PEDESTRIAN CROSSING BEHAVIOR AT SIGNALIZED CROSSWALKS

S.Q. XIE (Corresponding Author)

Department of Civil Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong, Tel.: 852-2859-2662; fax: 852-2517-0124; e-mail: seakay@connect.hku.hk

S.C. WONG

Department of Civil Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong, Tel.: 852-2859-1964; fax: 852-2559-5337; e-mail: hhecwsc@hku.hk

Tsz Man NG

Department of Civil Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong, e-mail: utszman@connect.hku.hk

William H. K. LAM

Department of Civil and Environmental Engineering, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, Tel.: 852- 2766-6045; fax: 852- 2365-9291; e-mail: william.lam@polyu.edu.hk

ABSTRACT

This study investigated pedestrian jaywalking at signalized crosswalks. Observational surveys were conducted at 7 crosswalks in different areas in Hong Kong, after which pedestrian information and site condition data were incorporated into a database. A binary logit model was used to identify possible factors that determine the probability of pedestrian jaywalking. To address the variation in the effects of the explanatory variables among pedestrians and the unobserved heterogeneity across sites, we used a random parameter model and a random effect model, respectively. The results showed that the random parameter model performed the best in terms of goodness-of-fit. It was found that the signal when a pedestrian arrives at the crosswalk is critical for decision making, and the jaywalking of surrounding pedestrians also influences the pedestrian's decision to cross. The gender and walking speed of the pedestrian, vehicle flow, and site location and condition of the crosswalk were also found to significantly determine the probability of pedestrian jaywalking.

Keywords: Pedestrian crossing behavior; Jaywalking; Signalized crosswalks; Random parameter model; Random effect model

INTRODUCTION

Signal control has been widely used around the world for more than 100 years. It provides a safe, economic, and efficient means of coordinating conflicting traffic flows at junctions, and is particularly popular in densely populated cities with heavy vehicle and pedestrian traffic loads. Signalized junctions are the most common type of junction in Hong Kong. Although pedestrian-vehicle collisions at signalized junctions have been reduced by 35% in the past 5 years, 387 pedestrian-vehicle crashes were still recorded, which comprised nearly 25% of the accidents that occurred at signalized junctions. Drivers should be aware of the traffic regulations as they must pass a written test on the Road Users' Code before obtaining their licenses, so pedestrian jaywalking is the most likely cause of pedestrian-vehicle accidents at signalized junctions in Hong Kong.

Studies have attempted to identify the factors that influence pedestrian crossing behavior. In terms of individual characteristics, it was observed that male pedestrians tend to jaywalk more than female pedestrians (Tiwari et al., 2007, Rosenbloom, 2009, Brosseau et al., 2013). A similar tendency was observed in questionnaire surveys based on the theory of planned behavior, which investigated pedestrians' attitudes toward jaywalking (Diaz, 2002, Zhou et al., 2009). However, Ren et al. (2011) observed that middle-aged female pedestrians had the lowest compliance rate in China. The elderly were also found to be more patient and less likely to jaywalk (Guo et al., 2011, Zhuang & Wu, 2011, Ren et al., 2011, Brosseau et al., 2013). Oxley (1997) conducted an experiment on pedestrian traffic judgment and observed that older adult pedestrians generally adopted a less safe crossing strategy and performed worse than younger pedestrians on two-way undivided roads, although their performance was similar to that of younger pedestrians on one-way divided roads. The differences associated with age-related physical, perceptual, and cognitive deficits were further discussed and validated in an experimental study of the age differences in pedestrians' gap selection (Oxley, 2005). Holland and Hill (2009) pointed out that driving experience also affected pedestrians' decisions to make unsafe crossings. Surprisingly, pedestrians with driving experience left smaller safety margins, although they were more likely to look both ways before crossing than non-drivers. Ren et al. (2011) suggested that a possible reason for the low compliance rate of female pedestrians in China was that fewer of them had driving licenses. They showed that individual characteristics affect pedestrians' judgement of the traffic conditions and gap selection. In view of this, Koh and Wong (2014) used a binary logit model to predict the proportion of pedestrians who accept a gap, and hence jaywalk. They found that the type of gap (location and sequence of oncoming vehicles) and the stage of crossing (near end or far end) influenced pedestrians' crossing decisions.

