

Received June 14, 2019, accepted June 24, 2019, date of publication July 11, 2019, date of current version August 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2928038

Designs for Safer Signal-Controlled Intersections by Statistical Analysis of Accident Data at Accident Blacksites

C. K. WONG 

Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong

e-mail: wongck@cityu.edu.hk

This work was supported by the GRF Funding Scheme from the Research Grants Council of the Hong Kong Special Administrative Region, China, under Project CityU 11213514.

ABSTRACT This paper describes the collection and statistical analysis of accident counts and intersection layout geometries at a range of signal-controlled intersections, with the aim of improving safety at these sites. Negative binomial regression analysis is conducted to relate the accident count data as a dependent variable, with various independent variables to capture the intersection layout and lane-marking patterns. Statistically significant variables are identified, and their individual effects on accident counts are analyzed. Although the accident-prediction models for signalized intersections have been extensively investigated, this paper also considers the effects of shared lane markings, which is a new approach. The results of this paper show that the shared lane markings are indeed a statistically significant predictor of the number of accidents. It was found that the accident counts at signal-controlled intersections could be reduced by altering the lane-marking patterns using a combination of well-established lane-based design methods and new governing constraint sets to enhance the safety controls for turning traffic derived from our statistical analysis. These new lane-marking patterns also satisfy engineering performance requirements. The intersections in Hong Kong were investigated as illustrative case studies, and the numerical results show a substantial decrease in the predicted accident counts, with an acceptable tradeoff in the reduction of overall intersection capacity.

INDEX TERMS Accident prevention, road accidents, statistical analysis, traffic signal control.

I. INTRODUCTION

According to an annual report published by the World Health Organization (WHO), road traffic injuries and deaths have significant effects on public healthcare systems and financial burdens on local governments for medical expenditures. According to 2018 figures, about 1.3 million people are killed and another 50 million injured annually worldwide due to road traffic crashes [1]. The published statistics in Hong Kong show that 15,725 traffic accidents in 2017 led to 19,888 casualties [2]. Of these accidents, 4205 (approximately 27%) occurred at or near a road intersection, and 3190 casualties (approximately 16%) involved pedestrians. According to the Hong Kong Police Force, the rates of accidents due to careless lane-changing and improper/illegal turns at intersections have

remained quite steady, at approximately 1100 and 800 per year, respectively.

As reported by Hao et al. (2018) [3], signalized intersections are conventionally designed to satisfy basic safety requirements, but their main focus is to minimize total traffic delays or maximize intersection throughputs with sophisticated optimization algorithms. However, it now seems that greater design standards should be introduced to improve safety at signalized intersections.

Ngo et al. (2019) investigated accidents related to the yellow-light dilemma and found that vehicle ad-hoc networks (VANETs) facilitated better communication between traffic lights and vehicles, which effectively reduced traffic accidents due to driver error in running a yellow light [4]. Li and Sun developed a multiobjective optimization problem (MOP) to consider four measures to improve network traffic performance, including maximizing system throughput, minimizing total delays, improving safety, and

The associate editor coordinating the review of this manuscript and approving it for publication was Zhengbing He.

preventing spillover [5]. This complex high-dimension problem was solved by genetic algorithms. A similar MOP approach was developed to optimize turning-lane assignment based on microscopic traffic simulation and a cell-mapping method [6].

Thus, there is clearly scope for optimization of lane markings in approach lanes, which indicate the permitted turns in the intersection lane, to increase safety. The aims of the present study were (1) to conduct a statistical analysis to establish relationships between the observed accident counts at signalized intersections and their respective 'lane usage' settings, (2) to identify statistically significant design settings to establish the lane-usage patterns needed to reduce accident risks, (3) to examine statistically significant variables to determine their contribution to accident counts, and (4) to revise existing lane-usage patterns and intersection layout settings based on the statistical analysis findings to propose safer designs for signalized intersections.

II. RELATED WORKS

Records of crash counts, crash rates, and crash injury severity have been analyzed in a variety of ways, including probabilistic, statistical, and other scientific methods. For instance, the severity levels of traffic accidents were analyzed using conditional probabilities, *t*-tests, and chi-square tests [7], [8]. A cross-tabulation method was applied [9] to assess distributional differences between injury levels and the sex and age of the victim groups. Accidents were also analyzed according to the location (at highway mid-block or at intersection), time (day or night), and type (single/multiple car or rear-end/head-on collision). The aim was to identify which driver-groups had a greater risk of involvement in severe accidents. Significant risk factors for injuries or fatalities at various locations were also found.

Al-Ghamdi (2003) collected and categorized data on 1774 accidents reported in Saudi Arabia and found that improper driving behaviors and red-light jumping were the major causes of accidents at urban intersections [10]. Mane and Pulugurtha (2018) statistically analyzed red-light violation crashes at signalized intersections in North Carolina and related these data to local land-use and demographic characteristics [11]. Shesterov and Mikhailov (2017) predicted accident rates at signalized intersections in St. Petersburg, which sees an average of 2.54 accidents for every million vehicles that cross intersections, based on total accident numbers and traffic flow-count data [12]. Crash-prediction models for predicting crashes on highway facilities were published by Oh *et al.* [13].

The causes of severe traffic accidents can be complicated by multiple contributing factors, which means that analyzing a single risk factor in isolation from others is likely insufficient to fully determine the causes of traffic accidents. Relationships between accident counts and contributing factors could be established by multivariate regression analysis. Thus, crash counts or rates could be set as dependent variables in regression analysis to relate these to

various contributing factors as independent variables. Tarko and Tracz (1995) developed a general regression-type accident-prediction model that uses traffic volume, pedestrian volume, and the green-light time proportion ratio (to cycle length) as independent variables to quantify traffic and signal impacts on pedestrian safety [14]. All parameters related to traffic and pedestrian volume showed a positive relationship with accident frequency. Greater model accuracy with a lower deviance was achieved by calibrating a higher slope parameter, which the authors inferred to mean that parameters related to traffic and pedestrian volume alone may be insufficient to explain changes in accident frequencies in their study. Greibe (2003) developed linear accident-prediction models mainly for 'black spot sites' and found that the variable most strongly correlated to accident numbers for both road links and junctions was motor vehicle flow [15]. Karlaftis and Golias (2002) applied hierarchical tree-based regression analysis to establish an accident prediction model for rural roadways and found that geometric design and pavement condition variables were the two most important factors correlated with the observed accident numbers in their data [16]. Elvik (2016) explored associations between various risk factors and accident occurrence and investigated the stabilities of these relationships over time [17]. Simple linear and nonlinear regression models were developed based on observed accident counts during different (1-year) time periods, and the model accuracies were mainly assessed by R-square values. In addition, Elvik fitted prediction models for junction accidents in Denmark using the data for daily traffic flows along major and minor roads.

Chen *et al* used severity-based analysis to classify crash outcomes in terms of accident severity [18]. This study also identified the traffic and highway design parameters that had a significant association with severe crashes on segments of multilane arterial highways. Nonsevere crashes were excluded in the study to avoid under-reporting biases. Severe and non-crash cases were modeled in a binary mode, with a target variable of '1' for severe crashes and '0' for non-crash cases. Driver age, sex, speed zone, traffic control type, time of day, crash type, and seat belt usage were identified with a logistic regression model to affect the accident severity of intersection crashes [18]. Yau (2004) used multivariate stepwise logistic regression to identify significant factors that contribute to the severity of traffic accidents in Hong Kong [19]. The site location, age and sex of the drivers, vehicle age, environmental factors, and congestion levels all showed a statistically significant relationship to single-vehicle accidents of private, goods, and motorcycle vehicles. Polders *et al.* (2015) studied signalized intersection safety by identifying crash types, locations, and factors associated with signalized intersections and used logistic regression modeling techniques to analyze 87 signalized intersections [20].

