traffic island (Cambon de Lavalette et al., 2009; Yan et al., 2016), the speed of the approaching cars (Lobjois et al., 2013; Sun et al., 2015), and the waiting time(Brosseau et al., 2013; Li, 2013, Diependaele, 2018).

The effects of weather/temporal conditions on pedestrian red-light violations have also been examined. An observational study by Li and Fernie (2010) showed that pedestrians walked faster and had lower compliance rates under cold and snowy conditions than under warm and dry pavement conditions. Liu and Tang (2014) indicated that, in contrast to bright headlights, the dim surroundings at dusk with poor visibility make pedestrians more cautious, thus increasing the number of safe road-crossing decisions. A survey study by Zhang et al. (2016) conducted in China showed a similar finding that pedestrians who crossed the street during the daytime were more likely to run the red light than those who crossed the street during other time periods (i.e., dusk, night, and dawn).

2.2 Pedestrian injury severity

Pedestrian–vehicle crashes are associated with a proportion of fatal or serious injuries. Several studies have investigated various contributing factors and their effects on the injury severity of pedestrian crashes with different modeling approaches. For example, Sarkar et al. (2011) identified pedestrian fatality risk factors along Bangladesh's roadways by using binary logistic regression models and found that pedestrians who crossed the road had a higher fatality risk than those who walked along the road. Meanwhile, pedestrian crashes at locations with no traffic control or stop control had a higher fatality risk than those at signalized intersections. Tarko and Azam (2011) linked police and hospital crash injury data to identify risk factors related to pedestrian injuries with a bivariate probit model. They found that rural roads and high-speed urban roads were dangerous for pedestrians, particularly when they were crossing such roads. Crossing a road between intersections (i.e., at mid-block locations) appeared to be a particularly dangerous behavior.

The random parameter model has been widely used to analyze crash injury severities because of its capability to account for unobserved predictors (Mannering et al., 2016). Abdul Aziz et al. (2013) developed separate random-parameter models for each borough (Manhattan, Bronx, Brooklyn, Queens, and Staten Island) of New York to explore the diversity of the determinants of pedestrian crash severity across the boroughs. The key findings showed that different parameters were found to be random for different boroughs. For example, the gender parameter was found to be random only for Manhattan and Queens, but it was fixed for Brooklyn. Similar studies were conducted by Islam and Jones (2014) and Haleem et al. (2015). Islam and Jones (2014) estimated random-parameter models to examine the factors that influence the severity of pedestrian at-fault crashes in urban and rural locations in Alabama. Their results showed that three variables—dark lighting conditions, two-lane roadways, and pedestrian age of 12 years or less—were significant in the cases of both urban and rural locations, whereas most variables were significant only in one location. Haleem et al. (2015) identified and compared the significant factors that affect pedestrian crash injury severity at signalized and unsignalized intersections. They found that very high pedestrian age, dark streets with no streetlights, and high speed limits increased the severity of pedestrian injury at both signalized and unsignalized intersections. Wang et al. (2016) developed a random parameter ordered probit model to identify the factors influencing the injury severity of pedestrians who are elderly involved in vehicle–pedestrian crashes. They found that the probabilities of fatality and serious injury in elder pedestrians during night time are more than double those occurred during daytime. Furthermore, Xin et al. (2017) used a random parameters generalized ordered probit model to analyze pedestrian crashes in Florida, and found that three factors (African American community, school zone, and bus stop area) related to neighborhood characteristics and the built environment have significant influences on pedestrian injury severity.

3. Data

The data for the statistical analysis in this study were obtained from the Traffic Accident Database System (TRADS) maintained by the Hong Kong Police Force and Transport Department. TRADS consists of three components: crash file, casualty file, and vehicle file. The crash file contains the precise crash date, time, location, number of vehicles and casualties involved, weather conditions, road type, traffic conditions, and the status of traffic control. The casualty file includes the role of the casualty (driver, passenger, or pedestrian) and the demographic characteristics of every victim, use of a seat belt or helmet, injury characteristics, location of the passenger and/or pedestrian, action of the pedestrian, and the contributory factor of the casualty (i.e., irregular maneuver of the casualty, such as pedestrian jay walking, crossing the road without paying attention to the traffic at the crossings, or being drunk). The vehicle file provides the vehicle and driver details, such as the age and the gender of the driver, the vehicle class, and the contributory factor of the driver/vehicle. In Hong Kong, injury severity is divided into three categories: fatal, serious, and slight (Sze and Wong, 2007).

The aim of this study was to examine the factors that contribute to pedestrian red-light violations and pedestrian injury severity at signalized crossings. Data for all pedestrian–motor vehicle crashes that occurred at signalized crossings in two years (2010 and 2012) were extracted from TRADS. Before conducting the statistical analysis, crashes with incomplete or missing data were removed from the dataset. The final dataset included 1752 pedestrian–motor vehicle crashes. Of these crashes, 2.8% or 49 crashes involved multiple pedestrians. In consistent with the previous study (Islam et al., 2014), if a crash involved multiple pedestrians, the pedestrian injury level and the pedestrian violation/non-violation were determined from the most severely injured pedestrian in the crash. Four categories of predictor variables were selected, including the time and environmental factors (i.e., season, time of day, rain, and natural light), roadways (i.e., road type and crossing type), pedestrian-related characteristics (i.e., pedestrian age and gender), and driver/vehicle-related characteristics (i.e., driver age and gender, and the class of the vehicle). Among the selected pedestrian–motor vehicle crashes, only 62 (3.64%) resulted in fatalities and were thus combined with serious crashes to form the category of fatal/serious crashes. A similar aggregation of pedestrian injury categories was commonly used in previous studies (Sze and Wong, 2007; Haleem et al., 2015). The response variables were set as follows: (a) whether there was a pedestrian red-light violation – “1” = yes and “0” = no; (b) whether a pedestrian suffered from a serious or fatal injury – “1” = yes and “0” = no. A pedestrian red-light violation was determined using a list of casualty contributory factors from the casualty file. Table 1 shows the summary statistics for these variables.