In addition to the pedestrian characteristics and types of gap, the environment and site conditions may also affect the decision making of pedestrians. Lavalette et al. (2009) found that the number of lanes of traffic, the presence of pedestrian crossing signals, and the presence of a central traffic island influenced pedestrians' decision making at crossings. Kruszyna and Rychlewski (2013) investigated the influence of approaching trams on pedestrian behavior at signalized crosswalks in Portland. Li and Fernie (2010) suggested that the weather also influenced the compliance rate, particularly for pedestrians crossing a signalized two-stage crossing with a center refuge island in the winter. The waiting time was also found to increase the probability of pedestrians jaywalking (Tiwari, 2007, Li and Ferinie, 2010), and Li (2013) proposed a model for pedestrians' intended waiting time. To reduce the high incidence of jaywalking and, hence, improve pedestrian safety at signalized crosswalks, pedestrian countdown signals have been introduced in recent years to prevent pedestrians from overestimating the waiting time (Keegan and Mahony, 2003) and taking the risk to jaywalk. This measure has been proven to effectively reduce the number of pedestrians starting to cross before the signal eventually turns green (Schattler et al., 2002).

Among the approaches used to identify the factors associated with pedestrian jaywalking behavior, ANOVA has been used to analyze the differences among groups of pedestrians (Li and Ferinie, 2010, Ren, et al., 2011) and logistic regression has been used to represent the effects of explanatory variables in determining the probability of jaywalking (Rosenbloom, 2009, Brosseau et al., 2013). ANOVA is useful for evaluating the influence of demographic factors, whereas logistic regression models are capable of linking the effects of the factors with the probability of jaywalking. However, the effects of explanatory variables are considered to be constant and fixed among all pedestrians in the simple logistic regression models, which may lead to misleading outcomes if considerable variation exists in the effects among individual pedestrians. In addition, although numerous previous studies have observed pedestrian crossing behavior at different sites, few studies have discussed the possible unobserved site differences.

In this study, observational surveys were conducted in 7 crosswalks in Hong Kong. The relevant individual-specific factors and site-specific factors were extracted and incorporated into a binary logit model to identify the contributory factors that determine the probability of jaywalking. To address the heterogeneity across pedestrians and sites, a random parameter model was used to accommodate the variation in the effects of the explanatory variables, and a random effect models was used to account for the unobserved heterogeneity across sites.

In Hong Kong, the sequence of pedestrian signals is a steady green signal, a flashing green signal, and a steady red signal. Pedestrians are only allowed to start crossing when the steady green signal is illuminated. The flashing green signal indicates that the pedestrians already on the crosswalk should continue and finish crossing at a reasonable speed. However, pedestrians who have not started crossing should wait until the next steady green signal. No pedestrians are allowed to cross during the red signal. In this study, pedestrians who entered a crosswalk during the flashing green signal or the red signal were regarded as jaywalkers according to the traffic regulations in Hong Kong. No countdown signals are provided at pedestrian crosswalks.

DATA

In this study, seven signalized junctions were randomly selected from different areas of Hong Kong (Table 1). There were 4 sites in urban areas, including 2 in Hong Kong Island and 2 in Kowloon, and the other 3 were in the New Territories. Video recording was conducted at each site for about 90 minutes, during which pedestrian movements were captured for further analysis. Preliminary analysis had previously been conducted based on the Travel Characteristic Survey 2011 to determine the period with the highest pedestrian flow on a typical working day from the video recording. In total, 7230 pedestrians who arrived during flashing green or red signals were recorded at the 7 sites. The number of observations varied from site to site, mainly depending on the populations of the areas. Table 1 lists the numbers of observations obtained at each site with the corresponding signal cycle time and average flow. The signal cycle times ranged from 90 seconds to 130 seconds. The crosswalk at Hung Hom had the lowest average pedestrian arrival rate at 2.7 ped/min, while the site at Tsuen Wan had the highest at 79.7 ped/min.

[Insert Table 1 Here]

To identify the factors that influenced the pedestrians' decisions to jaywalk, the pedestrian walking trajectories were manually tracked, and a series of variables were further extracted to build the dataset, including the demographic characteristics of the pedestrians, the pedestrian and traffic flow characteristics, the geometric design data, and the signal scheme of the junctions. The variables included are listed as Table 2. The upper part gives the proportions for the categorical variables, and the lower part provides the descriptive statistics of the continuous variables.

As shown in Table 2, only the pedestrians who arrived at the crosswalks during the flashing green (14.7%) or red (85.3%) signal were recorded. More than 60% of these pedestrians entered the crosswalks before the pedestrian signal finally turned green, i.e. jaywalked. It was found that 57% of the pedestrians who arrived at the crosswalks during the red signal jaywalked, and 100% of those who arrived during the flashing green signal jaywalked without waiting for another cycle. Some of the pedestrians may have thought that it was too long to wait for another cycle, and some may have been confused about the exact meaning of the flashing green signal and were unaware they were actually jaywalking.

The gender and age of the pedestrians were identified during the video tracking. The gender was easy to identify according to the pedestrians' appearance, and more than 90% of the observations were successfully distinguished. To accommodate the remaining unidentified cases, two dummy variables M, F were used to represent male pedestrians as M = 1 and F = 0, female pedestrians as M = 0 and F = 1, and the unidentified pedestrians as M = 0 and F = 0.