Accident severity may be represented by binary, ordinal, or multinomial target variables [21]. Patterns of severe crashes on multilane arterial segments were studied by Pande

and Abdel-Aty [22], and crashes involving lane-changes, rear-ending, and pedestrians, and single-vehicle/off-road related crashes were modeled at mid-block segments. Celik and Oktay (2014) applied a multinomial logit analysis to identify risk factors that influenced injury severity in road traffic accidents in Turkey [23] and found that a variety of risk factors had either a positive or a negative effect on the likelihood of injury severity, and the average pseudo-elasticities suggested that the installation of traffic lights would reduce the probability of fatal injuries by 41.8%.

Accident count data are generally non-negative and discrete in nature. Therefore applying ordinary (least-squares) regression analysis that assumes dependent variables to be continuous may be inappropriate. Poisson regression modeling could be a simple and direct method for accident data analysis, and Anjana and Anjaneyulu (2015) used a hierarchical Poisson regression model to identify major contributing factors to traffic accidents [24]. Again, the traffic flow volume, median width of approach lanes, and deviation in green time were found to be significant contributing factors. However, Lord and Mannering (2010) noted that the Poisson regression model cannot model raw data with over- and under-dispersion [25], which means that analysis results could be biased, especially with a small sample size.

To overcome potential over-dispersion problems in observed raw data (i.e., mean \ll variance), an extension for the Poisson-type models was proposed in which accident counts could be estimated using negative binomial regression models [26]–[28]. For example, accident prediction models for intersections were developed in the United States based on negative binomial regression [29]. The negative binomial regression model was shown to give better prediction and goodness-of-fit for accident data in Singapore in which traffic volumes, numbers of phases per cycle, uncontrolled left-turn lanes, and the presence of surveillance cameras were highly statistical significant factors in the traffic-accident occurrence [30]. Chen and Xie (2016) established statistical relationships to relate annual average daily traffic (AADT) and the numbers of multiple-vehicle crashes using generalized additive models and piecewise linear negative binomial regression models [31]. A negative binomial distribution model was applied to fit signalized intersection data to analyze red-light violation crashes [11]. In this study, predictor variables with a statistically significant relationship to the numbers of crash accidents with positive effects included traffic volume and the numbers of lanes on the minor approaches, and those with negative effects included industrial and/or large lots of residential land.

A zero-inflated negative binomial model was applied to predict vehicle crashes by drivers' characteristics and their records of violating traffic ordinances. Statistically significant factors that contributed to a higher likelihood of vehicle crashes included a record of violating regulations (such as drink-driving, hit-and-run, and speeding), driver age, sex, driving experience, and vehicle size [32]. Wang et al. (2016) developed and compared Poisson regression and negative

binomial regression to identify critical crash-prediction variables from 36 safety-related parameters, based on 5-year crash counts [33]. Poisson regression and negative binomial regression models were evaluated to establish relationships between accidents and road geometry designs [34]. Wong et al. (2007) studied some 262 signalized intersections in Hong Kong to relate accident numbers to various road environment factors, traffic-flow data, fixed geometric parameters, and control methods using Poisson and negative binomial regression models [35]. Finally, Lord et al. (2007) compared the performance of Poisson, negative binomial, and zero-inflated regression models in this context [36].

Most of these traffic accident studies have found accident rates or actual accident counts to be related to one or more of the same set of statistically significant factors that are difficult to control in daily operations (such as the traffic flow intensities) or that are unable to be varied in the design calculations (such as lane widths, turning radius, and total numbers of traffic lanes for approaching and departing traffic: these are already at their maximum values).

In this study, we conduct a new statistical analysis to relate lane-usage patterns at signalized intersections to observed accident counts. Based on regression-type analysis results, we attempt to modify the intersection layouts to reduce the (positively or negatively) statistically significant risky usages of lanes that are related to accidents. In the literature review, a basic statistical study was conducted to find contributory factors for crashes at signalized intersections. Geometric factors such as turning radius and average lane width were found to be statistically significant influences on the accident rates in their study. However, these two geometric factors are generally fixed and limited by the actual on-site dimensions of the intersections and are difficult to alter for accident prevention [35]. It follows that from an engineering design perspective, if design variables and relevant geometric characteristics at signalized intersections that can be varied or controlled in design stages could be included in a statistical analysis, the analysis results could be useful to guide improved design for future, safer signalized intersections.

To this end, in this study we extract various geometric and lane-based (design) parameters at intersections to establish the statistical relationships of these parameters with respect to the observed accident counts. The analysis will enable us to study the effects of individual design variables with either positive or negative relationships to accident counts. These findings could allow the refinement of intersection layouts to yield safer signalized intersections designs for practical implementation.

III. PROBLEM FORMULATION

A. BACKGROUND

Current signalized intersection designs usually optimize engineering performance by minimizing total vehicular delays or maximizing the reserve capacity (throughput) [37]–[40]. Basic safety measures in conventional design methods include implementation of (i) minimum green

light durations, (ii) minimum red light durations, and (iii) minimum clearance times to prevent front bumper–rear bumper collisions and vehicle clashes resulting from incompatible vehicular and pedestrian movements inside common areas at intersections [41].

It has traditionally been considered that basic safety requirements are sufficiently satisfied by separating the rights-of-way of conflicting traffic movements by displaying proper traffic signals to road users. More recently, a design framework for signalized intersections has been advanced to optimize individual lane usage by painting arrows on the road to show road users their permitted turning movements in various approach lanes. Shared-lane markings could be designed to permit more than one turning direction in an approach traffic lane. Overall intersection performance could be improved because shared lane markings, if designed and implemented properly, can effectively distribute demand-for-turning flows across adjacent traffic lanes, thus generating balanced lane-flow patterns.

However, in existing designs, where and how to establish shared lane markings (i.e., in which approach lanes and for which turns) and the total number of approach lanes that have shared-lane markings are not well controlled. Moreover, the potential accident risks of using single- or shared-lane markings at signalized intersections have not been investigated. Therefore, a better understanding of their relationships could be achieved by conducting a statistical analysis to relate the observed accident numbers to the existing intersection geometries and lane-marking settings (in which lane marking patterns could be varied at design calculation stages). It is also expected that better intersection layout arrangements with safer lane-marking patterns could be optimized by introducing new safety constraint sets in the mathematical design framework.

Figure 1 presents a typical layout of a 4-arm signalized intersection. Different numbers of approach and exit lanes with various lane widths could be available for different arms where various lane marking arrows can be painted on the road to guide drivers when turning at an intersection. In Figure 1, approach lanes with numbers 2, 3, 7, and 8 are the nearside lanes (i.e. next to the pavement kerb) and other approach lanes with numbers 1, 4, 5, and 6 are non-nearside lanes. Figure 2 shows realistic traffic arms from various signalized intersections consisting of approach and exit lanes. Lane markings for various turns and turn combinations are painted on the approach lanes. A shared-lane marking is highlighted to permit both left-turn and straight-ahead movements. Practically, shared-lane markings can also be designed to permit up to three turning movements on an approach lane, as shown in Figure 2. Shared-lane markings can be designed and implemented on either nearside or non-nearside lanes in different arms. At-grade pedestrian crossings can be used that are perpendicular to approach and exit lanes and are represented as yellow strips painted on the ground. For other busy intersections, pedestrian crossings could be disabled so that longer green-times could be designed for heavy vehicular

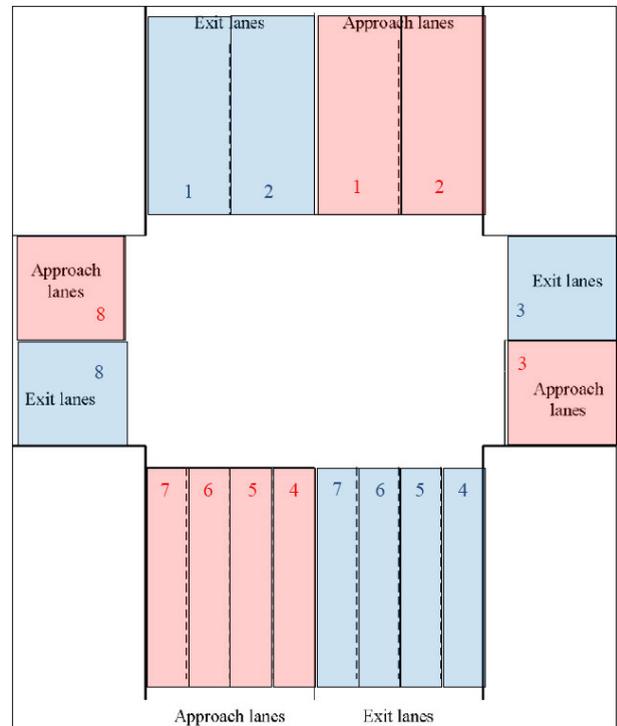


FIGURE 1. A typical 4-arm intersection with approach and exit lanes (left-hand side traffic).



