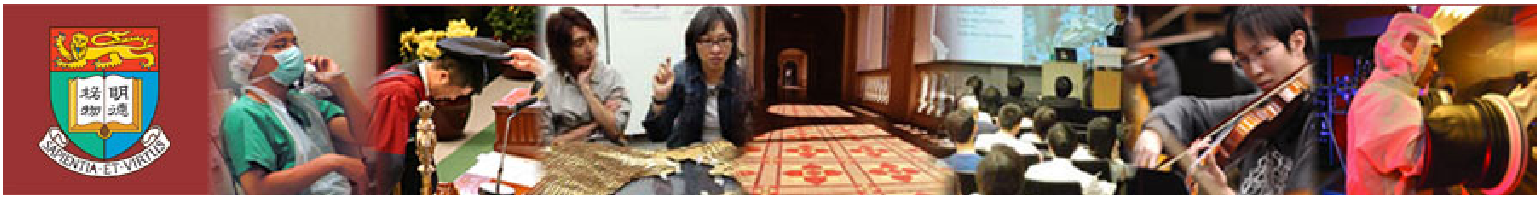

Title	Bootstrap resampling approach to disaggregate analysis of road crashes in Hong Kong
Author(s)	Pei, X; Sze, NN; Wong, SC; Yao, D
Citation	Accident Analysis & Prevention, 2016, v. 95 n. pt. B, p. 512-520
Issued Date	2016
URL	http://hdl.handle.net/10722/231701
Rights	

BOOTSTRAP RESAMPLING APPROACH TO DISAGGREGATE ANALYSIS OF ROAD CRASHES IN HONG KONG

Authors:

Xin PEI*
Department of Automation, Tsinghua University, Beijing, China
Tel: (86) 10 6279 5043
Fax: (86) 10 6279 5043
Email: peixin@mail.tsinghua.edu.cn

N.N. SZE
Department of Civil and Natural Resources Engineering, University of Canterbury, Christchurch, New Zealand
Tel: (64) 3364 2250
Fax: (64) 3364 2758
Email: tony.sze@canterbury.ac.nz

S.C. WONG
Department of Civil Engineering, The University of Hong Kong, Hong Kong, China
Tel: (852) 2859 1964
Fax: (852) 2559 5337
Email: hhecwsc@hku.hk

Danya YAO
Department of Automation, Tsinghua University, Beijing, China
Tel: (86) 10 6279 5043
Fax: (86) 10 6279 5043
Email: yaody@mail.tsinghua.edu.cn

ABSTRACT

Road safety affects health and development worldwide; thus, it is essential to examine the factors that influence crashes and injuries. As the relationships between crashes, crash severity, and possible risk factors can vary depending on the type of collision, we attempt to develop separate prediction models for different crash types (i.e., single- versus multi-vehicle crashes and slight injury versus killed and serious injury crashes). Taking advantage of the availability of crash and traffic data disaggregated by time and space, it is possible to identify the factors that may contribute to crash risks in Hong Kong, including traffic flow, road design, and weather conditions. To remove the effects of excess zeros on prediction performance in a highly disaggregated crash prediction model, a bootstrap resampling method is applied. The results indicate that more accurate and reliable parameter estimates, with reduced standard errors, can be obtained with the use of a bootstrap resampling method. Results revealed that factors including rainfall, geometric design, traffic control, and temporal variations all determined the crash risk and crash severity. This helps to shed light on the development of remedial engineering and traffic management and control measures.

Keywords: single-vehicle crash, multiple-vehicle crash, injury severity, count data model, bootstrap resampling method

1. INTRODUCTION

Road safety is a global issue linked to health and development. With more than 1.3 million road deaths and 50 million injuries occurring every year, road crashes are expected to become the fifth leading cause of death worldwide by 2030 (WHO, 2009). Fatalities and injuries result in losses of life and property and decreased quality of life. A better understanding of the factors contributing to road crashes, injuries, and deaths is critical to the development of appropriate road safety measures.

Considering the differences in crash circumstances and collision mechanisms, the factors contributing to, and their effects on injury severities of, single-vehicle (SV) and multi-vehicle (MV) crashes can be differentiated. Therefore, the prediction performances of separate crash prediction models for different collision types are superior to that of a combined crash prediction model. Separate crash prediction models for SV and MV crashes also have the capability to reveal the distinctive relationships between risks of crash and various contributory factors (Mensah and Hauer, 1998). In particular, the associations between crash frequencies and possible risk factors, such as geometric design, weather, seasonal variations (Shankar et al., 1995), and day and night conditions on rural roads (Persaud and Mucsi, 1995), for both SV and MV crashes have been revealed.

The effects of possible risk factors on injury severity of SV and MV crashes can be differentiated. In particular, the different risk factors associated with fatality risks of SV and MV crashes in Northern Sweden have been significantly distinguished (Öström and Eriksson, 1993). Differences in the associations between injury risk and possible factors have also been identified for truck drivers on rural roads (Chen and Chen, 2011), motorways (Bham et al., 2012), and urban roadways (Yau, 2004, 2006; Jung et al., 2010, 2012; Fréchède et al., 2010; Xie et al., 2011; Kim et al., 2012). In this study, we attempt to identify the differences in the relationship between possible factors and risk of crashes of different types with respect to collision types (i.e., SV versus MV crashes) and crash severity (i.e., killed and severe injury versus slight injury crashes). Due to their non-negative and random nature, count data models, including Poisson regression and negative binomial regression models, have long been used to model crash frequencies (Ivan et al., 1999, 2004). Advanced modeling approaches have been developed to cope with the complicated natures of crash distributions, residual distributions, variations in parameter estimates, heterogeneous effects of risk factors, and the effects of different exposure measures (Qin et al., 2004, 2006; Geedipally and Lord, 2010; Lord and Mannering, 2010; Xiong and Mannering, 2013).

Crash frequencies are often aggregated to daily, monthly, or yearly levels. However, aggregate crash prediction models may suffer from the ecological fallacy problem (Robinson, 1950; Golob et al., 2004), in which inferences of individual attributes may be blurred after aggregation. In contrast, the disaggregated approach using a finer sample can reduce the degree of ecological fallacy (Sullivan, 1990; Abdel-Aty and Pande, 2007). To remove the above concern and to take advantage of the availability of hourly-based crash, traffic, and weather data on 112 urban roadway segments that are evenly distributed in the Hong Kong territory over a 5-year period from 2002 to 2006, we develop disaggregated crash prediction models to measure the association between possible factors, such as road geometrics, traffic control, temporal variation, and weather, and risks of SV and MV crashes in Hong Kong. In particular, vehicle kilometer (VKM) is used as a proxy of exposure in the proposed crash prediction models (Pei et al., 2012). However, the predominance of zero counts in such a disaggregated crash prediction model based on hourly crash data is of concern, the estimates may be biased.