However, most of the pedestrians (96.1%) could not be identified as either elderly or children, and they were thus generally regarded as adults. As previously mentioned, there were four sites (two in Hong Kong Island and two in Kowloon) in urban areas, and three sites in the New Territories. We obtained 5064 observations (70.0%) from the four urban sites and 2166 observations (30.0%) from the other three sites in the New Territories.

The walking speed of each pedestrian was measured at 1 s intervals and then the average walking speed was computed. The mean of the average walking speed was 1.22 m/s, as shown in Table 2, which is similar to the findings of Lam et al. (2002) on pedestrian walking speeds at crosswalks in commercial areas in Hong Kong (75.38 m/min, i.e. 1.26 m/s). However, according to the *Transport Planning and Design Manual* (Transport Department, 2001), an assumed walking speed of 1.2 m/s is generally used to determine the flashing green period for the distance between the curbs in Hong Kong, although a walking speed of 0.9 m/s may be considered in exceptional cases to accommodate the elderly, people with disabilities, or exceptionally heavy pedestrian flows. Of the 1061 pedestrians who arrived during the flashing green signal, 668 (63.0%) walked slower than 1.2 m/s, and 377 (35.5%) walked slower than 0.9 m/s. This implies that the majority of pedestrians normally walk slower than 1.2 m/s, and that they are at risk of a vehicle accident if they do not pay attention to the duration of the flashing green signal and fail to speed up.

The total number of pedestrians in the cycle was used in the dataset instead of the average pedestrian arrival rate, as a simple number of pedestrians is more straightforward and easy to observe. Russell et al. (1976) and Reed and Sen (2005) found that pedestrians were encouraged to follow when observing someone else jaywalking. The percentage of pedestrians who jaywalked in the same cycle was used as a proxy of a situation variable to represent the follower behavior and estimate the influence of other jaywalkers on a pedestrian. In addition to the surrounding pedestrians, the average vehicle flow in a cycle and the pedestrian crossing time were used to measure the risk of vehicle-pedestrian accidents. Finally, the geometric data of the junctions and the signal phasing scheme were taken into account, as these represent the site conditions and the corresponding waiting times.

[Insert Table 2 Here]

METHODS

Basic binary logit model

The binary logit model was used to represent how the individual-specific and site-specific factors influence the pedestrians' jaywalking behavior. The response variable for the ith pedestrian $Y_i = 1$ if he/she jaywalks, and $Y_i = 0$ if he/she does not. Denote the probability of $Y_i = 1$ as π_i, then it follows a binomial distribution as

$$\text{logit}(\pi_i) = \log(\frac{\pi_i}{1-\pi_i}) = \sum_{k=1}^{p} \beta_k X_{ik} \qquad (1)$$

where X_{ik} is the k^{th} explanatory variable for the i^{th} pedestrian, and β_k (k = 1,…, p) are the regression coefficients. In using the basic binary logit model, each pedestrian was regarded as an individual observation, and parameters $\boldsymbol{\beta}$ were assumed to be constant for all individuals at all sites, i.e., a fixed-parameter model.

The same set of parameters $\boldsymbol{\beta}$ were applied to all observations at all sites. However, random variations in the effects of the explanatory variables among pedestrians and random effects across sites could have existed. Therefore, the random parameter binary logit model was used to account for the effect of the heterogeneity among pedestrians, and the random effect binary logit model was used to accommodate the unobserved heterogeneity across sites.

Random parameter binary logit model

To account for individual pedestrian's taste variations, a randomly distributed term was introduced for each coefficient, and the random parameter binary logit model was thus formulated as

$$\begin{aligned} &\text{logit}(\pi_i) = \log(\frac{\pi_i}{1-\pi_i}) = \sum_{k=1}^{p} \beta_{ik} X_{ik} \\ &\beta_{ik} = \beta_k + \mu_{ik} \end{aligned} \qquad (2)$$

where β_{ik} is the coefficient of the k^{th} explanatory variable for the i^{th} pedestrian, and μ_{ik} is normally distributed with a mean of 0 and variance σ_k^2. In general practice, a random parameter β_{ik} is introduced if the corresponding standard deviation σ_k is significantly larger than 0, otherwise, a fixed coefficient β_k is used for the corresponding explanatory variable X_{ik}.