FIGURE 2. Practical settings at signalized intersections in Hong Kong with different lane marking patterns and pedestrian crossing arrangements.

movements. From the design perspective, lane-marking patterns and approach lane arrangements could be varied for practical implementation to achieve different levels of safety for pedestrians and motorists using signalized intersections.

Having illustrated the possible intersection settings and actual lane-marking patterns that can be implemented practically at signalized intersections, a vector (which represents the intersection number) can be formed that consists of a set of independent (or predictor) variables for the proposed negative binomial regression analysis. This vector captures the geometric characteristics and traffic lane usages at a signalized intersection. The signalized intersections have various numbers of approach and exit lanes, left-turn, right-turn, and straight-ahead lane markings, shared lane markings permitting two or three turning movements, shared lane markings on nearside and non-nearside lanes, and approach and exit lanes perpendicular to pedestrian crossings. These parameters can all be counted in data collection from direct observations. They are regarded as independent (predictor) variables in the proposed negative binomial regression analysis. For a study intersection s , the observed accident number can be set as a dependent variable, λ_s . Table 1 gives the details of these design parameters.

B. NEGATIVE BINOMIAL REGRESSION ANALYSIS

In this section, negative binomial regression analysis was conducted to establish the statistical relationships between observed accident counts and various lane-usage patterns for surveyed intersections. The traditional designs of signalized intersections have focused on optimizing signal timings, such as the cycle time and the start and duration of green-times of different phases, to maximize the overall intersection capacity or minimize the total delay experienced by users. These conventional stage-, group- (or phase-), and lane-based design optimization methods satisfy ‘basic’ safety requirements [37]–[40]. To enhance safety and prevent accidents at signalized intersections, it is expected that intersection geometries and lane usage patterns need to be refined.

A negative binomial regression analysis was used to relate the observed accident counts to the intersection geometries and lane-usage parameters to establish the statistical relationships between the accident counts and the geometric and lane-usage parameters. Parameters with a statistically significant influence on the accident counts were identified. Their individual effects were analyzed and used to modify intersection designs to reduce accident risks.

Accident numbers are given in the form of counts that are discrete in nature. Ordinary linear or nonlinear regression analysis for continuous dependent variables may be imperfect to model and predict these accident counts. Poisson regression and negative binomial regression are considered better ways to handle such discrete variables. One important assumption necessary for the use of Poisson regression is that the mean of the count is equal to its variance. However, this assumption may not always hold in raw count data in which the variance is always found to be larger than the mean. Negative binomial regression is an extension of the Poisson regression that allows a certain amount of overdispersion in

TABLE 1. Summary of collected data from surveyed intersections.

Dependent and 8 independent variables		Min	Max	Mean	Standard deviation
h_s	Total number of accidents observed at intersection s (2004-2014)	11	180	41.93	42.779
X_s	Number of approach lanes with left-turn lane marking	0	5	1.586	1.214
	Number of approach lanes with straight-ahead lane marking	2	9	5.707	1.816
	Number of approach lanes with right-turn lane marking	0	7	1.897	1.459
	Ratio of number of exit lanes to number of approach lanes	0.524	1.143	0.806	0.132
	Number of approach lanes with shared lane markings permitting 2 movements	0	6	1.500	1.454
	Number of approach lanes with shared lane markings permitting 3 movements	0	1	0.034	0.184
	Number of shared lane marking on nearside lane	0	4	0.776	0.956
	Number of traffic lanes perpendicular to pedestrian crossing	0	27	11.103	7.276

the raw data so that the variance (of a data set) can be greater than its mean.

Thus, an exponential regression model is proposed in Eq. (1) to predict the numbers of accidents λ_s , and Eq. (2) is applied to establish the statistical relationships between the observed numbers of accidents h_s that occurred at a signalized intersection s during a specific study period and its geometric characteristics comprising lane usage patterns X_s , which is a vector containing a set of geometric and lane usage parameters such as the numbers of approach lanes that permit different turning movements. β is a vector of numerical coefficients associated with each of the geometric design parameters, ϵ_s is an error term, and $\exp(\epsilon_s)$ is an exponential function that follows a gamma distribution with mean = 1 and variance = α^2 .

$$\lambda_s = \exp(\beta X_s + \epsilon_s), \quad \forall s = 1, 2, \dots, S \quad (1)$$

The negative binomial distribution Eq. (2) has the general form:

$$P(h_s) = \frac{\Gamma((1/\alpha) + h_s)}{\Gamma((1/\alpha) h_s!)} \left(\frac{1/\alpha}{(1/\alpha) + \lambda_s} \right)^{1/\alpha} \left(\frac{\lambda_s}{(1/\alpha) + \lambda_s} \right)^{h_s}, \quad \forall s = 1, 2, \dots, S \quad (2)$$

where Γ is a gamma function, and the respective likelihood function is given by Eq. (3):

$$L(\lambda_s) = \prod_{s=1}^S \frac{\Gamma((1/\alpha) + h_s)}{\Gamma((1/\alpha) h_s!)} \left(\frac{1/\alpha}{(1/\alpha) + \lambda_s} \right)^{1/\alpha} \left(\frac{\lambda_s}{(1/\alpha) + \lambda_s} \right)^{h_s} \quad (3)$$

The likelihood function in Eq. (3) is maximized to obtain the coefficient estimates vector β and the dispersion parameter α . This model form thus relaxes the strict restriction required by the simple Poisson regression model in which the mean must equal the variance. Overdispersion in raw data sets can then be modeled using $Var(h_s) = E(h_s) [1 + \alpha E(h_s)]$. Detailed explanations can be found in Washington *et al.* [42], and a similar approach has been applied by other researchers to deal with accident data [35], [43].

To verify the possible overdispersion of data, the ratio of the Pearson chi-square divided by the degree of freedom was checked. When this ratio is greater than 1.0, the model is considered to be overdispersed [44], and a Poisson regression model is consequently not suitable to fit the accident data. To assess the overall model fitness, the Akaike Information Criterion AIC (Akaike, 1973) – a benchmark indicator for assessing a regression model's relative quality – can be evaluated [45]. AIC is also a standard parameter to assess the corresponding quality of regression methods for a set of data that considers the information lost during the data analysis process, as a trade-off between goodness-of-fit and complexity of dataset. A smaller AIC value always gives a better-fitting model, and an AIC value can be optimized by increasing the number of explanatory variables and is given by Equation (4):

$$AIC = 2\kappa - 2 \ln(L) \quad (4)$$

where κ is the number of parameters in the model, and $\ln(L)$ is the maximum log-likelihood.