4. Method

Discrete choice models such as multinomial or binary logit/probit models are commonly applied to analyze the factors that influence crash injury severity and pre-crash violation behavior. These traditional models assume that the effect of each variable remains fixed across the observations. This restrictive assumption can be easily violated in the traffic safety analysis because some important variables that affect crash occurrence and injury severity are unavailable to the analyst (Mannering et al., 2016). In contrast, the random parameter model can address this issue by allowing the parameter estimates to randomly vary across the observations. In the absence of the data for unobserved predictors, the random parameter model becomes an attractive option.

The data collected for the analysis in this study were extracted from TRADS, which does not include some important variables that influence pedestrian behavior or injury severity. These important omitted variables mainly included the pedestrians' physical and mental status (e.g., walking speed and attention capacity), roadway and traffic conditions (e.g., traffic volume, speed of the approaching cars, and detailed signalization information), and the driver's risk perception (e.g., aggressive behavior and risky lifestyle). If these omitted variables are correlated to the observed exploratory variables, the effects of the corresponding observed variables will vary; this is called unobserved heterogeneity (Mannering et al., 2016). In such a case, it seems unrealistic to assume that the effects of the influential factors were fixed across all the observed accidents (as assumed by the traditional approaches with fixed parameters). Meanwhile, the dependent variables of both pedestrian red-light violation and injury severity models were the binary response outcomes in our study. Thus, the random parameter probit models (or more specifically, random parameter binary probit models) were selected to analyze pedestrian red-light violations and injury severity.

The standard probit model assumes that the response variable (i.e., injury severity level) can be represented by a latent and continuous variable y^* which is related to a vector of explanatory X, as follow (Yu and Abdel-Aty, 2014):

$$y^* = \beta' X + \varepsilon \qquad (1)$$

where y^* is the latent propensity variable; β' is the vector of estimable parameters; X is the vector of explanatory variables; and ε is the randomly distributed error term (assumed to be normally distributed with zero mean and unit variance). For the binary response variable case, we can specify that

$$\begin{aligned} y &= 0 \quad if \quad y^* \le 0 \\ y &= 1 \quad if \quad y^* > 0 \end{aligned} \qquad (2)$$

The predicted probability of y for given X can be estimates as

$$\begin{aligned} P(y=0|X) &= \Phi(-\beta' X) \\ P(y=1|X) &= 1-\Phi(-\beta' X) = \Phi(\beta' X) \end{aligned} \qquad (3)$$

where $\Phi(.)$ denotes the standard normal cumulative distribution function. For the injury severity model, $y=0$ indicates slight injury; $y=1$ indicates fatal/serious injury. For the pedestrian red-light violation model, $y=0$ indicates non-violation case; $y=1$ indicates violation case.

The random parameter probit model is a generalization of the standard probit model, which allows the parameter vector β_i to vary across the observations. Accordingly, the probability can be expressed as follows (Jalayer et al., 2018):

$$\beta_i = \beta + \varphi_i \tag{4}$$

where φ_i is a randomly distributed term (e.g., a normally distributed term with mean 0 and variance σ^2) that capture heterogeneity across observations. The model parameter could be estimated via the simulated maximum likelihood method and using Halton draws to maximize the simulated likelihood. The econometric and statistical software NLOGIT 4.0 was used for the model estimation. In this study, several random parameter distributions were tested, including normal, log-normal, triangular, and uniform distribution. In both red-light violations and injury severity models, the normal distribution was found to result in statistically superior outcomes and thus was selected to fit the final models.

5. Results

This study aimed to assess the association between risk factors and (i) pedestrian red-light violations and (ii) pedestrian injury severity. Before considering the detailed model estimates, we conducted a preliminary analysis to examine whether pedestrian red-light violations were correlated with pedestrian injury severity. Separating the data according to whether a pedestrian red-light violation is involved (N = 388) or not (N = 1364), we found that the proportion of fatal/serious injuries that occurred with pedestrian red-light violations (28.39%) did not differ significantly from that of fatal/serious injuries without pedestrian red-light violations (25.29%) (χ^2 = 1.492, p-value = 0.236). Thus, separate models using the random parameter probit approach were used for analyzing pedestrian red-light violations and pedestrian injury severity.

To focus on the most significant variables, only those with t-statistics corresponding to the 95% confidence interval or above were included in the final model. Meanwhile, the likelihood ratio test was used to guarantee that each additional variable significantly improved the overall model performance. The parameters were considered to be random across the observations if they yielded statistically significant standard deviations for the normal distribution. In contrast, if their standard deviations were not statistically different from zero, the parameters remained fixed.