Random effect binary logit model

The pedestrian movements were captured from 7 crosswalks with different characteristics in Hong Kong. Therefore, observations in the same site were grouped as panel data, and a random effect binary logit model was used to account for both the within-site correlations and the inter-site heterogeneity. Hence, the random effect binary logit model is as follows:

$$\text{logit}(\pi_{ij}) = \log(\frac{\pi_{ij}}{1-\pi_{ij}}) = \sum_{k=1}^{p} \beta_k X_{ijk} + \mu_j \qquad (3)$$

where π_{ij} is the probability that the i^{th} pedestrian at the j^{th} crosswalk jaywalks, X_{ijk} is the k^{th} explanatory variable for the i^{th} pedestrian at j^{th} crosswalk, and μ_j is the random intercept with a mean of 0 and variance σ_j^2. Hence, the random effects μ_j vary across different crosswalks but remain constant for all of the pedestrians at the same crosswalk.

Goodness-of-fit

The Akaike information criterion (AIC) is widely applied to evaluate the quality of models for a given set of data. Although the likelihood values of models can always be improved by adding predictors, a penalty term for the number of estimated parameters is introduced to deal with the trade-off between the goodness of fit and the model complexity. The formula for AIC is given as

$$\mathrm{AIC} = 2K - 2\ln(L) \tag{4}$$

where K is the number of estimated parameters in the model and L is the maximum likelihood of the given set of data for the model. Therefore, the model with a lower AIC value is considered to be a better statistical fit.

To further evaluate the overall fit of the model, McFadden's adjusted pseudo R^2 is used to compare the log-likelihood value of the model at convergence with that of the model with all parameters set to zero. The formula for the index is

$$R^2 = 1 - \frac{LL(\boldsymbol{\beta}) - K}{LL(\mathbf{0})} \tag{5}$$

where $LL(\boldsymbol{\beta})$ and $LL(\mathbf{0})$ are the log-likelihood values of the proposed and null models, respectively. The value of the index varies between 0 for no fit and 1 for a perfect fit. In practice, a value of around 0.4 is generally considered to be an excellent fit (Ortuzar and Willumsen, 2011).

Finally, the likelihood-ratio test can be used to compare the goodness-of-fit of two competing models and to decide whether the null model should be rejected in favor of the alternative model. The test statistic is defined as twice the difference between the log-likelihoods:

$$D = -2\left[LL(\boldsymbol{\beta}_{null}) - LL(\boldsymbol{\beta}_{alternative})\right] \tag{6}$$

Comparing the chi-square distribution with degrees of freedom $K_{alternative} - K_{null}$, the null model can be rejected if the value exceeds the critical value at the 95% confidence level.

In this study, the likelihood-ratio test was first conducted to compare the basic binary logit model with the random effect binary logit model. Then, a second likelihood-ratio test was conducted between the random effect binary logit model and the random parameter binary logit model. The degree of freedom had to be 1 for the first test, because the number of parameters in the random effect binary logit model was one more than that of the basic model. However, the degrees of freedom for the second test were dependent on the number of random parameters in the random parameter binary logit model.

RESULTS

STATA 13 was used to estimate the three binary logit models. Before the models were finalized, a Pearson's correlation test was conducted to identify the explanatory variables that were independent of each other and to eliminate the highly correlated variables to ensure an unbiased estimation (Table 3). Not surprisingly, the correlation analysis indicated that the two dummy variables of gender, M and F, were highly correlated. This meant that either of the two dummy variables could be included in the model to represent pedestrian gender. Average vehicle flow was found to be highly correlated with cycle time and pedestrian red signal time, as the longer the time for vehicles, the larger the average vehicle flow. High correlations also existed among the geometric design variables and the signal phasing variables, including the number of lanes at the crosswalk, the numbers of approaches and approach lanes at the junction, the number of traffic streams at the junction, the number of signal stages, the cycle time, and the red signal time. These are all related to the size of the junction; i.e., the larger the junction, the longer the time to clear the vehicle traffic, and hence the longer the red signal time for pedestrians and the longer the cycle time. Only one or two of these variables can be included in the model.

[Insert Table 3 Here]

Finally, 8 explanatory variables (gender, signal at arrival, walking speed, number of pedestrian in the cycle, percentage of pedestrian jaywalking in the cycle, average vehicle flow in the cycle, crossing time, and number of stage) were included in the model. The estimation results and the average marginal effects for the basic (fixed parameter), random effect, and random parameter binary logit models are shown in Tables 4 and 5, respectively. The parameter estimates of all modeling approaches are significant at the 5% level. The signs of all parameters are consistent across the three models.

In terms of goodness-of-fit, all three models have acceptable overall fit with the McFadden's adjusted pseudo R2 values in the 0.26 ~ 0.29 range. Both the random effect and random parameter binary logit models have lower AIC values and larger values of McFadden's adjusted pseudo R2 than the basic binary logit model. Unobserved heterogeneities thus exist across sites and also among pedestrians, and hence the two models provide statistically superior fit compared to the basic binary logit model. The statistic of the likelihood-ratio test between the basic model and the random effect models is 240.66, which is much greater than $\chi 2$ (1, 99%) = 6.64, i.e., the basic model is rejected in favor of the random effect model. Similarly, the statistic of the likelihood-ratio test between the random effect model and the random parameter model is 103.50, which again is much larger than $\chi 2$ (3, 99%) = 11.34. This result shows that the random parameter model is statistically superior to the random effect model. Therefore, we mainly focus on the latter model in the following section.