The problem related to the issue of excess zeros in a traditional Poisson process was recognized in previous research (Shankar et al., 1997; Washington et al., 2011). To address this issue in crash counts, alternate model formulations, such as zero-inflated count data models (Miaou, 1994; Lee and Mannering, 2002; Shankar, 2003; Huang and Chin, 2010) were proposed and adopted for crash prediction modeling. These models assumed that zero counts were derived from the dual-state process in normal- and zero-count states, which means that there are two safety states for road entities. These models often outperformed traditional count data models with a better goodness-of-fit. However, the validity of the zero-inflated model and its application in crash prediction models were criticized by Lord et al. (2005, 2007), considering the zero-generating process of zero crash count. They argued that a road link should never be judged as being in an inherently safe state and that a zero crash count could be avoided by developing a suitable and manageable database with reasonable space and time scales. In concern of the safety variation over time for each road entity, Malyshkina et al. (2009) proposed Markov switching count data models, which allow the safety state of roadway to switch between two states. Their model achieved a superior statistical fit in contrast to traditional models. However, the fundamental assumption of this modeling approach is still based on two safety states of roads.

Apart from the improvement in statistical fit, the recognition of significant contributory factors to crash risk is essential and useful in practice. Some possible risk factors may not be recognized by traditional statistical methods when the crash counts are subject to excess zeros. Bootstrap resampling approach is capable of reducing the bias in parameter estimates and standard errors of crash prediction models with excess zeros. The bootstrap method is widely used in classification trees for road safety analysis (Harb et al., 2009; Chung, 2013) but rarely in regression models. It is expected that the standard error and confidence intervals of parameters of crash regression models obtained from bootstrap resampling approaches can be improved (Efron, 1979).

The remainder of this paper is organized as follows. We first describe the study design and data collection method in Section 2. Then, we discuss the methods of analysis in

Section 3. The results are presented in Section 4 and their implications are discussed in section 5. Section 6 presents the concluding remarks and recommendations for future research.

2. STUDY DESIGN AND DATA

We first establish a comprehensive crash database containing traffic volume, road geometrics and traffic control factors, weather conditions, and temporal distribution on 112 urban roadway segments in Hong Kong using geographical information system (GIS) techniques.

Extensive traffic count data are obtained from the Hong Kong Annual Traffic Census system (Transport Department, 2002-2006), which consists of over 1,500 stations and covers 86.8% of all motorways in Hong Kong (Tong et al., 2003; Lam et al., 2003). In particular, directional traffic flows are measured continuously at 112 core stations throughout the study period. These 112 core stations are evenly and widely distributed across the territory and cover 164.6 km (i.e., 8.0%) of all of the motorways in Hong Kong. As these locations are selected for transport planning purposes, there is unlikely to be any safety-related bias. The roadway segments that are considered in this research are defined according to the standards of ATC, of which the traffic flow and geometric design characteristics are consistent throughout the segments. With respect to the time interval, as mentioned by Rothrock and Keefer (1957), it is difficult to distinguish different traffic states (free-flow or congestion regime) if more detailed traffic flow data, e.g. less than 1 hour interval is used. In concern of the traffic volume variation along time, we derive directional traffic volumes for all of the road segments adjacent to the core stations for every 4-hour period [07:00–11:00 (morning), 11:00–15:00 (noon), 15:00–19:00 (afternoon), 19:00–23:00 (evening), 23:00–03:00 (middle of the night), and 03:00–07:00 (dawn)] every day 2002 to 2006. This gives 10,956 4-hour time units and yields a sample comprising 2,230,314 observations. The VKM is obtained by multiplying road segment length by traffic volume and is used as the proxy of exposure measure for every road segment in each time period.

Crash data are obtained from the Traffic Information System (TIS) maintained by the Transport Department, which captures precise information on crash circumstances, road environment, and vehicles and casualties involved in every road crash that involves personal injury. Using the GIS technique, 7,790 crashes that occur during 2002-2006 are accordingly mapped onto 210 corresponding spatial units. As the TIS database captures information on collision types and numbers of vehicles involved, it reveals that 3,393 crashes are SV crashes (43.6%) and 4,397 are MV crashes (56.4%).

In Hong Kong, crashes are categorized into three types with respect to crash severity based on the injury severity of the most seriously injured person in a crash: fatal, serious, and slight. A fatal crash refers to a crash in which at least one person is killed immediately or is injured and subsequently dies within 30 days of the crash. A serious injury crash refers to a crash in which one or more persons are injured and detained in hospital for more than 12 hours. A slight injury crash is one in which one or more persons are injured but not to the extent that a hospital stay of more than 12 hours is required. In this study, we group fatal and serious injury crashes together as killed and seriously injured (KSI) crashes in the subsequent analysis. We segregate the dataset into two with respect to crash severity levels: KSI and slight injury crashes. Of the 7,790 crashes, 1,634 (21.0%) are KSI crashes, of which 859 are SV crashes and 775 are MV crashes, and 6,156 (79.0%) are slight injury crashes, of which 2,534 are SV crashes and 3,622 are MV crashes.

Road geometric designs and traffic controls are also incorporated in the study, specifically including lane-changing opportunities (which refers to the total number of possible lane-cuttings based on those set out by different lane markings, see Pei et al., 2012); average lane width; road curvature (average change in angle); uphill gradient and downhill gradient; number of junctions (including ramps, signal junctions, yield junctions, stop sign junctions, and roundabouts); the presence of a central divider; the presence of a hard shoulder; the presence of a bus stop; and the presence of on-street parking. In addition, rainfall is also an important environmental factor likely to have a significant influence on road safety due to its effects on driver visibility and vehicle braking performance. In this study, detailed rainfall data for each geographical location and period are obtained from the Hong Kong Observatory. To represent the unobserved heterogeneity among time periods, the influences of temporal distribution on crash risk, in terms of year, day of the week, and time of day, are also controlled for.

Table 1 summarizes the characteristics of the 2,230,314 observations analyzed in the proposed model.

TABLE 1 Summary of the observations incorporated into the proposed model

	Min	Max	Mean	S.D.	Count	Proportion
Number of observations = 2,230,314						
Slight injury crashes						
Slight-SV crashes	0	2	0.002	0.04		
Slight-MV crashes	0	3	0.002	0.05		
KSI crashes						
KSI-SV crashes	0	2	0.0004	0.02		
KSI-MV crashes	0	2	0.0004	0.02		
Traffic volume (veh)	27	26,745	3,684	3,990		
Road length (km)	0.15	9.07	1.47	1.55		
Ln (VKM)	3.55	11.55	7.56	1.55		
Rainfall	0	216	1.04	5.90		
Lane changing opportunity	0	7.8	2.43	1.61		
Average lane width	2.40	7.30	3.63	0.64		
Curvature	0	85	21.92	17.54		
Uphill gradient	0	0.11	0.01	0.02		
Downhill gradient	0	0.11	0.01	0.02		
Number of junctions	0	33	4.48	4.21		
Presence of central divider					1,570,398	70.4%
Presence of hard shoulder					298,008	13.4%
Presence of bus stop					1,419,924	63.7%
Presence of on-street parking					1,145,280	51.4%
Year 2002					424,494	19.0%
Year 2003					426,696	19.1%
Year 2004					461,160	20.7%
Year 2005					459,900	20.6%
Year 2006					458,064	20.5%
Monday					291,726	13.1%
Tuesday					306,522	13.7%
Wednesday					305,442	13.7%
Thursday					307,950	13.8%

	Min	Max	Mean	S.D.	Count	Proportion
Friday					300,474	13.5%
Saturday					295,554	13.3%
Sunday or public holiday					422,646	19.0%
07:00–11:00					371,719	16.7%
11:00–15:00					371,719	16.7%
15:00–19:00					371,719	16.7%
19:00–23:00					371,719	16.7%
23:00–03:00					371,719	16.7%
03:00–07:00					371,719	16.7%

A multicollinearity test for the independent variables is conducted prior to the determination of association measures for crash occurrence. In the multicollinearity test, if the VIF value of any independent variable is greater than 10, that variable is removed from the model to avoid biased parameter estimates. The results of the multicollinearity test reveal that the VIF values of the independent variables are all less than 10. Therefore, no evidence is established for the existence of a multicollinearity problem.