Table 2 summarizes the parameter estimates for the random parameter probit models for pedestrian red-light violations and pedestrian injury severity. To directly assess the impacts of explanatory variables on probability of each outcome, marginal effects were also computed (Table 2). The marginal effects represent the change in the resulting probability of a particular outcome due to one unit change (or change from 0 to 1in the case of indicator variables) in an explanatory variable while holding all other variables constant. The results of the goodness-of-fit for the two random parameter probit models, including log-likelihood at convergence, log-likelihood at zero, and McFadden's pseudo R^2, are also presented in Table 2. Both models were considered to have a good fit because the McFadden's pseudo R^2 of the models was greater than 0.1 (Washington.et al., 2011).

From a methodological point of view and the application of the random parameter probit models, we found five parameter estimates to be statistically significant as the random parameters for the two estimation models. As shown in the pedestrian red-light violation model, rain, pedestrian age of 66 to 80 years, and driver age of more than 66 years were statistically significant random parameters. For the pedestrian injury severity model, driver age of less than 30 years and that of more than 66 years were significant random parameters. The effects of the rest of the factors were fixed across the populations for the pedestrian–motor vehicle crashes at signalized crossings.

In the pedestrian red-light violation model, the estimated parameter of rain was normally distributed (mean, −0.469; SD, 0.891). Given this estimate, 70.1% of the distribution was less than 0 and 29.9% of the distribution was greater than 0. We inferred that in 70.1% of the crash observations, the rainy condition reduced the lower likelihood of pedestrian red-light violations. Furthermore, its marginal effect was −0.114. This indicated that, for crashes occurring in rainy conditions, the probability of crashes involved pedestrian red-light violations decreased by 0.114 on average, compared to the crashes in sunny and cloudy conditions. The parameter for younger-old pedestrians aged between 66 and 80 years was normally distributed (mean, −0.291; SD, 1.118). For this parameter, 60.3% of the distribution was less than 0 and the remaining 39.7% was greater than 0. This implied that in 60.3% of the crash observations, younger-old pedestrians were associated with a lower likelihood of red-light violations. The parameter for drivers above 66 years of age was normally distributed (mean, −0.347; SD, 1.033). Given this estimate, 63.2% of the distribution was less than 0 and 36.8% of the distribution was greater than 0. Therefore, in 63.2% of the crash observations, drivers over 66 years of age were associated with a decreased likelihood of pedestrian red-light violations.

In the case of the injury severity model, the estimated parameters for drivers below 30 years of age and those over 66 years of age appeared to be random across crash observations. These two parameters were normally distributed (means, −0.854 and −0.307; SDs, 1.938 and 0.691, respectively). Given these estimates, the parameter values were less than 0 for 67.0% and 67.2% of the observed crashes, and greater than 0 for 33.0% and 32.8% of the observations, respectively. These results indicate that in a majority of the crash observations, both younger and older drivers were associated with decreased probabilities of fatal/serious injuries.

6. Discussion

On the basis of the estimated results of the two random parameter probit models, we can simultaneously identify the important factors that have significant effects on pedestrian red-light violations and the injury severity of pedestrians. As a summary, rain, one-way, dual carriageway, pedestrian age of less than 11 years, pedestrian age of 66 to 80 years, driver age of less than 30 years, and driver age of more than 66 years are found to be associated with both pedestrian red-light violations and pedestrian injury severity. Summer, mid-block crossing, and pedestrian age of 12 to 25 years are solely related to the likelihood of pedestrian red-light violations. Furthermore, the time periods 19:00–21:59 and 22:00–06:59, pedestrian age of 46 to 65 years, pedestrian age of more than 81 years, and buses are solely associated with the likelihood of pedestrian injury severity. These results could help in the recommendation of appropriate countermeasures and intervention strategies to curb pedestrian red-light violations (so as to prevent crash occurrence) and to reduce the severity of pedestrian crashes.

6.1 Time and environment

The day is divided into five periods: morning (07:00–09:59), mid-day (10:00–15:59), afternoon (16:00–18:59), evening (19:00–21:59), and night (22:00–06:59). Two periods are found to be significant for pedestrian injury severity. Compared with those that occur at mid-day, crashes that occur in the evening are associated with a decreased probability of fatal/serious injuries, which may be explained by slower speed and greater caution in light of the heavy pedestrian activity in this time period. In contrast, crashes that occur at night are associated with an increased likelihood of fatal/serious injuries. This could be attributed to the poor conspicuity of pedestrians and high vehicular speed because of light traffic flows during night time (Haleem et al., 2015; Wang et al., 2016). *Implicat*ions: In Hong Kong, pedestrians should be encouraged to wear high visibility clothing (or to place reflective markers on pedestrians' major joints) by education and outreach efforts. The appropriate use of reflective clothing has been proved to greatly enhance the pedestrian conspicuity at night and reduce severe pedestrian crashes in previous studies (Balk et al., 2007; Pour-Rouholamin and Zhou, 2016). Moreover, drivers should be educated to use more cautious driving behaviors during night, such as reducing driving speed.