DISCUSSION

In the random parameter model, 4 of the 8 variables (gender, walking speed, percentage of pedestrians jaywalking, and crossing time) produced statistically significant random parameters (all were normally distributed). Table 4 shows that the gender variable (M: 0, F: 1) resulted in a random parameter with a mean of − 0.360 and a standard deviation of 0.156 (98.95 % of the distribution is negative). This suggests that male pedestrians were less patient and more likely to jaywalk than female pedestrians, which is in line with the findings of most previous studies (Tiwari et al., 2007, Rosenbloom, 2009, Brosseau et al., 2013). The average marginal effect shows that female pedestrians are 5% less likely to jaywalk.

The signal at arrival resulted in a fixed parameter and was found to significantly determine the probability of jaywalking. According to the marginal effects in Table 5, the pedestrians who arrived at the crosswalk during the red signal were 33.9% less likely to jaywalk. This suggests that those who arrived at the crosswalks during the flashing green periods probably seized the remaining time before the vehicle traffic discharged, and directly walked across to avoid waiting for one more cycle time.

The average walking speed resulted in a random parameter with a mean of 3.251 and a standard deviation of 0.978 (nearly 100% of the distribution is greater than 0), which implies that there was considerable variation in the effect of walking speed. However, jaywalking pedestrians were found to walk faster, as they had to seize gaps in the traffic flow when crossing to avoid having accidents. The average walking speed was 1.22 m/s, as reported in Table 2, so the marginal effect (0.349 in the random parameter model) can be interpreted as indicating that a 0.1 m/s increase in walking speed resulted in a 3.49% increase in the probability of jaywalking.

In addition to the abovementioned individual-specific factors, individual pedestrians were likely to be influenced by surrounding pedestrians who arrived during the same cycle. The results in Table 4 indicate that both a larger number of pedestrians in the cycle and a higher percentage of those jaywalking in the cycle increased the probability that a particular pedestrian would jaywalk. The number of pedestrians in the cycle resulted in a fixed parameter of 0.005, and the percentage of pedestrians jaywalking in the cycle resulted in a random parameter with a mean of 5.276 and a standard deviation of 0.964 (nearly 100% of the distribution is greater than 0). The marginal effects of the random parameter model (0.001 for the total number of pedestrians and 0.567 for the percentage of pedestrians jaywalking) indicated that 1 additional jaywalking pedestrian resulted in a much greater increase in the probability of a particular pedestrian jaywalking than simply one more pedestrian in the same cycle. The two parameter estimates imply that the more pedestrians in a cycle, the greater the likelihood an individual will jaywalk, and the other pedestrians would then be encouraged by the first rule breaker and proceed to jaywalk. This result is the opposite of Rosenbloom's (2009) finding that the tendency to cross on a red signal is lower when there are more people waiting at the curb, due to the power of social control.

[Insert Table 4 Here]

[Insert Table 5 Here]

Pedestrians also typically observe and assess the site conditions. The average vehicle flow resulted in a fixed parameter of −0.025, indicating that a higher average vehicle flow decreased the probability of jaywalking, as the higher the vehicle flow, the shorter the gaps between vehicles, and hence the higher the risk of an accident. Crossing time was also found to be crucial in determining the probability of jaywalking and resulted in a random parameter with a mean of 0.194 and a standard deviation of 0.054 (nearly 100% of the distribution is greater than 0). The marginal effect (0.021 in the random parameter model) implies that a second increase in crossing time resulted in a 2% increase in the probability of jaywalking. The number of stages resulted in a fixed parameter of 1.734. The marginal effect (0.186 in the random parameter model) implies that one additional stage of the signal scheme resulted in an 18.6% increase in the probability of jaywalking. The results of both crossing time and number of stages indicated that pedestrians may be more likely to jaywalk at larger signalized intersections with longer kerb-to-kerb distance and more signal stages.

CONCLUSION

This study investigated the contributory factors of pedestrians' jaywalking behavior at signalized crosswalks. The crossing movements of 7230 pedestrians were captured at 7 crosswalks in Hong Kong. The information on the pedestrian behavior, the vehicle traffic flow, and the site-specific factors were incorporated into our proposed binary logit models to determine the probability of pedestrian jaywalking. To address the heterogeneity issues, the random parameter model was used to accommodate the variation in the effects of the explanatory variables among pedestrians, while the random effect model was used to account for the unobserved heterogeneity across sites.