3. METHOD

3.1 Count data modeling approach

The objective of this study is to estimate and differentiate the effects of contributory factors such as geometric design, weather conditions, and temporal distribution on the incidence of SV and MV crashes at different severity levels, controlling for the effect of exposure. As the number of crashes is discrete, non-negative, and random, count data models are considered in this study (Washington et al., 2011). The Poisson regression model, the basic model form of count data models, is usually used as a starting point in the development of crash predictive models (Miaou et al., 1993). Accordingly, the crash frequency for each road entity in a given period can be assumed to follow a Poisson distribution, with the probability function defined as

$$P(y_{it}) = \exp(-\lambda_{it})\lambda_{it}^{y_{it}}/y_{it}!,$$

where y_{it} is the observed number of crashes at the *i*th entity in period t, and λ_{it}, the expected number of crashes at the *i*th entity for period *t*, which represents the log-linear relationship between expected crash counts and possible risk factors as $\lambda_{it} = \exp(\boldsymbol{\beta}\mathbf{X}_{it})$, where $\mathbf{X}_{it}$ is the vector of possible risk factors, and $\boldsymbol{\beta}$ the vector of corresponding coefficients.

When the data are subject to over-dispersion (i.e., the variance of the count is significantly greater than its mean), a gamma-distributed error term with mean and variance of 1, and α should be incorporated into the Poisson parameter to account for the over-dispersion. The Poisson-gamma model, which is also known as the negative binomial (NB) model, is widely used to deal with the over-dispersion problem (Miaou, 1994; Poch and Mannering, 1996; Milton and Mannering, 1998).

Hence, the probability function of the NB model can be formulated by

$$P(y_{it}) = \frac{\Gamma[(1/\alpha) + y_{it}]}{\Gamma(1/\alpha) y_{it}!} \left(\frac{1/\alpha}{(1/\alpha) + \lambda_{it}} \right)^{1/\alpha} \left(\frac{\lambda_{it}}{(1/\alpha) + \lambda_{it}} \right)^{y_{it}},$$

where $\Gamma(\cdot)$ is a gamma function, y_{it} is the observed number of crashes at the *i*th entity in period *t*, λ_{it} is still linked with the explanatory factors with a log-linear function, as $\lambda_{it} = \exp(\boldsymbol{\beta}\mathbf{X}_{it})$, and α is referred to as the over-dispersion parameter.

Logarithmic transformation of exposure Ln(VKM) is included in the model, and the crash frequency can be given by,

$$\lambda_{it} = \exp(\boldsymbol{\beta X_{it}} + \theta \mathrm{Ln}(\mathrm{VKM}_{it})) = \mathrm{VKM}_{it}^{\theta} * \exp(\boldsymbol{\beta X_{it}}).$$

Therefore, the crash rate/risk is given by

$$risk_{it} = \lambda_{it} \,/\, \boldsymbol{exposure}_{it} = \mathrm{VKM}_{it}^{\theta-1} \exp(\boldsymbol{\beta X_{it}}).$$

The coefficients can be estimated by the maximum likelihood method by maximizing the logarithm of the likelihood function, which is formulated as a joint density function for all observations. The chi-square distributed likelihood ratio (*LR*) statistic is commonly used to assess the goodness-of-fit of the maximum likelihood estimation model. A significant *LR* statistic indicates a good fit for the proposed model.

3.2 Bootstrap resampling method

To resolve the problem of bias in parameter estimates and standard errors of crash prediction models with excess zeros, a bootstrap resampling approach is proposed.

Generally, bootstrapping provides a resampling simulation approach to estimate standard errors and other measures of statistical precision by repeatedly and randomly sampling subsets of data from the original dataset. The bootstrap method was first introduced in 1979 to estimate the variance of sample mean (Efron, 1979) and was then applied in more complicated problems, such as the parametric model and in estimating regression parameters (Efron and Tibshirani, 1993).

We consider a general one-sample problem like this. Let $R(\mathbf{X}, F)$, a function of $\mathbf{X}$, be a random variable of interest, where $\mathbf{X} = (X_1, X_2, ..., X_n)$ indicates the entire independent and identically distributed sample $X_1, X_2, ..., X_n$ from a population having $F(x)$ as the distribution function. On the basis of having observed $\mathbf{X} = \mathbf{x}$, we wish to estimate some characteristic $Q(R)$ of the distribution of R. Usually, the function of $Q(R)$ does not have an explicit form and a Monte Carlo algorithm is used to proceed to the bootstrap method.

1. Let $\hat{F}$ be the MLE of F (i.e., it assigns probability mass 1/n at each observation x_i, i=1,2,…,n).
2. Draw a bootstrap sample from $\hat{F}$, namely $X_1^*, X_2^*, ..., X_n^*$ distributed identically as $\hat{F}$, and compute $R^*(\mathbf{X}^*, \hat{F})$.

3. Independently repeat step 2 $\boldsymbol{B}$ times (where $\boldsymbol{B}$ is large) and obtain the bootstrap replications $R_1^*, R_2^*, ..., R_n^*$ and calculate $Q(R^*)$.

The bootstrap method also has the potential to provide a more accurate estimate for regression problems, which are one of the most important and popular applications (Shao and Tu, 1995; Davison and Hinkley, 1997). The traditional bootstrap method draws the bootstrap sample randomly and evenly with an identical probability with replacement. In this study, due to concern for a possible bias caused by an imbalanced database with excess zeros, we use a resampling method to neutralize the inferences of excess zeros on the precision of the parameter estimates and standard errors for the effects of contributory factors on crash risk. In particular, a three-step resampling approach (Andrews and Buchinsky, 2000; Chernick and Labudde, 2011) can be set out as follows.

1. Divide the database into two strata according to the occurrence of crashes. Thus, two strata, one consisting of the zero crash counts and the other consisting of the non-zero crash counts, are defined.
2. Randomly draw k samples (where k is the number of observations with non-zero crash counts) from both the zero crash count and non-zero crash count strata in each bootstrap replication. A subset of observations with balanced zero and non-zero crash counts is thus extracted and a Poisson and/or Poisson-gamma regression is conducted to compute the estimates of $\boldsymbol{\beta}$.
3. Repeat step 2 by drawing samples of another k observations by 1,500 times. Then, calculate the bootstrap standard errors and percentile confidence intervals of $\boldsymbol{\beta}$ based on the 1,500 estimates.