With respect to the seasonality effect, pedestrians involved in crashes were associated with a higher likelihood of pedestrian red-light violations if the crash occurred in summer compared to spring. The need to hurry or the desire to keep moving is the main reason for the lack of compliance with pedestrian signals (Sisiopiku and Akin, 2003). Because Hong Kong's climate is subtropical, with bright sunshine and high temperature in summer, people are more likely to lose patience when exposed to scorching sunlight and consequently, are at a higher risk of red-light violations in summer than in the other seasons (Larsen and Sunde, 2008). *Implications*: Campaigns to combat pedestrian red-light violations when crossing the street in summer are required. For example, sunshield, a special traffic facility that is used only in summer and is designed to help protect riders/pedestrians from sunlight and high temperatures (Zhang and Wu, 2013), can be built at signalized intersections with high pedestrian volumes.

Consistent with previous studies, crashes that occur in rainy weather are associated with an increased probability of fatal/serious injuries (Islam and Jones, 2014; Haleem et al., 2015). In contrast, the rainy weather decreases the probability of pedestrian red-light violations, although this effect is heterogeneous. Because the model is conditional on crashes that have already occurred, a lower likelihood of pedestrian red-light violations implies a higher likelihood of driver violations. Therefore, two possible reasons can account for this result. One is that drivers are prone to be at fault under rainy conditions because of the poor visibility of the drivers to notice pedestrians and the long stopping distances under wet pavement conditions (Konstantopoulos et al., 2010). The other possible reason would be that pedestrians tend to comply more with the traffic signals under rainy conditions. The suspicion may fall on the latter, because it seemed to contradict the previous observational study by Li and Fernie (2010) that pedestrians have lower compliance under snowy conditions than under dry conditions. This result also implies that the crash risk of pedestrians crossing signalized crosswalks may be higher during raining. *Implications*: Campaigns focused on making pedestrians more aware of the fact that it is difficult for drivers to see them in rainy weather are required. Meanwhile, drivers are suggested to adapt their driving behavior to the prevailing conditions by reducing their driving speed and increasing the distance between vehicles during rain.

6.2 Roadways

Compared with crashes at intersection crossings, crashes that occur at mid-block crossings are associated with a lower likelihood of pedestrian red-light violations. This could be due to the lower compliance of the drivers at mid-block crossings with less complicated traffic conditions or the higher compliance of the pedestrians at mid-block crossings. Consistent with previous studies, crashes on one-way roads are associated with a lower likelihood of pedestrian red-light violations (Cambon de Lavalette et al., 2009; Dunbar, 2012) and decreased injury severity (Sze and Wong, 2007; Abdul Aziz et al., 2013) than those on two-way roads (without a central traffic island). Crashes on dual carriageways (with a central traffic island) have an increased probability of pedestrian red-light violations but a decreased probability of fatal/serious injuries. This implies that the safety effects of the central traffic island on pedestrian crossing behaviors are bidirectional, both positive and negative. A central traffic island can provide a certain level of protection from being hit by vehicles when crossing a trafficway. In contrast, this protection may discourage pedestrians from waiting until the crossing signal turns green (Cambon de Lavalette et al., 2009). *Implications*: We found that pedestrians are more likely involved in severe injury crashes when they cross two-way roads (without a central traffic island) and more than two carriageways, compared to crossing one-way roads and dual carriageways. In addition, we also found that the installation of central traffic island cannot significantly decrease the pedestrian red-light violations. In this case, overhead pedestrian bridges or underground pedestrian passageways, if it is possible, are recommended to replace some two and more lanes signalized crossings where there are many pedestrians.

6.3 Pedestrian-related factors

In this study, we divided the pedestrians into six age groups: below 11 years (children), 12 to 25 years (young), 26 to 45 years (younger-middle), 46 to 65 years (older-middle), 66 to 80 years (younger-old), and above 81 years (older-old).

Many studies have demonstrated that children are one of the most at-risk populations with respect to involvement in traffic accidents as pedestrians, particularly in urban environments (Toroyan and Peden, 2007; Rosenbloom et al., 2008; Verzosa and Miles, 2016). Our study agrees that children are at a high risk, as not only did the variable of children's age of less than 11 years increase the likelihood of pedestrian red-light violations, but it also increased the injury severity of pedestrian crashes as compared to pedestrian age of 26 to 45 years. Poorer traffic safety knowledge and lower capability for dealing with moving traffic situations are the main reasons for the higher risk of red-light violations by children (Rosenbloom et al., 2008). Furthermore, once the collision occurs, children are associated with a higher likelihood of fatal/serious injuries. A plausible explanation is that children experience more head injuries than adults when hit by a passenger car. To validate this finding, an investigation is conducted on the relationship between the percentage of head injuries and pedestrian age groups, and the results are shown in Fig. 1. This figure shows that children below 11 years of age have the highest percentage of head injuries among all age groups, whereas the older-young pedestrians 26 to 45 years of age have the lowest percentage of head injuries. Similarly, young pedestrians 12 to 25 years of age are associated with a higher likelihood of red-light violations but exhibit no significant association with the injury severity.