C. LANE-BASED OPTIMIZATION FOR SIGNALIZED INTERSECTIONS

In the existing lane-based design framework, design variables are given as follows. Let $\Lambda = (\Lambda_B, \Lambda_C)$ be the set of control variables, where Λ_B is a subset of the binary variables, and Λ_C is a subset of continuous variables. The subset Λ_B contains binary variables: lane markings (arrows) $\Delta_{i,j,k}$ ($\forall i = 1, \dots, N_T; j = 1, \dots, N_T; k = 1, \dots, K_i$) and successor functions, $\Omega_{i,j,l,m}$ ($\forall ((i,j), (l,m)) \in \Psi_s$). Another subset, Λ_C , includes continuous variables: lane flows, $q_{i,j,k}$ ($\forall i = 1, \dots, N_T; j = 1, \dots, N_T; k = 1, \dots, K_i$), common flow multiplier, μ , cycle length (reciprocal of cycle time), ζ , green start-times for turning movements,

$\theta_{i,j}$ ($\forall i = 1, \dots, N_T; j = 1, \dots, N_T$), green duration-times for turning movements, $\Theta_{i,k}$ ($\forall i = 1, \dots, N_T; k = 1, \dots, K_i$), green start-times on approach traffic lanes, and green duration-times for approach traffic lanes, $\Phi_{i,k}$ ($\forall i = 1, \dots, N_T; k = 1, \dots, K_i$).

Linear constraint sets have been developed to outline the feasible solution region that governs all binary and continuous variables in the lane-based design framework. Thus, Eq. (5) represents the conservation of demand flows in approach lanes that are required to ensure that the assigned lane flows in approach lanes are matched with the users' given demand flow-patterns:

$$\mu Q_{ij} = \sum_{k=1}^{K_i} q_{i,j,k}, \quad \forall i = 1, \dots, N_T; j = 1, \dots, N_T \quad (5)$$

Eq. (6) represents the minimum permitted movement from approach lanes to ensure that every approach lane is used to permit at least one movement turn:

$$\sum_{\substack{j=1 \\ j \neq i}}^{N_T} \Delta_{i,j,k} \geq 1, \quad \forall i = 1, \dots, N_T; k = 1, \dots, K_i \quad (6)$$

Eq. (7) represents the maximum movements permitted at the exit to avoid an excess of lane marking arrows (to permit movement turns), which can result in undesirable merging of traffic at (downstream) exit lanes.

$$\xi_{Z(i,j)} \geq \sum_{k=1}^{K_i} \Delta_{i,j,k}, \quad \forall i = 1, \dots, N_T; j = 1, \dots, N_T \quad (7)$$

Eq. (8) describes the permitted movements across adjacent approach lanes to prevent potential conflicts from turning movements being made from the same arm. A right-turn lane marking should not be designed on left-hand lanes because the resulting right-turn traffic movement will conflict with the left-turn traffic movement resulting from left-turn lane marking on right-hand lanes.

$$1 - \Delta_{i,j,k+1} \geq \Delta_{i,m,k}, \quad \forall i = 1, \dots, N_T; j = 1, \dots, N_T; m = j + 1, \dots, N_T; k = 1, \dots, K_i - 1 \quad (8)$$

Eq. (9) represents the cycle length required to ensure that the operating cycle time is always bounded by the users' specified minimum and maximum ranges.

$$\frac{1}{c_{\min}} \geq \zeta \geq \frac{1}{c_{\max}} \quad (9)$$

Eq. (10) and Eq. (11) describe the lane-signal settings needed to synchronize the traffic signal timings, including starts of green and durations of green, for all approach lanes and movement-turns matched with the lane-marking designs.

$$M(1 - \Delta_{i,j,k}) \geq \Theta_{i,k} - \theta_{i,j} \geq -M(1 - \Delta_{i,j,k}), \quad \forall i = 1, \dots, N_T; j = 1, \dots, N_T; k = 1, \dots, K_i \quad (10)$$

$$M(1 - \Delta_{i,j,k}) \geq \Phi_{i,k} - \varphi_{i,j} \geq -M(1 - \Delta_{i,j,k}), \quad \forall i = 1, \dots, N_T; j = 1, \dots, N_T; k = 1, \dots, K_i \quad (11)$$

Eq. (12) is the start of green, and Eq. (13) is the duration of green required to regulate the starts and durations of green times within a signal cycle.

$$1 \geq \theta_{i,j} \geq 0, \quad \forall i = 1, \dots, N_T; j = 1, \dots, N_T \quad (12)$$

$$1 \geq \varphi_{i,j} \geq g_{i,j}\zeta, \quad \forall i = 1, \dots, N_T; j = 1, \dots, N_T \quad (13)$$

Eq. (14) describes the order of signal displays required to ensure that any pair of incompatible movements including vehicular (i.e., traffic from different arms) and pedestrian phases are well-separated and ordered by the signal timings within a signal cycle.

$$\Omega_{i,j,l,m} + \Omega_{l,m,i,j} = 1, \quad \forall ((i,j), (l,m)) \in \Psi_s \quad (14)$$

Eq. (15) describes the clearance time constraint needed to provide sufficient inter-green times to separate any pair of incompatible movements within a signal cycle:

$$\theta_{l,m} + \Omega_{i,j,l,m} + M(2 - \Delta_{i,j,k} - \Delta_{l,m,n}) \geq \theta_{i,j} + \varphi_{i,j} + \omega_{i,j,l,m}\zeta, \quad \forall ((i,j), (l,m)) \in \Psi \quad (15)$$

where M is an arbitrary large positive constant number, and $\omega_{i,j,l,m}$ is a user-specified minimum intergreen or clearance time for the conflicting traffic-turning movements (i,j) and (l,m) .

Eq. (16) and Eq. (17) represent the prohibited movements, where this prohibition prevents the need for redundant lane markings for turns for which there is no demand.

$$MQ_{i,j} \geq \sum_{k=1}^{K_i} \Delta_{i,j,k}, \quad \forall i = 1, \dots, N_T; j = 1, \dots, N_T \quad (16)$$

$$M\Delta_{i,j,k} \geq q_{i,j,k} \geq 0, \quad \forall i = 1, \dots, N_T; j = 1, \dots, N_T; k = 1, \dots, K_i \quad (17)$$

Eq. (18) describes the identical flow factor across adjacent approach lanes that equalizes the flow factors for any two adjacent approach lanes from the same arm with the same lane-marking arrow.

$$\begin{aligned} & M(2 - \delta_{i,j,k} - \delta_{i,j,k+1}) \\ & \geq \frac{1}{\bar{v}_{i,k}} \sum_{j=1, \dots, N_T-1} \left(1 + \frac{1.5}{r_{i,j,k}}\right) q_{i,j,k} \\ & \quad - \frac{1}{\bar{v}_{i,k+1}} \sum_{j=1, \dots, N_T-1} \left(1 + \frac{1.5}{r_{i,j,k+1}}\right) q_{i,j,k+1} \\ & \geq -M(2 - \delta_{i,j,k} - \delta_{i,j,k+1}), \quad \forall k = 1, \dots, K_i - 1; i = 1, \dots, N_T \quad (18) \end{aligned}$$

Eq. (19) describes the maximum acceptable degree of saturation permitted, which ensures that the traffic streams are less than the maximum allowable flow-to-capacity ratios given by users.

$$\Phi_{i,k} + e\zeta \geq \frac{1}{p_{i,k}\bar{v}_{i,k}} \sum_{j=1}^{N_T-1} (q_{i,j,k}), \quad \forall i = 1, \dots, N_T; k = 1, \dots, K_i \quad (19)$$

where $P_{i,k}$ is the maximum permitted degree of saturation on lane k from arm i ($= 0.9$). For a traffic lane k from arm i , the degree of saturation can be expressed as:

$$\bar{\rho}_{i,k} = \frac{y_{i,k}}{\Phi_{i,k} + e\zeta} \leq p_{i,k}, \quad \forall i = 1, \dots, N_T; k = 1, \dots, K_i$$

where $\bar{\rho}_{i,k}$ is the degree of saturation on lane k from arm i and e is the extra effective green time from the difference between the actual and effective greens (measured in time units, which are usually taken as 1 s).