To trade-off between prediction performance and computation time, optimum bootstrap simulation at 1500 was used (Andrews and Buchinsky, 2000). Results indicated that parameter estimates converge when the number of simulation increased up to 1,500.

Compared to point estimates of parameters by the nominal maximum likelihood approach, the bootstrap method is capable of producing more reliable parameter estimates with smaller variance for a dataset with imbalanced observed outcomes. This is essential for crash prediction models with excess zeros.

4. RESULTS

4.1 Overall

Separate crash prediction models are developed to identify the possible factors related to the respective risks of SV and MV crashes. As shown in Table 2, association measures for slight injury crashes and KSI crashes are conducted. As the slight injury crash counts, both of SV and MV crashes, are subject to over-dispersion, NB regression models are used. A bootstrap resampling method is conducted to obtain more accurate standard errors and confidence intervals based on 1,500 bootstrap replications. Table 2 presents the results of parameter estimates with the 95% confidence intervals obtained by both the nominal maximum likelihood estimation method and bootstrap resampling method.

TABLE 2 Results of crash frequency prediction models

(a) Slight injury crash prediction by the NB regression model

	Slight injury crash					
	SV crash			MV crash		
	β	(95% CI)	(Bootstrap 95% CI)	***β***	(95% CI)	(Bootstrap 95% CI)
Constant	-9.62	(-10.09, -9.15)*	(-9.94, -9.30)*	-11.36	(-11.79, -10.92)*	(-11.68, -11.03)*
Ln(VKM)	0.26	(0.22, 0.30)*	(0.23, 0.29)*	0.56	(0.52, 0.60)*	(0.53, 0.59)*
Rainfall	0.01	(0.01, 0.01)*	(0.01, 0.01)*	0.01	(0.01, 0.01)*	(0.01, 0.01)*
Lane changing opportunity	0.07	(0.04, 0.10)*	(0.05, 0.09)*	0.12	(0.10, 0.15)*	(0.11, 0.14)*
Average lane width	-0.02	(-0.09, 0.05)	(-0.06, 0.03)	-0.08	(-0.16, -0.01)*	(-0.14, -0.03)*
Curvature	-0.002	(-0.004, 0.001)	(-0.004, 0.001)	0.001	(-0.001, 0.003)	(-0.001, 0.003)
Uphill gradient	-4.74	(-7.45, -2.02)*	(-6.64, -2.84)*	-2.40	(-4.80, 0.001)	(-4.04, -0.76)*
Downhill gradient	-0.88	(-3.45, 1.70)	(-2.55, 0.80)	0.11	(-2.18, 2.40)	(-1.42, 1.65)
Number of junctions	0.04	(0.03, 0.05)*	(0.03, 0.04)*	0.01	(0.001, 0.02)*	(0.004, 0.01)*
Presence of central divider	-0.13	(-0.27, 0.01)	(-0.23, -0.03)*	-0.43	(-0.56, -0.30)*	(-0.52, -0.34)*
Presence of hard shoulder	-0.09	(-0.23, 0.06)	(-0.19, 0.02)	-0.43	(-0.53, -0.32)*	(-0.50, -0.35)*

Presence of bus stop	0.43	(0.33, 0.54)*	(0.36, 0.51)*	0.39	(0.30, 0.47)*	(0.33, 0.45)*
Presence of on-street parking	0.20	(0.08, 0.32)*	(0.12, 0.29)*	0.17	(0.07, 0.28)*	(0.09, 0.25)*
Year 2002		(control)				
Year 2003	-0.04	(-0.16, 0.09)	(-0.12, 0.05)	-0.08	(-0.18, 0.02)	(-0.15, -0.01)*
Year 2004	-0.19	(-0.32, -0.07)*	(-0.28, -0.11)*	-0.19	(-0.29, -0.08)*	(-0.26, -0.12)*
Year 2005	-0.10	(-0.22, 0.02)	(-0.18, -0.02)*	-0.13	(-0.23, -0.03)*	(-0.20, -0.06)*
Year 2006	-0.09	(-0.21, 0.03)	(-0.18, -0.01)*	-0.01	(-0.11, 0.09)	(-0.07, 0.06)
Monday	-0.05	(-0.19, 0.10)	(-0.15, 0.05)	0.03	(-0.09, 0.15)	(-0.05, 0.11)
Tuesday	0.01	(-0.13, 0.14)	(-0.09, 0.10)	0.03	(-0.09, 0.15)	(-0.05, 0.11)
Wednesday	-0.05	(-0.19, 0.09)	(-0.15, 0.05)	0.08	(-0.04, 0.20)	(-0.01, 0.16)
Thursday	-0.12	(-0.26, 0.02)	(-0.22, -0.02)*	0.06	(-0.06, 0.18)	(-0.02, 0.14)
Friday	-0.08	(-0.22, 0.06)	(-0.18, 0.02)	0.14	(0.02, 0.25)*	(0.06, 0.22)*
Saturday	0.06	(-0.08, 0.19)	(-0.03, 0.15)	0.17	(0.05, 0.29)*	(0.09, 0.25)*
Sunday or public holiday		(control)				
07:00-11:00	0.56	(0.39, 0.73)*	(0.43, 0.68)*	0.58	(0.42, 0.74)*	(0.45, 0.71)*
11:00-15:00	0.48	(0.31, 0.65)*	(0.35, 0.61)*	0.47	(0.31, 0.63)*	(0.34, 0.60)*
15:00-19:00	0.59	(0.42, 0.76)*	(0.47, 0.71)*	0.44	(0.28, 0.61)*	(0.31, 0.57)*
19:00-23:00	0.29	(0.11, 0.46)*	(0.16, 0.41)*	0.39	(0.22, 0.55)*	(0.26, 0.52)*
23:00-03:00	0.24	(0.06, 0.41)*	(0.10, 0.37)*	0.31	(0.14, 0.48)*	(0.18, 0.45)*
03:00-07:00		(control)				
LR statistic	696			2097		
α	2.79			2.11		