The random parameter model was found to be more suitable for addressing the heterogeneous effects of the explanatory variables among pedestrians. The pedestrian characteristics (gender, walking speed), the behavior of surrounding pedestrians (total number of pedestrians and the proportion of jaywalkers), the vehicle traffic, the timing of arrival and the length of signal, and the location of the crosswalk were found to significantly determine the probability of pedestrian jaywalking. The results imply that pedestrians with superior physical ability are generally less patient and more likely to take the risk of jaywalking.

The results also revealed some critical issues relating to the current policies and design of signalized pedestrian crosswalks in Hong Kong. The significance of the flashing green signal is ambiguous to some pedestrians, as it seems that most pedestrians are not aware that starting to cross during the flashing green signal period is also illegal. Because Hong Kong is a densely populated city, it would be well worth considering providing more informative signals rather than simply promoting the regulation. Furthermore, it was also found that the majority of pedestrians normally walked slower than 1.2 m/s, which is the speed commonly used in the design of signalized pedestrian crossings to determine the length of the flashing green signal. This may lead pedestrians to overestimate the remaining time before the vehicle traffic streams discharge, and hence rather take the risk of jaywalking than wait for the length of another cycle. A possible measure that policy makers could consider is to introduce a signal countdown with the conventional graphic signal, which has been shown to significantly increase the proportion of pedestrians who start to cross during the green signal (Keegan and O'Mahony, 2003) and to effectively enhance pedestrian safety (Schattler et al., 2002). The text “Don’t walk/Walk” may also be considered to give clear instructions to pedestrians.

Overall, our findings show that pedestrian crossing behavior is dependent on individual-specific factors and site-specific factors. In the future, observational surveys conducted at more sites with different geometric features and signal phasing schemes would enable further insights to be obtained on the effects of site-specific factors. Other environmental factors, including weather, temperature, noise, and type of land use, would be well worth investigating with a more comprehensive dataset.

ACKNOWLEGEMENTS

The research described here was supported by a Research Postgraduate Scholarship and grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project Nos.: 717512, 524313).

REFERENCES

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.

Dıaz, E. M. (2002). "Theory of planned behavior and pedestrians' intentions to violate traffic regulations." *Transportation Research Part F: Traffic Psychology and Behaviour*, 5(3), 169-175.

Guo, H., Gao, Z., Yang, X., & Jiang, X. (2011). "Modeling pedestrian violation behavior at signalized crosswalks in China: A hazards-based duration approach." *Traffic Injury Prevention*, 12(1), 96-103.

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(4), 1097-1106.

Keegan, O., & O'Mahony, M. (2003). "Modifying pedestrian behaviour." *Transportation Research Part A: Policy and Practice*, 37(10), 889-901.

Koh, P. P., & Wong, Y. D. (2014). "Gap acceptance of violators at signalised pedestrian crossings." *Accident Analysis & Prevention*, 62, 178-185.

Kruszyna, M., & Rychlewski, J. (2013). "Influence of approaching tram on behaviour of pedestrians in signalised crosswalks in Poland." *Accident Analysis & Prevention*, 55, 185-191.

Lam, W. H., Lee, J. Y., & Cheung, C. Y. (2002). "A study of the bi-directional pedestrian flow characteristics at Hong Kong signalized crosswalk facilities." *Transportation*, 29(2), 169-192.

Lavalette, B. C., Tijus, C., Poitrenaud, S., Leproux, C., Bergeron, J., & Thouez, J. P. (2009). "Pedestrian crossing decision-making: A situational and behavioral approach." *Safety Science*, 47(9), 1248-1253.

Li, B. (2013). "A model of pedestrians' intended waiting times for street crossings at signalized intersections." *Transportation Research Part B: Methodological*, 51, 17-28.

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(4), 1156-1163.

Ortuzar, J., & Willumsen, L. G. (2011). *Modelling Transport*. John Wiley & Sons.

Oxley, J., Fildes, B., Ihsen, E., Charlton, J., & Day, R. (1997). "Differences in traffic judgements between young and old adult pedestrians." *Accident Analysis & Prevention*, *29*(6), 839-847.

Oxley, J. A., Ihsen, E., Fildes, B. N., Charlton, J. L., & Day, R. H. (2005). "Crossing roads safely: An experimental study of age differences in gap selection by pedestrians." *Accident Analysis & Prevention*, *37*(5), 962-971.

Reed, R., & Sen, S. (2005). Racial differences and pedestrian safety: Some evidence from Maryland and implications for policy. *Journal of Public Transportation*, *8*(2), 3.

Ren, G., Zhou, Z., Wang, W., Zhang, Y., & Wang, W. (2011). "Crossing behaviors of pedestrians at signalized intersections: Observational study and survey in China." *Transportation Research Record: Journal of the Transportation Research Board*, (2264), 65-73.