D. GOVERNING CONSTRAINT SETS FOR SAFETY ENHANCEMENTS

In this section, new governing constraint sets are proposed for addition to the conventional lane-based design framework given in Eq. (5) – Eq. (19) to further refine the feasible solution region to enhance safety performance. Additional constraint sets could be developed to establish lane-marking patterns, such as numbers and positions of shared lanes, to improve safety at signalized intersections. Developing the new constraint sets depends on the results of the proposed statistical analysis. Statistically, the positions of shared-lane markings, the total numbers of straight-ahead movements, and the numbers of permitted movements on shared lanes may all influence safety at intersections. Detailed results of our analysis are given in the case study section.

1) RESTRICTING SHARED LANE MARKINGS ONLY ON NEARSIDE LANES

Based on the intersection layout (for left-hand side traffic) and model notations, the first traffic lane from the left-hand side ($k = 1$) is regarded as a nearside lane, which is an approach lane next to the pavement kerb as given in Figure 1. Eq. (20) is designed to ensure that the lane marking can only be a single arrow, to permit only one direction of movement turn to be allowed from all non-nearside approach lanes. Mathematically, an equality constraint set Eq. (20) mainly controls all approach lanes $k \in \{2, \dots, K_i\}$ from all arms $i \in \{1, 2, \dots, N_T\}$. All nearside lanes where $k = 1$ will be excluded from this equality requirement in all arms.

$$\sum_{j=1}^{N_T-1} \Delta_{i,j,k} = 1, \quad \forall i = 1, \dots, N_T; k = 2, \dots, K_i. \quad (20)$$

2) RESTRICTING MAXIMUM NUMBER OF STRAIGHT-AHEAD LANE MARKINGS

Eq. (21) is developed to control the numbers of lane markings for straight-ahead movements from all individual arms where η_i is a user-specified maximum number of the straight-ahead lane-markings from arm i . To model different types (or shapes) of intersections, a mathematical function $\tau(i,j)$ is developed to identify which specific movement turn j from arm i is a straight-ahead movement U .

$$\sum_{k=1}^{K_i} \Delta_{i,j,k} \leq \eta_i, \quad \forall i = 1, \dots, N_T; \forall j \in \tau(i,j) \subset U. \quad (21)$$

3) RESTRICTING MAXIMUM NUMBER OF TURNS ON SHARED LANES

Practically, shared lanes are designed to permit more than a single-movement turn (i.e., two or even three movement turns, simultaneously) on approach lanes. Eq. (22) can be added to the formulation to restrict the maximum numbers of turn movements to two in all approach lanes to prohibit the most complex shared-lane marking (i.e., that for three turns; see Figure 2) to improve safety.

$$\sum_{j=1}^{N_i-1} \Delta_{i,j,k} \leq 2, \quad \forall i = 1, \dots, N_T; k = 1, \dots, K_i \quad (22)$$

4) RELOCATING ALL AT-GRADE PEDESTRIAN MOVEMENTS TO GRADE SEPARATED INFRASTRUCTURES

In practice, all roundabout, priority, and signalized intersections are designed following the well-established Hong Kong codes of practices, which are based on EU or UK precedent. Traffic-signal settings for pedestrian movements do not generally influence the engineering performance of the entire intersection as long as the minimum green duration-times for the pedestrian phases are provided in the design framework.

Elevated flyovers or underground walkway tunnels can be built to completely separate the pedestrian movements and vehicular traffic movements, which further reduces the accident risk. Given the safety enhancements they provide, such grade-separated structures should always be proposed, depending on financial constraints and the availability of space. In the lane-based model, all governing constraints for the minimum green-durations and clearance times in Eqs. (12-15) could be removed because they did not exist.

IV. CASE STUDY

In this section, a case study is given to demonstrate the proposed method for designing signalized intersections to reduce accident risk. First, the observed accident count numbers and relevant statistical records were collected from the Traffic Annual Reports 2004-2014 published by the Hong Kong Police Force. Intersections that are grouped as ‘accident blacksite’ were retrieved. Intersection names, types, locations, and observed accident counts h_s were recorded. Geometric layouts and lane marking patterns were collected for surveyed signalized intersections. The vector \mathbf{X}_s for each intersection s was compiled numerically and contains eight design parameters, as listed in Table 1.

A. COLLECTION OF ACCIDENT COUNT DATA AND GEOMETRIC LAYOUT DETAILS OF INTERSECTIONS

Accident count data for analysis in this study were collected from the Hong Kong Police Force publication entitled ‘Traffic Annual Report’. Accident counts from 2004-2014 have been retrieved from the official website. On this website, a list of ‘accident blacksites’ are updated annually, containing

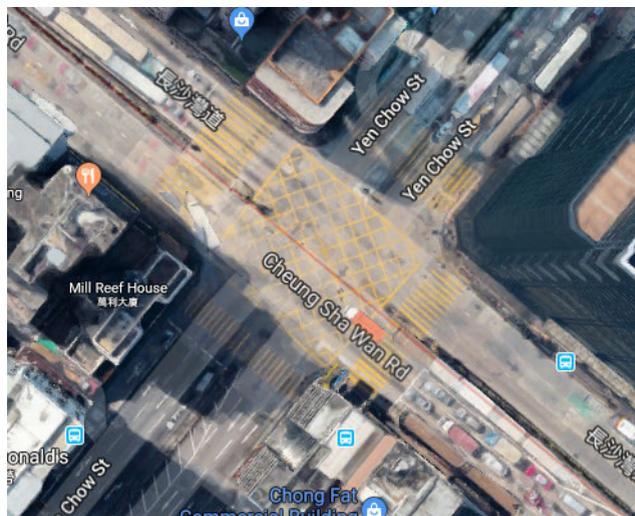


FIGURE 3. Signalized intersection between Cheung Sha Wan Road and Yen Chow Street.

mainly signalized intersections in Hong Kong. To be categorized as a ‘traffic accident blacksite’, there must be at least 9 vehicle injury accidents or at least 6 pedestrian injury accidents at the intersection within the previous 12-month period. Since 2010, an intersection at which two or more fatal traffic accidents occur within 5 years is also classified as an ‘accident blacksite’.

This study established statistical relationships between the observed accident numbers and respective geometric layouts and lane-marking patterns for various surveyed signalized intersections. Relevant geometric intersection layouts and lane-marking patterns from those blacksites were collected for statistical analysis. Information sources included local traffic aids, the Google map, and site visits. The required numerical figures relating to geometric layouts and lane-marking patterns were counted manually.

It was expected that this investigation would show that layout settings of signalized intersections could be varied in practical operations to reduce accident risks. To establish the statistical relationships between these intersection layouts of different lane marking patterns and their observed accident counts, we needed to collect the geometric features and lane-marking patterns for all surveyed signalized intersections for analysis, and this was done by site-visits and photocapturing from Google maps.