Pedestrian ages of 46 to 65 years, 66 to 80 years, and above 81 years are associated with increased probabilities of fatal/serious injuries as compared to pedestrian age of 26 to 45 years, mainly because of the high percentage of head injuries (see Fig.1) and the weak physical conditions of the pedestrians in the former age groups (Sze and Wong, 2007; Abdul Aziz et al., 2013; Verzosa and Miles, 2016). Meanwhile, the parameter for the pedestrian age of 66 to 80 years is found to be random in the red-light violation model. It is intuitive that older pedestrians tend to be more cautious and select a greater time gap to cross the road, which lead to a reduced probability of red-light violations. However, such an effect appears to not be uniform across the observations. This random effect may be attributed to some unmeasured factors not included in the current study, such as a decline in physical function (e.g., greater frailty, slower walking speeds, and poorer cognitive function) results in older pedestrians being sometimes more prone to violate red-lights when crossing the road (Holland and Hill, 2010).

Implications: Children exhibit a significantly higher risk of red-light violations and injury severity; this finding implies that children's safety should be considered a top priority for enhancing pedestrian safety at signalized crossings. Practical experience from some developed countries, such as Australia, has suggested that traffic and pedestrian safety education in elementary schools is effective in curbing children's violations and enhancing children's safety (Zegeer and Bushell, 2012). We recommend conducting safety awareness and education campaigns on road-crossing skills and road safety knowledge in elementary schools. Furthermore, elderly pedestrians could be the second target age group for pedestrian safety improvement because they demonstrate a significantly higher probability of fatal/serious injuries. Studies suggest that high vehicle speed is the most important contribution factor to elderly pedestrian crashes because of these people's slower walking speeds and poorer cognitive judgment of an approaching car's speed (Lobjois et al., 2013). Effective speed reduction measures (i.e., speed ramps and street narrowing) are recommended in areas with a high population of elderly pedestrians or a significant high number of pedestrian crashes.

6.4 Driver/vehicle-related factors

Drivers have been divided into three age groups, i.e., below 30 years (young), 31 to 65 years (middle), and above 66 years (old). Compared with middle-aged drivers, both young and old drivers are less likely to be involved in crashes in which the pedestrian violated the red light, although the effect of old drivers is heterogeneous. This result implies that young and old drivers have a higher likelihood of driving violations, which is in agreement with the findings of Ulfarsson et al. (2010) regarding the excess odds of being considered at fault of both young and old drivers in pedestrian–motor crashes. The parameters for both young and old drivers in the estimated model for pedestrian injury severity are found to be random. The random nature of the driver age parameters can be attributed to the fact that young (new) and old drivers in general drive slowly, but the relatively poor driving skills of the young drivers and the relatively weak physical condition of the old drivers lead to increased probabilities of fatal/serious injuries (McGwin and Brown, 1999). However, more in-depth research on the effect of driver age on pedestrian safety at signalized crossings is needed. A vehicle-related variable for buses is associated with a significantly increased probability of fatal/serious injury. This could be explained by the large size and heavy weight of a bus. Moreover, bus drivers tend fail to detect the crossing pedestrians.

6.5 Limitations and future research

The main limitation of our study is related to the source of information: a police-based register of crashes. This database does not include the pedestrians' physical and mental status (e.g., walking speed and attention capacity), roadway and traffic conditions (e.g., traffic volume, speed of the approaching cars, and waiting time), and drivers' risk perception (e.g., aggressive behavior and risky lifestyle), although these factors are deemed important. Although the random parameter probit models used in our study will mitigate the adverse effects of these potential variables, we still found it difficult to track the original source of heterogeneity and quantify the safety effects of the unobserved factors by using the resulting model estimates. Thus, one extension of this study is to incorporate police-reported data and other data sources (such as questionnaire surveys and field observations) to achieve a more explicit understanding of the casual mechanism underlying pedestrian crashes at signal crossings. Another possible extension relates to the advanced statistical method. There may be some same unobserved factors that influence the pedestrian red violations are likely to influence the injury severity. A seemingly unrelated regression models that correlated disturbance terms (Xu et al., 2018) will be further explored to investigate this issue.

Acknowledgments

This work was jointly supported by 1) the Joint Research Scheme of National Natural Science Foundation of China/Research Grants Council of Hong Kong (Project No. 71561167001 & N_HKU707/15), 2) the Natural Science Foundation of China (No. 713711921, 71671100); 3) the Research Fund for Fok Ying Tong Education Foundation of Hong Kong (142005); 4) Research Grants Council of Hong Kong (Project No. 17203017); and 5) Francis S Y Bong Professorship in Engineering.

References

Abdul Aziz, H. M., Ukkusuri, S. V., Hasan, S., 2013. Exploring the determinants of pedestrian–vehicle crash severity in New York City. Accident Analysis & Prevention, 50, 1298-1309.

Balk, S. A., Graving, J. S., Chanko, R. G., Tyrrell, R.A., 2007. Effects of retroreflector placement on the nighttime conspicuity of pedestrians: an open-road study. Human Factors & Ergonomics Society Annual Meeting Proceedings, 51(24), 1565-1568.

Brosseau, M., Zangenehpour, S., Saunier, N., Miranda-Moreno, L., 2013. The impact of waiting time and other factors on dangerous pedestrian crossings and violations at signalized intersections: A case study in Montreal. Transportation Research Part F: Traffic Psychology and Behaviour, 21, 159-172.

Cambon de Lavalette, B., Tijus, C., Poitrenaud, S., Leproux, C., Bergeron, J., Thouez, J., 2009. Pedestrian crossing decision-making: A situational and behavioral approach. Safety Science, 47, 1248-1253.