(b) KSI crash prediction by the Poisson regression model

	KSI crash					
	SV crash			MV crash		
	$\boldsymbol{\beta}$	(95% CI)	(Bootstrap 95% CI)	$\boldsymbol{\beta}$	(95% CI)	(Bootstrap 95% CI)
Constant	-10.66	(-11.45, -9.87)*	(-11.24, -10.09)*	-12.95	(-13.91, -11.99)*	(-13.67, -12.24)*
Ln(VKM)	0.29	(0.22, 0.37)*	(0.24, 0.35)*	0.69	(0.61, 0.78)*	(0.63, 0.76)*
Rainfall	0.01	(-0.01, 0.01)	(-0.01, 0.01)	0.02	(0.01, 0.02)*	(0.01, 0.02)*
Lane changing opportunity	0.06	(0.01, 0.11)*	(0.02, 0.09)*	0.07	(0.01, 0.12)*	(0.03, 0.10)*
Average lane width	0.08	(-0.03, 0.20)	(-0.01, 0.17)	0.03	(-0.13, 0.18)	(-0.09, 0.14)
Curvature	0.003	(-0.002, 0.007)	(-0.001, 0.006)	0.001	(-0.004, 0.006)	(-0.002, 0.004)
Uphill gradient	-5.09	(-9.69, -0.49)*	(-8.44, -1.73)*	0.10	(-4.66, 4.85)	(-3.30, 3.50)
Downhill gradient	-0.94	(-5.30, 3.41)	(-4.21, 2.33)	4.32	(-0.03, 8.67)	(1.14, 7.49)*
Number of junctions	0.05	(0.04, 0.06)*	(0.04, 0.06)*	0.03	(0.02, 0.05)*	(0.02, 0.04)*
Presence of central divider	-0.14	(-0.39, 0.10)	(-0.32, 0.03)	-1.01	(-1.32, -0.70)*	(-1.24, -0.79)*
Presence of hard shoulder	0.07	(-0.17, 0.31)	(-0.10, 0.24)	0.21	(-0.02, 0.45)	(0.06, 0.37)*
Presence of bus stop	0.36	(0.18, 0.54)*	(0.23, 0.49)*	0.25	(0.06, 0.44)*	(0.12, 0.38)*
Presence of on-street parking	0.31	(0.10, 0.52)*	(0.16, 0.46)*	0.23	(-0.04, 0.50)	(0.02, 0.44)*
Year 2002		(control)				
Year 2003	-0.23	(-0.44, -0.03)*	(-0.38, -0.09)*	-0.28	(-0.49, -0.07)*	(-0.42, -0.14)*
Year 2004	-0.31	(-0.51, -0.10)*	(-0.45, -0.16)*	-0.51	(-0.73, -0.29)*	(-0.67, -0.36)*
Year 2005	-0.24	(-0.44, -0.04)*	(-0.38, -0.10)*	-0.30	(-0.51, -0.09)*	(-0.44, -0.16)*
Year 2006	-0.49	(-0.71, -0.28)*	(-0.64, -0.35)*	-0.44	(-0.66, -0.22)*	(-0.58, -0.29)*
Monday	-0.29	(-0.53, -0.04)*	(-0.47, -0.11)*	-0.06	(-0.33, 0.21)	(-0.25, 0.14)
Tuesday	-0.16	(-0.40, 0.07)	(-0.32, -0.01)*	0.07	(-0.19, 0.32)	(-0.11, 0.25)
Wednesday	-0.27	(-0.51, -0.03)*	(-0.44, -0.10)*	0.04	(-0.22, 0.30)	(-0.13, 0.22)
Thursday	-0.22	(-0.46, 0.02)	(-0.38, -0.05)*	0.12	(-0.13, 0.37)	(-0.05, 0.30)
Friday	-0.04	(-0.26, 0.19)	(-0.19, 0.12)	0.02	(-0.24, 0.28)	(-0.16, 0.19)
Saturday	-0.14	(-0.37, 0.10)	(-0.30, 0.02)	0.13	(-0.12, 0.38)	(-0.04, 0.30)
Sunday or public holiday		(control)				
07:00-11:00	-0.18	(-0.44, 0.09)	(-0.36, 0.01)	-0.49	(-0.77, -0.21)*	(-0.69, -0.29)*
11:00-15:00	-0.27	(-0.54, 0.00)	(-0.47, -0.08)*	-0.80	(-1.10, -0.51)*	(-1.01, -0.59)*
15:00-19:00	-0.14	(-0.40, 0.12)	(-0.32, 0.04)	-0.61	(-0.90, -0.32)*	(-0.81, -0.41)*
19:00-23:00	-0.24	(-0.51, 0.02)	(-0.42, -0.06)*	-0.65	(-0.94, -0.36)*	(-0.86, -0.45)*
23:00-03:00	0.15	(-0.09, 0.39)	(-0.02, 0.32)	-0.27	(-0.55, 0.01)	(-0.47, -0.08)*
03:00-07:00		(control)				
LR-statistic	224			576		

** Statistically significant at the 5% level*

As indicated by the results of the likelihood ratio test, the four proposed models all fit well with the observations at the 99% confidence level. As shown in Table 2, with the use of a bootstrap resampling approach, the standard errors of the parameters are all lower than those by the nominal maximum likelihood approach, and the ranges of 95% confidence intervals are all smaller. Therefore, the null hypothesis that the parameter is zero is more likely to be rejected in the bootstrap resampling models. Therefore, more contributory factors could be identified.

4.2 Role of exposure

The logarithmically transformed VKM is incorporated into the models to indicate the role of exposure in the association measures. As shown in Table 2, VKM significantly determines the risks of crashes regardless of collision types and crash severities, all at the 5% level. For instance, when exposure increases, the crash frequencies of slight injury-SV crash ($\beta = 0.26$), slight injury-MV crash (0.56), KSI-SV crash (0.29), and KSI-MV crash (0.69) all increase at a less than proportionate rate at the 5% level of significance. The marginal increases in crash frequency diminish when traffic volume increases given that the road segment length remains constant.

4.3 Weather conditions

As shown in Table 2, rainfall correlates significantly with the risk of crash for all collision types and crash severities, except for that of a KSI-SV crash, all at the 5% level. In particular, rainfall relates positively to the occurrence of slight injury-SV crashes ($\beta = 0.01$), slight injury-MV crashes (0.01), and KSI-MV crashes (0.02).

4.4 Geometric design and traffic control

With respect to geometric design and traffic control characteristics, certain factors including lane-changing opportunity, lane width, gradient, and the presence of local access may contribute to crash occurrence regardless of crash types. For instance, lane-changing opportunity correlates positively with the risks of slight injury-SV ($\beta = 0.07$), slight injury-MV (0.12), KSI-SV (0.06), and KSI-MV crashes (0.07), all at the 5% level of significance. The number of junctions also correlates positively with crash occurrence, regardless of crash type, at the 5% significance level, with the coefficients equal to 0.04, 0.01, 0.05, and 0.03 for slight injury-SV, slight injury-MV, KSI-SV, and KSI-MV crashes, respectively.

In contrast, the presence of a bus stop increases the frequencies of slight injury-SV (β = 0.43), slight injury-MV (0.39), KSI-SV (0.36), and KSI-MV (0.25) crashes, all at the 5% level of significance. Further, the presence of on-street parking increases the risks of slight injury-SV (0.20), slight injury-MV (0.17), KSI-SV (0.31), and KSI-MV (0.23) crashes, all at the 5% significance level.

Certain factors have significant effects on one or two particular crash types only. For instance, average lane width correlates negatively with the risk of slight injury-MV crashes ($\beta = -0.08$) at the 5% significance level. Increases in uphill gradient reduce the frequencies of slight injury-SV (-4.74), slight injury-MV (-2.40), and KSI-SV crashes (-5.09), whereas increases in downhill gradient increase the risk of KSI-MV crashes (4.32), all at the 5% significance level. As expected, the presence of central dividers reduces the risks of slight injury-SV (-0.13), slight injury-MV (-0.43), and KSI-MV (1.01) crashes, all at the 5% level of significance. A road with a hard shoulder has a lower risk of slight injury-MV crashes (-0.43) but a higher risk of KSI-MV crashes (0.21), both at the 5% significant level.