In data collection, eight independent variables in the vector \mathbf{X}_s listed in Table 1 for all intersection s were counted and recorded manually. For example, the intersection at Cheung Sha Wan Road/Yen Chow Street in Figure 3 was captured from Google Maps. By enlarging the view of the intersection from various directions, details of the intersection layout could be recorded, including the lane arrangements, lane markings, and pedestrian crossing details. Figure 4 was compiled to show the necessary information to define the eight independent variables in the vector \mathbf{X}_s based on the lane-marking patterns and pedestrian crossings. Table 2 collates the numerical values for the eight independent variables

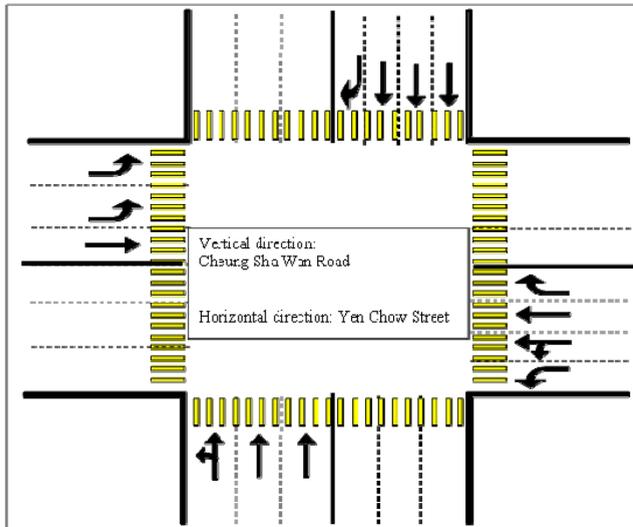


FIGURE 4. Details of lane markings and pedestrian crossings at intersection between Cheung Sha Wan Road and Yen Chow Street.

TABLE 2. Numerical values of the eight independent variables at intersection between Cheung Sha Wan Road/Yen Chow Street.

Independent variables, X_s	Numerical value
1. Number of approach lanes with left-turn lane marking	3
2. Number of approach lanes with straight-ahead lane marking	7
3. Number of approach lanes with right-turn lane marking	2
4. Ratio of number of exit lanes to number of approach lanes	0.8571 (=12/14)
5. Number of approach lanes with shared lane markings permitting 2 movements	2
6. Number of approach lanes with shared lane markings permitting 3 movements	0
7. Number of shared lane markings on nearside lane	1
8. Number of traffic lanes perpendicular to pedestrian crossing	26

at the intersection between Cheung Sha Wan Road and Yen Chow Street.

Similarly, these numerical values for the eight independent variables could be retrieved from the other surveyed intersections. If road maintenance work occurred during the time period in which the photos were taken (as shown in Figure 5), alternative site-visit times were arranged to collect the missing layout details at the intersection.

After collecting the accident counts and the geometric layouts of all surveyed signalized intersections, basic statistics listing their minimum, maximum, mean, and standard deviation figures were summarized and are presented in Table 2. Table 3 shows the relationships between pairs of independent variables that are considered to be acceptable for regression analysis. This study used regression-type analysis to relate the accident counts to various lane-usage patterns, so it is necessary to ensure low correlations among the independent variables to avoid multicollinearity in the statistical analysis.



FIGURE 5. Alterations of existing lane markings due to temporary roadworks.

TABLE 3. Correlations of all independent variables for signalized intersections in the case study.

X_s	1	2	3	4	5	6	7	8
1	1							
2	-0.128	1						
3	0.114	0.167	1					
4	-0.283	-0.092	-0.388	1				
5	-0.119	-0.236	0.132	-0.007	1			
6	0.065	-0.232	0.079	-0.000	0.000	1		
7	-0.353	-0.140	-0.055	0.042	0.701	0.144	1	
8	-0.086	0.317	-0.039	0.002	0.219	-0.147	0.208	1

B. CASE STUDY RESULTS OF THE NEGATIVE BINOMIAL REGRESSION ANALYSIS

Having checked the correlations among pairs of independent variables, the data set was analyzed using the statistical package ‘R’. A negative binomial regression model was chosen to generate the regression equation through the ‘glm.nb’ function. The ratio of the Person’s Chi square (χ^2) over the degree of freedom was 1.35, which verified that overdispersion exists in the raw data and that the negative binomial regression analysis should be used instead of Poisson regression analysis. The AIC value for measuring the overall model fitness was calculated to be 390.53. Other relevant statistical results are collected in Table 4 for further investigation. Using the regression analysis, the levels of significance of the independent variables were evaluated. This result could guide the design of safer signalized intersections by controlling the intersection layouts and lane-marking patterns.

To demonstrate the applications of the negative binomial regression analysis results for designing a safer signalized intersection, this technique was applied to the case study intersection depicted in Figure 4. This intersection was chosen because it is an accident blacksite, with ($h_s =$) 93 observed accidents.

The previous section identified statistically significant independent lane-marking variables correlating to accident

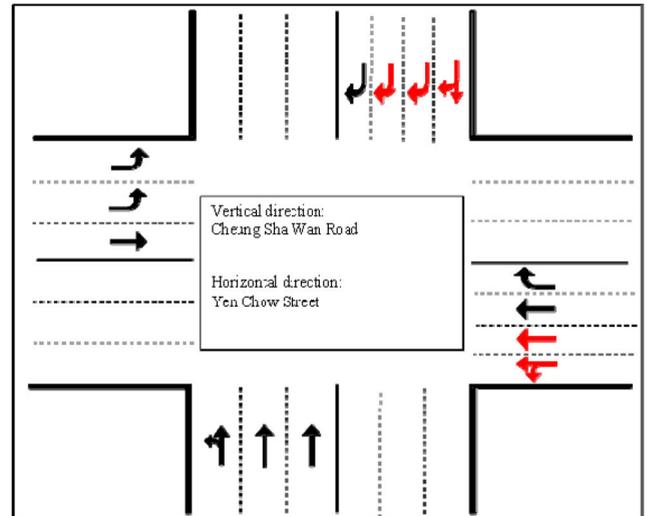
TABLE 4. Results of the Negative binomial regression analysis using the statistical package R.

X_s	Estimated coefficient	Standard error	z-value	p-value	Significance
1	-0.14434	0.08953	-1.612	0.106929	
2	0.20949	0.06849	3.059	0.002223	0.01
3	-0.09507	0.08075	-1.177	0.239042	
4	0.45116	0.89828	0.502	0.615496	
5	0.11847	0.10356	1.144	0.252656	
6	1.46593	0.48618	3.015	0.002568	0.01
7	-0.51975	0.13837	-3.756	0.000172	0.001
8	0.07626	0.01561	4.885	1.04e-06	0.001

Remarks: Statistically significant variables are those with z-value >2 (or < -2). Constant term: Estimated coefficient = 1.71645 with z-value 1.723 and p-value = 0.084932.

numbers, including the (a) number of approach lanes with straight-ahead lane markings [positive coefficient]; (b) total number of shared lanes with three permitted traffic movements [positive coefficient]; (c) number of shared-lane marking on nearside lanes [negative coefficient]; and (d) number of traffic lanes perpendicular to a pedestrian crossing [positive coefficient]. The positive or negative signs of the coefficients derive from the fact that it was found that the independent variables in (a), (b), and (d) were positively related to accident numbers whereas the independent variable in (c) showed a negative relationship with the accident numbers. It follows that reducing the values of variables in (a), (b), and (d) in the intersection design should reduce accident risks, while conversely, increasing the value of variable in (c) should reduce the accident risk.

With respect to the findings in (a), a new constraint set in Eq. (21) was proposed to match the requirements in the numbers of straight-ahead movements. From Table 1, the averaged total number of straight-ahead lane markings for all surveyed intersections was 5.707. To reduce accident risks, we set $\sum_{i=1}^4 \eta_i \leq 5$ for the case study intersection in which the maximum number is just below the average value. For special requirements, we may further restrict the total number of straight-ahead lane markings from each arm i by setting $\eta_i \leq K_i - 1$ or $\eta_i \leq K_i - 2$ depending on the actual sizes of the intersection. In the case study 4-arm intersection in Figure 4, $\eta_{i=1} \leq K_i - 1 = 2$ (northbound direction) and $\eta_{i=2} \leq K_i - 1 = 2$ (eastbound direction) are used for the arms with three approach lanes. Similarly, $\eta_{i=3} \leq K_i - 2 = 2$ (southbound direction) and $\eta_{i=4} \leq K_i - 2 = 2$ (westbound direction) are used for the arms with four approach lanes.