Diependaele, K., 2018. Non-compliance with pedestrian traffic lights in Belgian cities. Transportation Research Part F: Traffic Psychology and Behaviour. https://doi.org/10.1016/j.trf.2016.11.017

Dommes, A., Cavallo, V., Dubuisson, J., Tournier, I., Vienne, F., 2014. Crossing a two-way street: comparison of young and old pedestrians. Journal of Safety Research, 50, 27-34.

Dommes, A., Granié, M. A., Cloutier, M. S., Coquelet, C., Huguenin-Richard, F., 2015. Red light violations by adult pedestrians and other safety-related behaviors at signalized crosswalks. Accident Analysis & Prevention, 80, 67-75.

Dunbar G., 2012. The relative risk of nearside accidents is high for the youngest and oldest pedestrians. Accident Analysis & Prevention, 45, 517-521.

Guo, H., Gao, Z., Yiang, X., Jiang, X., 2011. Modeling pedestrian violation behavior at signalized crosswalks in China: A hazards-based duration approach. Traffic Injury Prevention, 12, 96-103.

Haleem, K., Alluri, P., Gan, A., 2015. Analyzing pedestrian crash injury severity at signalized and non-signalized locations. Accident Analysis & Prevention, 81, 14-23.

Holland, C., Hill, R., 2010. Gender differences in factors predicting unsafe crossing decisions in adult pedestrians across the lifespan: A simulation study. Accident Analysis & Prevention, 42, 1097-1106.

Islam, S., Jones, S. L., 2014. Pedestrian at-fault crashes on rural and urban roadways in Alabama. Accident Analysis & Prevention, 72, 267-276.

Jalayer, M., Shabanpour, R., Pourrouholamin, M., Golshani, N., Zhou, H., 2018. Wrong-way driving crashes: A random-parameters ordered probit analysis of injury severity. Accident Analysis & Prevention, 2018, 117, 128-135.

Konstantopoulos, P., Chapman, P., Crundall, D., 2010. Driver's visual attention as a function of driving experience and visibility. Using a driving simulator to explore drivers' eye movements in day, night and rain driving. Accident Analysis & Prevention, 42, 827-834.

Larsen, O. I., Sunde, Ø., 2008. Waiting time and the role and value of information in scheduled transport. Research in Transportation Economics, 23, 41-52.

Li, Y., Fernie, G., 2010. Pedestrian behavior and safety on a two-stage crossing with a center

refuge island and the effect of winter weather on pedestrian compliance rate. Accident Analysis & Prevention, 42, 1156-1163.
Li, B., 2013. A model of pedestrians' intended waiting times for street crossings at signalized intersections. Transportation Research Part B: Methodological, 51, 17-28.
Liu, Y., Tung, Y., 2014. Risk analysis of pedestrians' road-crossing decisions: Effects of age, time gap, time of day, and vehicle speed. Safety Science, 63, 77-82.
Lobjois, R., Benguigui, N., Cavallo, V., 2013. The effects of age and traffic density on street-crossing behavior. Accident Analysis & Prevention, 53, 166-175.
Mannering, F. L., Shankar, V., Bhat, C. R., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Analytic Methods in Accident Research, 11, 1-16.
McGwin, G., Brown, D., 1999. Characteristics of traffic crashes among young, middle-aged, and old drivers. Accident Analysis and Prevention, 31, 181-198.
Mullen, B., Cooper, C., Driskell, J. E., 1990. Jaywalking as a function of model behaviour. Personality and Social Psychology Bulletin, 16, 320-330.
Pour-Rouholamin, M., Zhou, H., 2016. Investigating the risk factors associated with pedestrian injury severity in Illinois. Journal of Safety Research, 57, 9-17.
Poó, F. M., Ledesma, R. D., Trujillo, R., 2018. Pedestrian crossing behavior, an observational study in the city of Ushuaia, Argentina. Traffic Injury Prevention, 19(3), 305-310.
Ren, G., Zhou, Z., Wang, W., Zhang, Y., Wang, W., 2011. Crossing behaviors of pedestrians at signalized intersections. Transportation Research Record, 2264, 65-73.
Rosenbloom, T., Ben-Eliyahu, A., Nemrodov, D., 2008. Children's crossing behavior with an accompanying adult. Safety Science, 46, 1248-1254.
Rosenbloom, T., 2009. Crossing at a red light: Behaviour of individuals and groups. Transportation Research Part F: Traffic Psychology and Behaviour, 12, 389-394.
Sarkar, S., Richard, T., Hunt, J., 2011. Logistic regression model of risk of fatality in vehicle–pedestrian crashes on national highways in Bangladesh. Transportation Research Record, 2264, 128-137.
Sisiopiku, V. P., Akin, D., 2003. Pedestrian behaviors at and perceptions towards various pedestrian facilities: An examination based on observation and survey data. Transportation Research Part F: Traffic Psychology and Behaviour, 6, 249-274.
Sun, R., Zhuang, X., Wu, C., Zhao, G., Zhang, K., 2015. The estimation of vehicle speed and stopping distance by pedestrians crossing streets in a naturalistic traffic environment. Transportation Research Part F: Traffic Psychology and Behaviour, 30, 97-106.
Sze, N. N., Wong, S. C., 2007. Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes. Accident Analysis & Prevention, 39, 1267-1278.
Tarko, A., Azam, M., 2011. Pedestrian injury analysis with consideration of the selectivity bias in linked police–hospital data. Accident Analysis & Prevention, 43, 1689-1695.
Transport Department, 2018. Road Traffic Accident Statistics, 2017. https://www.td.gov.hk/sc/road_safety/road_traffic_accident_statistics/2017/index.html
Toroyan, T., Peden, M. (Eds.), 2007. Youth and Road Safety. World Health Organization, Geneva.
Ulfarsson, G. F., Kim, S., Booth, K. M., 2010. Analyzing fault in pedestrian–motor vehicle crashes in North Carolina. Accident Analysis & Prevention, 42, 1805-1813.
Verzosa, N., Miles, R., 2016. Severity of road crashes involving pedestrians in Metro Manila,