Rosenbloom, T. (2009). "Crossing at a red light: Behaviour of individuals and groups." *Transportation Research Part F: Traffic Psychology and Behaviour*, 12(5), 389-394.

Russell, J. C., Wilson, D. O., & Jenkins, J. F. (1976). "Informational properties of jaywalking models as determinants of imitated jaywalking: An extension to model sex, race and number." *Sociometry*, 270-273.

Schattler, K. L., Wakim, J. G., Datta, T. K., & McAvoy, D. (2007). "Evaluation of Pedestrian and Driver Behaviors at Countdown Pedestrian Signals in Peoria, Illinois." *Transportation Research Record: Journal of the Transportation Research Board*, 2002(1), 98-106.

Tiwari, G., Bangdiwala, S., Saraswat, A., & Gaurav, S. (2007). "Survival analysis: Pedestrian risk exposure at signalized intersections." *Transportation Research Part F: Traffic Psychology and Behaviour*, 10(2), 77-89.

Transport Department, HKSAR, China (2001). *Transport Planning and Design Manual*.

Zhou, R., Horrey, W. J., & Yu, R. (2009). "The effect of conformity tendency on pedestrians' road-crossing intentions in China: An application of the theory of planned behavior." *Accident Analysis & Prevention*, 41(3), 491-497.

Zhuang, X., & Wu, C. (2011). "Pedestrians' crossing behaviors and safety at unmarked roadway in China." *Accident Analysis & Prevention*, 43(6), 1927-1936.

Table 1 Site locations

Area	District (Land use)	Junction	No. of Obs.	Signal cycle time (s)	Average pedestrian arrival rate (ped/min)
West Island	Central (Commercial)	Queen's Rd. Central* Pedder St.	1832	120	31.0
East Island	Causeway Bay (Commercial)	Morrison Hill Rd.* Leighton Rd.	1984	120	25.2
West Kowloon	Jordan (Commercial)	Jordan Rd.* Nathan Rd.	1142	130	73.5
East Kowloon	Hung Hom (Residential)	Hung Lok Rd.* Hung Lai Rd.	106	90	2.7
West New Territories	Tsuen Wan (Commercial /Residential)	Sha Tsui Rd.* Chung On St.	1142	95	79.7
Middle New Territories	Sha Tin (Industrial)	Ngan Shing St.* Siu Lek Yuen Rd.	128	110	5.1
East New Territories	Tseung Kwan O (Residential)	King Ling St.* Choi Ming St.	896	110	8.6

**The selected crosswalk*

Table 2 Summary of data

Categorical variables	Attributes	Count (Proportion)	
Jaywalking	Yes:1	4586 (63.4%)	
	No: 0	2644 (36.6%)	
Gender	Male: (1, 0)	3357 (46.4%)	
Represented by two dummy variables (M, F)	Female: (0, 1)	3265 (45.2%)	
	Unidentified: (0, 0)	608 (8.4%)	
Age	Adults: 0	6948 (96.1%)	
	Kids: 1	140 (1.9%)	
	Elderly: 2	142 (2.0%)	
Signal at arrival	Flashing green: 0	1061 (14.7%)	
	Red: 1	6169 (85.3%)	
District	Urban: 1	5064 (70.0%)	
	New Territories (NT): 0	2166 (30.0%)	
Continuous variables	**Range**	**Mean**	**S.D.**
Walking speed (m/s)	Min: 0.16; Max: 4.56	1.22	0.41
Total number of pedestrians in the cycle	Min: 1; Max: 207	84.44	58.26
Percentage jaywalking in the cycle	Min: 0; Max: 1	0.42	0.19
Average vehicle flow in the cycle (veh/min)	Min: 0.6; Max: 20.4	10.80	4.10
Crossing time (s)	Min:2; Max: 85	10.5	7.98
Geometric design			
Number of lanes at the crosswalk	Min: 1; Max: 6	3.14	1.42
Number of approaches at the junction	Min: 1; Max: 4	2.79	1.13
Number of approach lanes at the junction	Min: 3; Max: 13	8.51	3.94
Number of traffic streams at the junction	Min: 1; Max: 9	4.50	2.47
Signal phasing scheme			
Number of signal stages	Min: 2; Max: 4	3.34	0.85
Cycle time (s)	Min: 90; Max: 130	119.72	6.56
Pedestrian red signal time (s)	Min: 67; Max: 100	93.90	8.61