**FIGURE 6. Enhanced layout and lane markings at the intersection between Cheung Sha Wan Road and Yen Chow Street for reducing accident risks.**

With respect to results in (b), we prohibited shared-lane markings permitting three movement turns using Eq. (22) in all arms in redesigning the case-study intersection.

For the results obtained in (c), shared-lane marking should not be permitted on non-nearside lanes that can be implemented by Eq. (20), such that lane markings with only single movement turns could be incorporated in non-nearside lanes.

With respect to the results in (d), all pedestrian movements were removed from the optimization framework. As a demonstration, the lane marking patterns were re-optimized by adding the new safety constraint sets proposed in this study with users' inputs to the relevant lane marking parameters, according to the guidance from the above statistical analysis results. In general, we were able to continue optimizing the intersection capacity in which the common flow multiplier μ could be set as the objective function for optimization (found in Eq. (5)) subject to linear constraints in Eqs. (5-22). This binary-mixed-integer-linear-programming (BMILP) problem can be solved by standard branch-and-bound routine using CPLEX solver. Taking the intersection at Cheung Sha Wan Road and Yen Chow Street in Figure 4 as the case study, the revised lane marking patterns are given in Figure 6 and summarized in Table 5.

It was found that, due to the new constraint sets on restricting the total number of lane markings for straight-ahead movements in (a), the resultant total number of straight-ahead lane markings were reduced from seven to five. Thus, all the three straight-ahead movements have been removed from Cheung Sha Wan Road (southbound direction) but a new straight-ahead lane marking was added along Yen Chow Street (eastbound direction). Because the number of right-turn lane markings is not a significant variable, no new constraint set was added to control the right-turn lane-marking pattern, and hence two more right-turn lane markings were added to the arm. Also, because shared lane-markings that

TABLE 5. Numerical values of the eight independent variables at intersection between Cheung Sha Wan Road/Yen Chow Street after re-designing the lane markings with the new safety constraint sets.

Independent variables, X_c	Numerical value
1. Number of approach lanes with left-turn lane marking	2
2. Number of approach lanes with straight-ahead lane marking	5
3. Number of approach lanes with right-turn lane marking	4
4. Ratio of number of exit lanes to number of approach lanes	0.8571 (=12/14)
5. Number of approach lanes with shared-lane markings permitting two movements	3
6. Number of approach lanes with shared-lane markings permitting three movements	0
7. Number of shared-lane markings on nearside lane	3
8. Number of traffic lanes perpendicular to pedestrian crossing	0

permit two-movement turns are allowed only along nearside lanes in (c), one new shared lane marking was added along Cheung Sha Wan Road (southbound direction), and the original shared lane marking from a non-nearside lane along Yen Chow Street (westbound direction) was then required on the next nearside lane. Thus, there were a total of three shared lane-markings on nearside lanes in the revised design.

Another statistically significant independent variable in (b) was the total number of shared lanes that permitted three traffic movements that remained ‘0’ before and after the alterations. From the statistical analysis results, it was also found that pedestrian crossings are dangerous, leading to higher accident risks in (d). There were originally 26 traffic lanes perpendicular to the pedestrian crossings, as shown in Figure 4 and Table 2. To reduce conflicts between vehicular traffic and pedestrian movements at a signalized intersection, the pedestrian crossing could be removed and a new flyover could be built to enable safer pedestrian movement (Figure 7). Moreover, depending on the financial viability, all four at-grade pedestrian crossings at the intersection could be removed and replaced by grade-separated infrastructures such as underground walkways or elevated flyovers. Thus, the number of traffic lanes perpendicular to pedestrian crossings could probably be reduced from 26 to zero. Having altered the lane marking patterns and intersection layout according to the changes above, the predicted number of accidents for the intersection at Cheung Sha Wan Road and Yen Chow Street number, based on the regression model result in Table 4 and the numerical values for the vector X_c in Table 5, was reduced to four.

Of course, restructuring all at-grade pedestrian movements to grade-separated facilities would be a special case. In urban areas, acquiring sufficient space to build grade-separated facilities for pedestrians is not simple. Nevertheless, if the at-grade pedestrian pathways are not altered and remain in the site, the predicted accident counts may well increase to 26. A compromise solution is offered by the lane-marking



FIGURE 7. Upgrading of at-grade pedestrian crossing to a grade-separated flyover.

alteration described above, which was predicted to reduce the accident count to four at this signalized intersection.

V. CONCLUSIONS

The lane-based optimization framework for traffic signal settings has been applied for isolated intersections and signal-controlled networks. Linear constraint sets have been developed to govern the lane-marking design and traffic signal settings. Conventional lane-based designs focus on optimizing the engineering performance, such as maximizing the intersection or network capacity or minimizing the total delays for motorists. To find an alternative solution, a statistical analysis for the accident counts and intersection layouts and lane-marking patterns was conducted. A negative binomial regression model was developed to relate the accident counts to various statistically significant independent variables. Their variable effects were analyzed to guide the lane-marking designs by establishing new governing constraint sets. Using this framework, accident blacksites in Hong Kong were studied. It was found that accident risks at signal-controlled intersections could be reduced effectively by (i) reducing the number of approach lanes in which straight-ahead movements were permitted, (ii) reducing the total number of shared lanes with three types of traffic movements permitted, (iii) reducing the number of pedestrian crossings perpendicular to traffic lanes, and (iv) increasing the shared-lane markings in nearside lanes.

One limitation of this study is the fixed directional structure of the intersections. The accident survey was conducted for accident blacksites, which are the busiest intersections in Hong Kong, where motorists are permitted to make complex turning movements. However, if the study intersection only involves straight-ahead movements, then the straight-ahead lane markings must be preserved: there is no scope for this to be altered. Therefore, future work must involve expanded intersection survey sampling, to cover intersections with different directional structures, so that the statistical analysis can be correctly categorized, to provide regression equations and findings applicable to each type of intersection, and thus yield safety designs and guidelines specific to each type of intersection.

To investigate the directional structures in intersection levels, the new design constraint sets for safety enhancement could also be applied to signalized network design. Lane markings could be optimized to ensure that sufficient paths exist to connect the origin-destination pairs across all intersections. It may even be feasible to replace the dangerous straight-ahead movement at a blacksite intersection with a series of left- and right-turns across several downstream intersections.