Philippines. Accident Analysis & Prevention, 94, 216-226.
Washington, S. P., Karlaftis, M. G., Mannering, F. L., 2011. Statistical and Econometric Methods for Transportation Data Analysis, second ed., Chapman &Hall/CRC, Boca Raton, Florida.
Wang, Y., Haque, M. M., Chin, H. C., 2016. Elderly pedestrian injuries in Singapore. Journal of Transportation Safety & Security, 9(3), 273-300.
Xin, C., Guo, R., Wang, Z., Lu, Q., Lin, P.S., 2017. The effects of neighborhood characteristics and the built environment on pedestrian injury severity: A random parameters generalized ordered probability model with heterogeneity in means and variances. Analytic Methods in Accident Research, 16, 117-132.
Xu, X., Xie, S., Wong, S. C., et al, 2016a. Severity of pedestrian injuries due to traffic crashes at signalized intersections in Hong Kong: a Bayesian spatial logit model. Journal of Advanced Transportation, 50, 2015-2028.
Xu, X., Wong, S. C., Choi, K., 2016b. Quasi-induced exposure method for pedestrian safety at signalized intersections. Journal of Transportation Safety & Security, 8(2), 129-147.
Xu, X., Šarić, Ž., Zhu, F., Babićc, D., 2018. Accident severity levels and traffic signs interactions in state roads: a seemingly unrelated regression model in unbalanced panel data approach. Accident Analysis & Prevention, 120, 122-129.
Yan, F., Li, B., Zhang, W., Hu, G., 2016. Red-light running rates at five intersections by road user in Changsha, China: An observational study. Accident Analysis & Prevention, 95, 381-386.
Yu, R., Abdel-Aty, M., 2014. Using hierarchical Bayesian binary probit models to analyze crash injury severity on high speed facilities with real-time traffic data. Accident Analysis & Prevention, 62(2), 161-167.
Zegeer, C. V., Bushell, M., 2012. Pedestrian crash trends and potential countermeasures from around the world. Accident Analysis & Prevention, 44, 3-11.
Zhang, W., Wang, K., Wang, L., Feng, Z., Hu, Y., 2016. Exploring factors affecting pedestrians' red-light running behaviors at intersections in China. Accident Analysis & Prevention, 96, 71-78.
Zhuang, X., Wu, C., Ma, S., 2018. Cross or wait? Pedestrian decision making during clearance phase at signalized intersections. Accident Analysis & Prevention, 111, 115-124.
Zhang, Y., Wu, C., 2013. The effects of sunshields on red light running behavior of cyclists and electric bike riders. Accident Analysis & Prevention, 52, 210-218.

Table 1 Descriptive statistics

Variable	Summary statistics
Response	
Pedestrian injury severity	Fatal/serious injury = 455 (26.0%); slight injury* = 1297 (74.0%)
Pedestrian red-light violation	Violation case = 388 (22.1%); non-violation case* = 1364 (77.9%)
Time and environment	
Season	Spring* = 424 (24.2%); summer = 448 (25.6%); autumn = 448 (25.6%); winter = 432 (24.7%)
Time of day	7:00–9:59 = 262 (15.0%); 10:00–15:59* = 637 (36.4%); 16:00–18:59 = 374 (21.3%); 19:00–21:59 = 234 (13.4%); 22:00–6:59 = 245 (14.0%)
Day of week	Weekday* = 1277 (72.9%); weekend = 475 (27.1%)
Rain	Raining = 237 (13.5%); not raining* = 1515 (86.5%)
Natural light	Daylight* = 1211 (69.7%); dawn/dusk = 52 (3.0%); dark = 479 (27.3%)
Roadway	
Road type	One-way = 766 (43.7%); two-way* = 437 (24.9%); dual carriageway (with a central traffic island) = 379 (21.6%); more than two carriageways = 170 (9.7%)
Crossing type	Mid-block crossing = 876 (50.0%); intersection crossing* = 87.6 (50.0%)
Pedestrian-related	
Pedestrian age (years)	Child (≤11) = 110 (6.3%); young (12–25) = 268 (15.3%); younger-middle (26–45)* = 474 (27.1%); older-middle (46–65) = 519 (29.6%); younger-old (66–80) = 255 (14.6%); older-old (≥81) = 126 (7.2%)
Pedestrian gender	Male* = 889 (50.7%); female = 863 (49.3%)
Driver/vehicle-related	
Driver age (years)	Young (≤30) = 225 (12.8%); younger-middle (31–45)* = 517 (29.5%); older-middle 46–65 = 785 (44.8%); old (≥66) = 225 (12.8%)
Driver gender	Male = 1633 (93.2%); female = 119 (6.8%)
Class of vehicle	Motorcycle = 76 (4.3%); private car* = 552 (31.5%); taxi = 422 (24.1%); bus = 257 (14.7%); truck = 337 (19.2%); other = 108 (6.2%)

Note: * represents the variables treated as the control.