Table 3. Pearson correlation test of variable

	M	F	Age	Signal at arrival	District	Walking speed	No. of pedestrians in the cycle	Percentage of jaywalking	Average vehicle flow	Crossing time	No. of lanes at the crosswalk	No. of approaches at the junction	No. of approach lanes at the junction	No. of traffic streams at the junction	No. of signal stage	Cycle time	Pedestrian red signal time
M	1.00																
F	-0.84	1.00															
Age	0.04	-0.03	1.00														
Signal at arrival	0.04	0.00	0.02	1.00													
District	0.00	0.06	-0.03	0.01	1.00												
Walking speed	0.16	-0.06	-0.07	0.08	0.19	1.00											
No. of pedestrians in the cycle	-0.14	-0.08	0.09	-0.07	-0.10	-0.46	1.00										
Percentage of jaywalking	0.07	0.05	-0.05	-0.02	0.07	0.15	-0.51	1.00									
Average vehicle flow	-0.07	-0.03	0.04	0.03	0.32	0.00	0.45	-0.40	1.00								
Crossing time	-0.10	-0.03	0.10	-0.01	0.08	-0.54	0.49	-0.30	0.13	1.00							
No. of lanes at the crosswalk	-0.02	-0.07	0.05	0.02	0.14	-0.11	0.30	-0.30	0.11	0.78	1.00						
No. of approaches at the junction	0.01	-0.11	0.07	0.06	-0.36	-0.04	0.19	-0.35	0.03	0.49	0.69	1.00					
No. of approach lanes at the junction	0.07	-0.07	0.04	0.14	0.11	0.32	-0.11	-0.26	0.24	0.26	0.57	0.76	1.00				
No. of traffic streams at the junction	0.05	-0.09	0.03	0.07	-0.43	0.09	-0.10	-0.18	-0.19	0.35	0.62	0.93	0.73	1.00			
No. of signal stage	0.05	-0.09	0.05	0.11	-0.51	0.19	-0.05	-0.26	0.09	-0.05	0.10	0.74	0.71	0.73	1.00		
Cycle time	-0.08	-0.04	0.06	0.00	0.44	-0.18	0.67	-0.39	0.60	0.51	0.49	-0.02	0.10	-0.24	-0.31	1.00	
Pedestrian red signal time	-0.08	0.02	0.02	-0.02	0.61	-0.12	0.51	-0.21	0.60	0.13	-0.01	-0.50	-0.24	-0.71	-0.55	0.84	1.00

Table 4 Estimates and goodness-of-fit for the basic, random effect, and random parameter binary logit models

	Basic	Random Effect	Random Parameter
Variables			
Gender (M:0, F:1)	− 0.408*	− 0.405*	− 0.360*
s.d. Gender			*0.156**
Signal at arrival (Flashing green:0, Red:1)	− 5.266*	− 8.700*	− 12.905*
Walking speed (m/s)	1.267*	2.654*	3.251*
s.d. Walking speed			*0.978**
No. of pedestrians in the cycle	0.161*	0.005*	0.005*
Percentage of jaywalking	4.890*	4.580*	5.276*
s.d. Percentage of jaywalking			*0.964**
Average vehicle flow (veh/min)	− 0.094*	− 0.029*	− 0.025*
Crossing time (s)	0.047*	0.110*	0.194*
s.d. Crossing time			*0.054**
Number of stage	0.417*	0.954*	1.734*
σ_j		1.095*	
Goodness-of-fit			
No. of observations	7230	7230	7230
No. of parameters, K	8	9	12
Log likelihood at zero, $LL(0)$	− 5011.45	− 5011.45	− 5011.45
Log likelihood at convergence, $LL(\beta)$	− 3686.96	− 3566.63	− 3514.88
AIC	7389.91	7151.25	7066.47
McFadden's adjusted pseudo R^2	0.26	0.29	0.30
Likelihood-ratio test		*vs. basic model*	*vs. random effect model*
$\chi^2 = -2\left[LL(\boldsymbol{\beta}_{null}) - LL(\boldsymbol{\beta}_{alternative})\right]$		240.66	103.50
Degrees of freedom		1	3
Significance level		< 0.01	< 0.01

*Note: * = Significance at the 5% level*

Table 5 Average marginal effects for the basic, random effect, and random parameter binary logit models

Variables	Basic	Random Effect	Random Parameter
Gender (M:0, F:1)	− 0.072*	− 0.405*	− 0.039*
Signal at arrival (Flashing green:0, Red:1)	− 0.421*	− 8.699*	− 0.339*
Walking speed (m/s)	0.221*	2.654*	0.349*
No. of pedestrians in the cycle	0.003*	0.005*	0.001*
Percentage of jaywalking	0.839*	4.580*	0.567*
Average vehicle flow (veh/min)	− 0.016*	− 0.029*	− 0.003*
Crossing time (s)	0.008*	0.110*	0.021*
Number of stage	0.073*	0.954*	0.186*

*Note: * = Significance at the 5% level*