4.5 Temporal variation

Regardless of the crash types, the crash risks are generally lower during 2003-2006 than in 2002, except for the risks of slight injury-SV crashes in 2003 and slight injury-MV crashes in 2006.

As shown in Table 2, with respect to the time of day, risks of slight injury-SV and slight injury-MV crashes during dawn (03:00-07:00) are lower than those during other times of the day. In contrast, the risk of a KSI-MV crash during dawn (03:00-07:00) is higher than that during other times of the day. There is no general trend for the temporal variation in crash risks with respect to the day of the week, except that the risks of slight injury-SV crashes on Thursdays ($\beta =$-0.12) and KSI-SV crashes on Mondays (-0.29) and Wednesdays (-0.16) are lower than those on Sundays or public holidays, and the risk of a slight injury-MV crash on Fridays (0.14) and Saturdays (0.17) is higher than that on weekends or public holidays, both at the 5% level of significance. We use the factors of time of day, day of the week, and year to account for the unobserved heterogeneity across different times. However, it may not be possible to explicitly elaborate and interpret the revealed coefficient estimates of the time factors.

5. DISCUSSION

Four separate crash prediction models are established to measure the association between weather conditions, geometric designs, traffic controls, and temporal distribution, and the risks of slight injury-SV, slight injury-MV, KSI-SV, and KSI-MV crashes, based on the disaggregated traffic flow and crash data in Hong Kong during 2002 to 2006. To remove the problem of excess zeros, a bootstrap resampling approach is applied.

5.1 Role of exposure

The results indicate that crash frequencies, regardless of collision type and crash severity, increase at a less than proportionate rate with the increase in the VKM. This result implies that when exposure is incorporated into the crash prediction model, crash rates should decrease with the traffic volume. This is consistent with the findings of Rothrock and Keefer

(1957), who considered that the vehicular speed might decrease with the increase in traffic concentration, and thus the risk of losing control and other speed-related impaired behavior could be reduced. This would in turn reduce the risks of possible crashes (Mountain et al., 1996; Qin et al., 2004, 2006). Increases in the frequencies of MV crashes, both slight injury and KSI crashes, are more sensitive to the increase in VKM than are increases in the risks of SV crashes. This can be attributed to the stronger correlation between the likelihood of MV crashes and the possibilities of traffic conflicts, which are in turn more closely related to the traffic intensity. This is consistent with the findings of Mensah and Hauer (1998).

5.2 Weather conditions

Rainfall is found to relate positively to crash risk. The risks of slight injury-SV, slight injury-MV, and KSI-MV crashes remarkably increase with the increase in rainfall. This is possibly due to slippery road surfaces and low visibility, which may induce more crashes (Fridstrøm et al., 1995; Hermans et al., 2006). Furthermore, the risk of a MV crash is seemingly higher than that of a SV crash under such unfavorable conditions. An empirical study on Wisconsin interstate highways revealed that the risk of a MV crash was 2.5 times higher than that of a SV crash under rainy conditions (Jung et al., 2011). The vehicular speed is usually lower to offset the effects of lowered visibility and poor skid resistance under rainy conditions. However, the required stopping distance is increased in emergency situations and therefore the possibility of front/rear-end collisions and the risks of MV crashes will increase.

5.3 Geometric design and traffic control

Increases in lane-changing opportunities are found to correlate with the increase in crash risks, especially for MV crashes. This is likely to be attributable to the increase in the possibilities of vehicular interactions, with the expansion of permissible lane-cutting and overtaking opportunities. However, the magnitudes of the effects of lane-changing opportunity on the risks of KSI crashes are lower than those of slight injury crashes. This can be attributed to defensive driving maneuvers that offset the potential hazards induced by possible increases in lane-cutting behavior, thus reducing the impact force on collision. This may in turn reduce the crash severity.

Similarly, an increase in the average lane width correlates with a reduction in the risk of slight injury-MV crashes. This favorable finding can be attributed to the increase in room, due to the increase in road space, for defensive driving behavior, and in return avoidance of collisions between vehicles travelling on different lanes and/or on the same lane.

Gradient also plays a noticeable role in road safety. We distinguish the effects of upward and downward slopes on crash risks by incorporating two variables, upward gradient and downward gradient, in the proposed crash prediction models. The results indicate that an increase in upward slope obviously reduces the crash risks of slight injury-SV, slight injury-MV, and KSI-SV crashes. In particular, the reduction in the risk of slight injury crashes is more dramatic than that of KSI crashes. This is reasonable because vehicles will be slowed by the slope and the required stopping distance will be shorter on an upward slope. However, as the risks of MV crash are sensitive to vehicular interactions and traffic usually jams on upward sloping roads, the favorable effects on the risks of MV crash risk by the increase in upward slope may be minimal. Indeed, increases in the risk of KSI-MV crashes may be obvious due to the increase in the magnitude of the downward slope, causing both higher vehicular speed and higher traffic concentration. The above findings are consistent with those of previous studies (Shankar et al., 1995; Yu et al., 2013).

As expected, the presence of a central divider favorably reduces the risks of slight injury-SV, slight injury-MV, and KSI-MV crashes, due to the separation of opposite traffic streams. This is particularly essential for the reduction of opposing vehicle collisions, which are normally more serious. The presence of a hard shoulder reduces the risk of slight injury-MV crashes but increases that of KSI-MV crashes. A possible reason may be that road segments with hard shoulder are generally in the higher road hierarchy, e.g. freeway and major arterial road, which usually have higher speed limits. Therefore, the collision speed and thus the risk of more severe crash could have been higher. It is worth investigating the effect of speed on the risk of more severe crash when the relevant data are available in future research.

Road junctions, including ramps, signal junctions, yield junctions, stop sign junctions, and roundabouts, play an important role in traffic control through the channelization of different traffic streams by mode and direction with respect to time and space. Road junctions are essential in reducing the risks of traffic conflicts and associated crashes. The results of the present study indicate that the crash risks increase with the number of junctions on the road. In particular, the effect of the number of junctions on the risk of a SV crash is seemingly higher than that for a MV crash. This result implies that vehicles are less likely to collide with other vehicles in such circumstances. This could be attributable to the higher likelihood of loss of control, vehicle runaway, or collision with other road features due to the complicated maneuvers required at or near junctions. The presence of a bus stop also increases the risk of SV crashes. This is possibly due to frequent pedestrian activities near bus stops and thus a higher risk of vehicle-pedestrian conflicts. Indeed, of the crashes involving pedestrians in this study, over 90% occur on road segments that have at least one bus stop. Moreover, there may be potential for more conflict between buses entering or leaving bus bays and other vehicles in the main traffic stream, which would induce MV crashes. The presence of on-street parking is also associated with a higher crash risk, especially of SV crashes. This may be attributable to more frequent roadside activities near parking areas.