Table 2 Random parameter probit estimation results for pedestrian red-light violations and injury severity

Variable	Red-light violation				Injury severity			
	Coefficient	St. error	*t*-statistic	Marginal effects	Coefficient	St. error	*t*-statistic	Marginal effects
Constant	-0.627^{**}	0.085	−7.37		-0.740^{**}	0.083	−8.96	
Time and Environment								
Summer	0.125^{*}	0.065	1.97	0.030	-			
19:00–21:59	-				-0.506^{**}	0.126	−4.01	-0.143
22:00−6:59	-				0.264^{**}	0.102	2.58	0.075
Raining	-0.469^{**}	0.141	−3.33	−0.114	0.293^{**}	0.098	2.99	0.083
s.d. Raininng	*0.891^{**}*	*0.142*	*6.30*					
Roadways								
Mid-block crossing	-0.282^{**}	0.076	−3.72	−0.069	-			
One-way	-0.589^{**}	0.088	−6.66	−0.143	-0.397^{**}	0.080	−4.94	−0.112
Dual carriageway (with central traffic island)	0.578^{**}	0.093	6.20	0.141	-0.477^{**}	0.098	−4.89	−0.135
Pedestrian-related								
Children (≤11 years)	0.377^{**}	0.133	2.83	0.092	0.313^{*}	0.147	2.13	0.089
Young (12–25 years)	0.315^{**}	0.102	3.09	0.077	-			
Older-middle (46–65 years)	-				0.252^{**}	0.088	2.84	0.071
Younger-old (66–80 years)	-0.291^{*}	0.142	−2.05	−0.071	0.751^{**}	0.102	7.35	0.212
s.d. Younger-old	*1.118^{**}*	*0.148*	*7.56*					
Older-old (≥81 years)	-				0.894^{**}	0.134	6.65	0.252
Driver/vehicle-related								
Young (≤30 years)	-0.367^{**}	0.119	−3.09	−0.089	-0.854^{**}	0.205	−4.16	−0.241
s.d. Young					*1.938^{**}*	*0.254*	*7.62*	
Old(≥66 years)	-0.347^{*}	0.142	−2.44	−0.084	-0.307^{**}	0.118	−2.60	−0.087
s.d. Old	*1.033^{**}*	*0.149*	*6.93*		*0.691^{**}*	*0.119*	*5.82*	
Bus	-				0.397^{**}	0.094	4.24	0.112
Number of observations	1752				1752			
Number of parameters	14				15			
Log-likelihood at convergence	−833.85				−922.46			
Log-likelihood at zero	−931.32				−1033.54			
McFadden’s pseudo R^2	0.105				0.107			

Note: s.d. denotes the abbreviation of standard deviation and the italicized text represents the estimates for the variables resulting in random parameters. Only variables that are significant at the 95% confidence level are presented herein. - indicates that the coefficient is statistically insignificant. *Level of significance >95%. **Level of significance >99%.

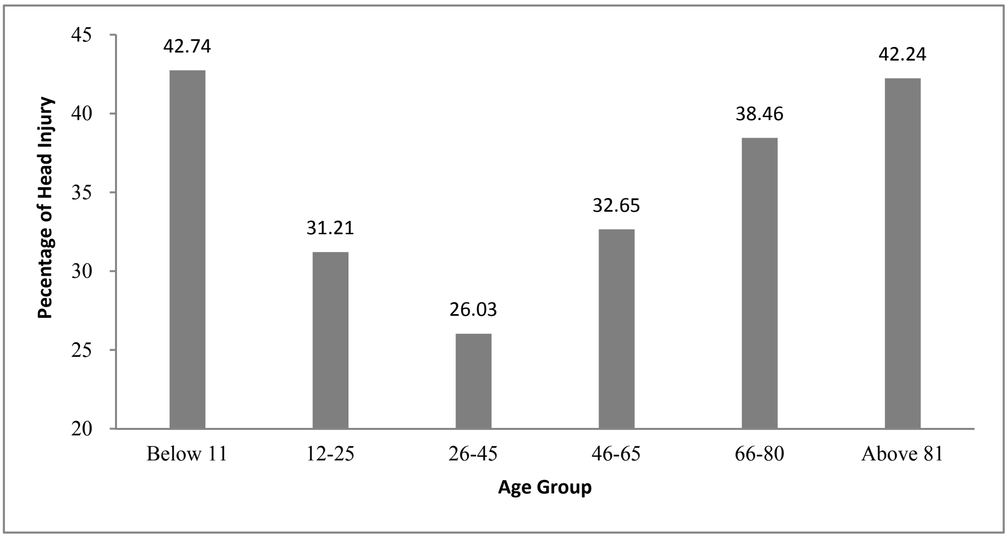

Fig.1. Distribution of Percentage of Head Injuries for Different Age Groups