REFERENCES

- [1] *Global Status on Road Safety*, World Health Organization (WHO), Geneva, Switzerland, 2018.
- [2] Transport Department. (2017). *Road Traffic Accident Statistics*. [Online]. Available: https://www.td.gov.hk/en/road_safety/road_traffic_accident_statistics/2017/index.html
- [3] W. Hao, C. Ma, B. Moghimi, Y. Fan, and Z. Gao, "Robust optimization of signal control parameters for unsaturated intersection based on tabu search-artificial bee colony algorithm," *IEEE Access*, vol. 6, pp. 32015–32022, 2018.
- [4] T.-T. Ngo, T. Huynh-The, and D.-S. Kim, "A novel VANETs-based traffic light scheduling scheme for greener planet and safer road intersections," *IEEE Access*, vol. 7, pp. 22175–22185, 2019.
- [5] X. Li and J.-Q. Sun, "Signal multiobjective optimization for urban traffic network," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 11, pp. 3529–3537, Nov. 2018.
- [6] X. Li and J.-Q. Sun, "Turning-lane and signal optimization at intersections with multiple objectives," *Eng. Optim.*, vol. 51, no. 3, pp. 484–502, 2019.
- [7] A. C. Malliaris, J. H. DeBlois, and K. H. Digges, "Light vehicle occupant ejections—A comprehensive investigation," *Accident Anal. Prevention*, vol. 28, no. 1, pp. 1–14, 1996.
- [8] M. Öström and A. Eriksson, "Pedestrian fatalities and alcohol," *Accident Anal. Prevention*, vol. 33, no. 2, pp. 173–180, 2001.
- [9] O. T. Holubowycz, C. N. Kloeden, and A. J. McLean, "Age, sex, and blood alcohol concentration of killed and injured drivers, riders, and passengers," *Accident Anal. Prevention*, vol. 26, no. 4, pp. 483–492, 1994.
- [10] A. S. Al-Ghamdi, "Analysis of traffic accidents at urban intersections in Riyadh," *Accident Anal. Prevention*, vol. 35, no. 5, pp. 717–724, 2003.
- [11] A. S. Mane and S. S. Pulugurtha, "Influence of on-network, traffic, signal, demographic, and land use characteristics by area type on red light violation crashes," *Accident Anal. Prevention*, vol. 120, pp. 101–113, Nov. 2018.
- [12] E. Shesterov and A. Mikhailov, "Accident rates at signalized intersections," *Transp. Res. Procedia*, vol. 20, no. 2017, pp. 613–617, 2017.
- [13] J. Oh, C. Lyon, S. Washington, B. Persaud, and J. Bared, "Validation of FHWA crash models for rural intersections: Lessons learned," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1840, no. 1, pp. 41–49, 2003.
- [14] A. Tarko and M. Tracz, "Accident prediction models for signalized crosswalks," *Saf. Sci.*, vol. 19, pp. 109–118, Jun. 1995.
- [15] P. Greibe, "Accident prediction models for urban roads," *Accident Anal. Prevention*, vol. 35, no. 2, pp. 273–285, 2003.
- [16] M. G. Karlaftis and I. Golias, "Effects of road geometry and traffic volumes on rural roadway accident rates," *Accident Anal. Prevention*, vol. 34, no. 3, pp. 357–365, 2002.
- [17] R. Elvik, "Does the influence of risk factors on accident occurrence change over time?" *Accident Anal. Prevention*, vol. 91, pp. 91–102, 2016.
- [18] H. Chen, L. Cao, and D. B. Logan, "Analysis of risk factors affecting the severity of intersection crashes by logistic regression," *Traffic Injury Prevention*, vol. 13, no. 3, pp. 300–307, 2012.
- [19] K. K. W. Yau, "Risk factors affecting the severity of single vehicle traffic accidents in Hong Kong," *Accident Anal. Prevention*, vol. 36, no. 3, pp. 333–340, 2004.
- [20] E. Polders, S. Daniels, E. Hermans, T. Brijs, and G. Wets, "Crash patterns at signalized intersections," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2514, no. 1, pp. 105–116, 2015.
- [21] V. Shankar and F. Mannering, "An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity," *J. Saf. Res.*, vol. 27, no. 3, pp. 183–194, 1996.
- [22] A. Pande and M. Abdel-Aty, "A novel approach for analyzing severe crash patterns on multilane highways," *Accident Anal. Prevention*, vol. 41, no. 5, pp. 985–994, 2009.
- [23] A. K. Çelik and E. Oktay, "A multinomial logit analysis of risk factors influencing road traffic injury severities in the erzurum and kars provinces of turkey," *Accident Anal. Prevention*, vol. 72, pp. 66–77, Nov. 2014.
- [24] S. Anjana and M. V. L. R. Anjaneyulu, "Safety analysis of urban signalized intersections under mixed traffic," *J. Saf. Res.*, vol. 52, pp. 9–14, Feb. 2015.
- [25] D. Lord and F. Mannering, "The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives," *Transp. Res. A, Policy Pract.*, vol. 44, no. 5, pp. 291–305, 2010.
- [26] M. Abdel-Aty, "Analysis of driver injury severity levels at multiple locations using ordered probit models," *J. Saf. Res.*, vol. 34, no. 5, pp. 597–603, 2003.
- [27] M. W. Knuiman, F. M. Council, and D. W. Reinfurt, "Association of median width and highway accident rates (with discussion and closure)," *Transp. Res. Rec.*, vol. 1401, no. 1401, pp. 70–82, 1993.
- [28] A. C. Cameron and P. K. Trivedi, *Regression Analysis of Count Data*. Cambridge, U.K.: Cambridge Univ. Press, 1998.
- [29] J. Oh, S. Washington, and K. Choi, "Development of accident prediction models for rural highway intersections," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1897, no. 1, pp. 18–27, 2004.
- [30] H. C. Chin and M. A. Quddus, "Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections," *Accident Anal. Prevention*, vol. 35, no. 2, pp. 253–259, 2003.
- [31] C. Chen and Y. Xie, "Modeling the effects of AADT on predicting multiple-vehicle crashes at urban and suburban signalized intersections," *Accident Anal. Prevention*, vol. 91, pp. 72–83, Jun. 2016.
- [32] D.-H. Kim, L. M. Ramjan, and K.-K. Mak, "Prediction of vehicle crashes by drivers' characteristics and past traffic violations in Korea using a zero-inflated negative binomial model," *Traffic Injury Prevention*, vol. 17, no. 1, pp. 86–90, 2016.
- [33] K. Wang, J. K. Simandl, M. D. Porter, A. J. Graettinger, and R. K. Smith, "How the choice of safety performance function affects the identification of important crash prediction variables," *Accident Anal. Prevention*, vol. 88, pp. 1–8, Mar. 2016.
- [34] S.-P. Miaou, "The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions," *Accident Anal. Prevention*, vol. 26, no. 4, pp. 471–482, 1994.
- [35] S. C. Wong, N. N. Sze, and Y. C. Li, "Contributory factors to traffic crashes at signalized intersections in Hong Kong," *Accident Anal. Prevention*, vol. 39, no. 6, pp. 1107–1113, 2007.
- [36] D. Lord, S. P. Washington, and J. N. Ivan, "Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: Balancing statistical fit and theory," *Accident Anal. Prevention*, vol. 37, no. 1, pp. 35–46, 2005.
- [37] C. K. Wong and S. C. Wong, "Lane-based optimization of signal timings for isolated junctions," *Transp. Res. B, Methodol.*, vol. 37, no. 1, pp. 63–84, 2003.

- [38] C. K. Wong and B. G. Heydecker, "Optimal allocation of turns to lanes at an isolated signal-controlled junction," *Transp. Res. B, Methodol.*, vol. 45, no. 4, pp. 667–681, 2011.
- [39] Y. Liu and C. K. Wong, "Refining lane-based traffic signal settings to satisfy spatial lane length requirements," *J. Adv. Transp.*, vol. 2017, May 2017, Art. no. 8167530.
- [40] C. K. Wong and Y. Liu, "Lane-based optimization for macroscopic network configuration designs," *Discrete Dyn. Nature Soc.*, vol. 2017, Jul. 2017, Art. no. 1257569.
- [41] X. Zhuang, C. Wu, and S. Ma, "Cross or wait? Pedestrian decision making during clearance phase at signalized intersections," *Accident Anal. Prevention*, vol. 111, pp. 115–124, Feb. 2018.
- [42] S. P. Washington, M. Karlaftis, and F. L. Mannering, *Statistical and Econometric Methods for Transportation Data Analysis*. Boca Raton, FL, USA: Chapman and Hall, 2003.
- [43] L.-Y. Chang, "Analysis of freeway accident frequencies: Negative binomial regression versus artificial neural network," *Saf. Sci.*, vol. 43, pp. 541–557, 2005.
- [44] J. M. Hilbe, *Modeling Count Data*. Cambridge, U.K.: Cambridge Univ. Press, 2014.
- [45] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Proc. 2nd Int. Symp. Inf. Theory, Tsahkadsor, Armenia, USSR*, B. N. Petrov and F. Csaki, Eds. Budapest, Hungary: Akadémiai Kiadó, Sep. 1971, pp. 267–281.



C. K. WONG received the Ph.D. degree in civil engineering from The University of Hong Kong, in 2004. He is currently an Assistant Professor with the City University of Hong Kong. His current research interests include traffic signal control and optimization.

• • •