5.4 Temporal variation

With respect to time distribution, the crash risk from 2003 to 2006 is generally lower than in 2002. The reduction in crash risk in the latter period may be attributable to the implementation of various road safety measures, such as remedial engineering measures, enforcement against impaired driving behavior, and road safety education, during the period. In addition, factors including economic recession before 2003 could be deterministic to the variations in crash risk over the year. The reasons are, however, not explicitly revealed in the present study. An investigation of the issues related to variations in the associations between crash occurrence and possible contributory factors across years was attempted by Wong et al. (2007) and Sze et al. (2008) in Hong Kong. Noland (2003) incorporated "year" as a predictor variable in the proposed crash prediction models and revealed that year correlated negatively with the risk of fatality. However, when Noland incorporated a variable such as medical technology into the crash prediction model, no significant relationship between year and fatality risk was revealed. Regardless, the use of the dummy variable "Year *i*" may represent the effects of possible confounding factors over time, even if comprehensive information on the factors including implementation of road safety measures and advances in medical services are not available. Similarly, we use crash year as the dummy variable to proxy the corresponding heterogeneity effects on the association.

The risk of a slight injury crash during the period from 3:00 a.m. to 7:00 a.m. is found to be lower than that in all other periods, whereas the risk of KSI crashes, and especially the KSI-MV crash, during the period from 3:00 a.m. to 7:00 a.m. is found to be higher than that in all other time periods. This result implies that the risk of more severe crashes occurring during the dawn period is higher. This is possibly attributable to higher vehicular speeds, the high chance of impaired driving behavior, especially driving while fatigued, and poor lighting conditions, which is worth exploring in future study. Although it may not be possible to explicitly reveal the factors attributable to the variation in crash risk across different periods, influences of unidentified factors that vary over time can be controlled for in the proposed model by incorporating factors such as year, day of the week, and time of day.

6. CONCLUSIONS

Considering that the effects of possible factors on the risks of SV and MV crashes can be differentiated, separate crash prediction models have been developed in this study. By taking advantage of the availability of disaggregated information on traffic flow and weather conditions, highly disaggregated crash prediction models with respect to time and space could be devised. Accordingly, the effects of several time-varying factors, such as traffic volume and rainfall, on crash risk during particular period can thus be investigated in this study. To remove the problem of bias in the estimates of coefficients due to excess zeros in the highly disaggregated model, a bootstrap resampling approach has been applied to produce a more robust estimate of standard errors and confidence intervals of the parameters.

The results indicate that factors such as VKM, geometric design, weather conditions, and temporal distribution can have significant but distinctive effects on the risks of SV and MV crashes and the associated crash severity. This is essential to the formulation and implementation of cost-effective road safety measures and thus to the reduction of crash and injury risk over the long run. To further investigate the specific relationship between risk factors to crash occurrence, more explicit evidence with detailed information and carefully investigation are needed in future study.

The results of the current study also suggest that the bootstrap resampling approach is capable of dealing with an imbalanced dataset, in this case, one with excess zeros. It is worth exploring the application of the bootstrap resampling approach for crash prediction models for other road entities and collision types in future studies, given that comprehensive information on crash circumstances and vehicle characteristics are available.

ACKNOWLEDGEMENTS

The work described in this paper was supported by a grant from the National Natural Science Foundation of China (Grant No. 71301083), a grant from the National Basic Research Program of China (973 Project; No.2012CB725405), a research grant from the University Research Committee of the University of Hong Kong (Project No. 201109176069), and grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project Nos. 718811 and 717512).

REFERENCES

Andrews D. W. K. and Buchinsky M. (2000) A three-step method for choosing the number of bootstrap repetitions. *Econometrica*, **68**(1)**,** 13-51.

Abdel-Aty, M. and Pande, A. (2007) Crash data analysis: collective vs. individual crash level approach. *Journal of Safety Research*, **38**, 581-587.

Bham, G., Javvadi, B., and Manepalli, U. (2012) Multinomial logistic regression model for single-vehicle and multivehicle collisions on urban U.S. highways in Arkansas. *Journal of Transportation Engineering*, **138**(6), 786-797.

Cameron, A. and Trevedi, P. (1990) Regression-based tests for overdispersion in the Poisson model. *Journal of Econometrics*, **46**, 347-364.

Chang, L.Y. and Chen, W.C. (2005) Data mining of tree-based models to analyze freeway accident frequency. *Journal of Safety Research*, **36**, 365-375.

Chen, F. and Chen, S. (2011) Injury severities of truck drivers in single- and multi-vehicle accidents on rural highways. *Accident Analysis and Prevention*, **43**(5), 1677-1688.

Chernich, M.R. and Labudde, R.A. (2011) *An Introduction to Bootstrap Methods with Applications to R*. Hoboken, NJ: John Wiley & Sons, Inc.

Chung, Y.-S. (2013) Factor complexity of crash occurrence: An empirical demonstration using boosted regression trees. *Accident Analysis and Prevention*, **61**, 107-118.

Davison, A.C. and Hinkley, D.V. (2006) *Bootstrap Methods and their Application*. Cambridge University Press.

Efron, B. (1979) Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, **7**, 1-26.

Efron, B. and Tibshirani, R.J. (1993) *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Fréchède, B., McIntosh, A.S., Grzebieta, R., and Bambach, M.R. (2011) Characteristics of single vehicle rollover fatalities in three Australian states (2000-2007). *Accident Analysis and Prevention*, **43**(3), 804-812.

Fridstrøm, L., Ifver, J., Ingebrigtsen, S., Kulmala, R., and Thomsen, L.K. (1995) Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts. *Accident Analysis and Prevention*, **27**(1), 1-20.

Geedipally, S.R. and Lord, D. (2010) Investigating the effect of modeling single-vehicle and multi-vehicle crashes separately on confidence intervals of Poisson-gamma models. *Accident Analysis and Prevention*, **42**(4), 1273-1282.

Golob, T.F., Recker, W.W., and Alvarez, V.M. (2004) Freeway safety as a function of traffic flow. *Accident Analysis and Prevention*, **36**, 933-946.

Harb, R., Yan, X., Radwan, E., and Su, X. (2009) Exploring precrash maneuvers using classification trees and random forests. *Accident Analysis & Prevention*, **41**(1), 98-107.

Hermans, E., Wets, G., and Van den Bossche, F. (2006) Describing the evolution in the number of highway deaths by decomposition in exposure, accident risk, and fatality risk. *Transportation Research Record*, **1950**, 1-8.

Huang, H., Chin, H.C., and Haque, M. (2008) Severity of driver injury and vehicle damage in traffic crashes at intersections: A Bayesian hierarchical analysis. *Accident Analysis and Prevention*, **40**(1), 45-54.

Ivan, J.N., Pasupathy, R.K., and Ossenbruggen, P.J. (1999) Differences in causality factors for single and multi-vehicles crashes on two-lane roads. *Accident Analysis and Prevention*, **31**, 695-704.

Ivan, J. (2004) New approach for including traffic volumes in crash rate analysis and forecasting. *Transportation Research Record*, **1897**, 134-141.

Jung, S., Qin, X., and Noyce, D.A. (2010) Rainfall effect on single-vehicle crash severities using polychotomous response models. *Accident Analysis and Prevention*, **42**(1), 213-224.

Jung, S., Qin, X., and Noyce, D.A. (2011) Modeling highway safety and simulation in rainy weather. *TRB 2011 Annual Meeting*.

Jung, S., Qin, X., and Noyce, D.A. (2012) Injury severity of multi-vehicle crash in rainy weather. *Journal of Transportation Engineering*, **138**(1), 1-12.

Kim, J.-K., Ulfarsson, G.F., Kim, S., and Shankar, V.N. (2012) Driver-injury severity in single-vehicle crashes in California: A mixed logit analysis of heterogeneity due to age and gender. *Accident Analysis and Prevention*, **50**, 1073-1081.

Lam, W.H.K., Hung, W.T., Lo, H.K., Lo, H.P., Tong, C.O., Wong, S.C., and Yang, H. (2003) Advancement of the annual traffic census in Hong Kong. *Proceedings of the Institution of Civil Engineers - Transport*, **156**, 103-115.

Lee, J. and Mannering, F. (2002) Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. *Accident Analysis and Prevention*, **34**(2), 149-161.

Lord, D. and Mannering, F. (2010) The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A*, **44**, 291-305.

Lord, D., Washington, S.P., and Ivan, J.N. (2005) Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: Balancing statistical fit and theory. *Accident Analysis and Prevention*, **37**(1), 35-46.

Lord, D., Washington, S.P., and Ivan, J.N. (2007) Further notes on the application of zero-inflated models in highway safety. *Accident Analysis and Prevention*, **39**(1), 53-57.

Malyshkina, N.V., Mannering, F.L., and Tarko, A.P. (2009) Markov switching negative binomial models: an application to vehicle accident frequencies. *Accident Analysis and Prevention*, **41** (2), 217-226.

Mensah, A. and Hauer, E. (1998) Two problems of averaging arising in the estimation of the relationship between accidents and traffic flow. *Transportation Research Record*, **1635**, 37-43.

Miaou, S.P. and Lum, H. (1993) Modeling vehicle accidents and highway geometric design relationships. *Accident Analysis and Prevention*, **25**, 689-709.

Miaou, S.P. (1994) The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. *Accident Analysis and Prevention*, **26**(4), 471-482.

Milton, J. and Mannering, F. (1998) The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. *Transportation*, **25**(4), 395-413.

Mountain, L., Fawaz, B., and Jarrett, D. (1996) Accident prediction models for roads with minor junctions. *Accident Analysis and Prevention*, **28**(6), 695-707.

Noland, R.B. (2003) Traffic fatalities and injuries: The effect of changes in infrastructure and other trends. *Accident Analysis and Prevention*, **35**(4), 599-611.

Öström, M. and Eriksson, A. (1993) Single-vehicle crashes and alcohol: a retrospective study of passenger car fatalities in Northern Sweden. *Accident Analysis and Prevention*, **25**(2), 171-176.

Pei, X., Wong, S.C., and Sze, N.N. (2012) The roles of exposure and speed in road safety analysis. *Accident Analysis and Prevention*, **48**, 464-471.

Persaud, B. N. and Mucsi, K. (1995) Microscopic accident potential models for two-lane rural roads. *Transportation Research Record*, **1485**, 134-139.

Poch, M. and Mannering, F. (1996) Negative binomial analysis of intersection-accident frequencies. *Journal of Transportation Engineering*, **122**(2), 105-113.

Qin, X., Ivan, J.N., and Ravishanker, N. (2004) Selecting exposure measures in crash rate prediction for two-lane highway segments. *Accident Analysis and Prevention*, **36**(2), 183-191.

Qin, X., Ivan, J.N., Ravishanker, N., Liu, J., and Tepas, D. (2006) Bayesian estimation of hourly exposure functions by crash type and time of day. *Accident Analysis and Prevention*, **38**(6), 1071-1080.

Robinson, W.S. (1950) Ecological correlations and the behavior of individuals. *American Sociological Review*, **15**(3), 351-357.

Rothrock, C. A. and Keefer L.E. (1957) Measurement of Urban Traffic Congestion. *Highway Research Board Bulletin*, 1-11.

Shankar, V., Mannering, F., and Barfield, W. (1995) Effect of roadway geometric and environmental factors on rural freeway accident frequencies. *Accident Analysis and Prevention*, **27** (3), 371-389.

Shankar, V., Milton, J., and Mannering, F. (1997) Modeling accident frequencies as zero-altered probability processes: An empirical inquiry. *Accident Analysis and Prevention*, **29**(6), 829-837.

Shankar, V.N., Ulfarsson, G.F., Pendyala, R.M., and Nebergall, M.B. (2003) Modeling crashes involving pedestrians and motorized traffic. *Safety Science*, **41**(7), 627-640.

Shao, J. and Tu, D. (1995) *The Jackknife and Bootstrap*. New York: Springer.

Sullivan, E.C. (1990) Estimating accident benefits of reduced freeway congestion. *Journal of Transportation Engineering*, **116**(2), 167-180.

Sze, N.N., Wong, S.C., and Chan, W.F. (2008) Traffic crashes at toll plazas in Hong Kong. Proceedings of the Institution of Civil Engineers, *Transport*, **161**, 71-76.

Tong, C.O., Hung, W.T., Lam, W.H.K., Lo, H.K., Lo, H.P., Wong, S.C., and Yang, H. (2003) A new survey methodology for the annual traffic census in Hong Kong. *Traffic Engineering and Control*, **44**, 214-218.

Transport Department (2002-2006) *The Annual Traffic Census 2002, 2003, 2004, 2005, 2006.* Hong Kong: Government of HKSAR.

Washington, S.P., Karlaftis, M.G., and Mannering, F.L. (2011) *Statistical and Econometric Methods for Transportation Data Analysis, second edition*. New York: Chapman & Hall/CRC.

Wong, S.C., Sze, N.N., and Li, Y.C. (2007) Contributory factors to traffic crashes at signalized intersections in Hong Kong. *Accident Analysis and Prevention*, **39**, 1107-1113.

World Health Organization (WHO) (2009) *Global Status Report on Road Safety: Time for Action*. Department of Violence & Injury Prevention & Disability, Geneva, Switzerland.

Xie, Y., Zhao, K., and Huynh, N. (2011) Analysis of driver injury severity in rural single-vehicle crashes. *Accident Analysis and Prevention*, **47**, 36-44.

Xiong, Y. and Mannering, F. (2013) The heteroscedastic effects of guardian supervision on adolescent driver-injury severities: A finite mixture-random parameters approach. *Transportation Research Part B*, **49**, 39-54.

Yau, K.K.W. (2004) Risk factors affecting the severity of single vehicle traffic accidents in Hong Kong. *Accident Analysis and Prevention*, **36**(3), 333-340.

Yau, K.K.W., Lo, H.P., and Fung, S.H.H. (2006) Multiple-vehicle traffic accidents in Hong Kong. *Accident Analysis and Prevention*, **38**(6), 1157-1161.

Yu, R., Abdel-Aty, M., and Ahmed, M. Bayesian random effect models incorporating real-time weather and traffic data to investigate mountainous freeway hazardous factors. *Accident Analysis and Prevention*, **50**, 371-376.