Occupant-level injury severity analyses for taxis in Hong Kong: A Bayesian space-time logistic model

Fanyu Meng[a], Pengpeng Xu[a], S.C. Wong[a,1], Helai Huang[b], Y.C. Li[a]

[a] *Department of Civil Engineering, The University of Hong Kong, Hong Kong, China*

[b] *School of Traffic and Transportation Engineering, Central South University, Changsha, China*

Abstract

This study aimed to identify the factors affecting the crash-related severity level of injuries in taxis and quantify the associations between these factors and taxi occupant injury severity. Casualties resulting from taxi crashes from 2004 to 2013 in Hong Kong were divided into four categories: taxi drivers, taxi passengers, private car drivers and private car passengers. To avoid any biased interpretation caused by unobserved spatial and temporal effects, a Bayesian hierarchical logistic modeling approach with conditional autoregressive priors was applied, and four different model forms were tested. For taxi drivers and passengers, the model with space-time interaction was proven to most properly address the unobserved heterogeneity effects. The results indicated that time of week, number of vehicles involved, weather, point of impact and driver age were closely associated with taxi drivers' injury severity level in a crash. For taxi passengers' injury severity an additional factor, taxi service area, was influential. To investigate the differences between taxis and other traffic, similar models were established for private car drivers and passengers. The results revealed that

[1] Corresponding author.
E-mail addresses: fymeng91@gmail.com (Fanyu Meng), pengpengxu@yeah.net (Pengpeng Xu), hhecwsc@hku.hk (S.C. Wong), huanghelai@csu.edu.cn (Helai Huang), joeylyc@connect.hku.hk (Y.C. Li).

although location in the network and driver gender significantly influenced private car drivers’ injury severity, they did not influence taxi drivers’ injury severity. Compared with taxi passengers, the injury severity of private car passengers was more sensitive to average speed and whether seat belts were worn. Older drivers, urban taxis and fatigued driving were identified as factors that increased taxi occupant injury severity in Hong Kong.

Keywords: occupant injury severity; KSI risk; taxi safety; Bayesian hierarchical model; space-time interaction

1. Introduction

Taxis are key public transport service providers in Hong Kong, offering a personalized point-to-point service for passengers. In 2014, taxis accounted for 12% of the boardings among all public transport modes, and the number of daily taxi boardings was 950 (Transport Department, 2014). Taxi drivers were found to have a higher risk of being involved in crashes, particularly fatal ones, as their exposure to risk is relatively greater (Baker *et al.* 1976, Johnson *et al.* 1999). The Transport Department (2014) reported that 233 out of 1,000 taxis were involved in crashes in Hong Kong, second only to public light buses among all vehicular classes (compared to 15 for private cars). Both taxi-involved crash frequency and driver casualties have increased over the past decade. The number of taxis involved in crashes in 2014 was 4,211 and the driver casualty rate 37.09%, both ranking second among all vehicle types apart from private cars (Transport Department 2014). Taxi safety has become a severe problem in developed and motorized cities such as Hong Kong.

Taxis have been a subtopic of road safety studies since the 1990s, and the focus has mainly been on the psychological patterns of taxi drivers and factors affecting taxi crash risks. In a psychological sense, the unique driving behavior of taxi drivers has been attributed to aspects such as hazard perception, driving attitude and individual personalities (Burns and Wilde 1995, Machin and De Souza 2004, Rosenbloom and Shahar 2007, Shams *et al*. 2011). Rosenbloom and Shahar (2007) surveyed male taxi drivers' and nonprofessional drivers' attitudes toward traffic violation penalties, and found that nonprofessional drivers regarded traffic violation penalties as more just and appropriate than did taxi drivers. The potential hazards associated with these psychological patterns, and the significant differences between the driving attitudes of taxi drivers and nonprofessional drivers, have been identified. Other significant factors related to taxi crash risks have been explored, such as fatigued driving (Dalziel and Job 1997), use of safety measures (Routley *et al.* 2009, Sumner *et al.* 2014) and drivers' personal characteristics such as age, gender and income (Chin and Huang 2009, La *et al.* 2013). The psychological, physical and behavioral features of taxi drivers have been found to be distinct from those of nonprofessional private car drivers, and different risk factors have been identified for taxi-involved crashes, but little research has been conducted to examine the crash-related injury severity for taxis.

Lam (2004) performed Pearson chi-square tests and logistic regressions to quantify the relationship between taxi drivers' injuries and several environmental factors. Demographic factors (age and gender) were also included. Factors such as driving late at night and driving without passengers were found to have a significant effect on taxi injury. Although the study quantitatively analyzed taxi drivers' injury issues, limitations in terms of both generality and methodology remain. First, only five environmental and two demographic factors were incorporated into the model, and the influence of other factors such as taxi operational attributes and traffic information were not investigated. Second, the effects of crashes on taxi drivers and passengers can be very different, and those on passengers have rarely been analyzed. In the taxi service, the driver controls the taxi and serves the passenger, and the passenger simply accepts the service passively. Thus, two separate analyses should be conducted on an occupant level for taxi drivers and passengers, to investigate the differences in factors that influence their injury severity levels. Third, basic logistic regression is unable to capture spatial and temporal heterogeneity and spatial correlations, which have been found to be significant in crash injury severity modeling (Klassen *et al.* 2014; Chen *et al.* 2015, Wei *et al*. 2017; Xu *et al.* 2017b). A comprehensive study using a more rigorous modeling scheme and with more integrative information is therefore necessary for taxi injury severity analyses.

Unobserved heterogeneity is an issue in most road safety research cases, identified by both crash frequency and injury severity analyses. Correlations with observed factors, if not addressed in the model, will thus result in biased interpretations of the estimated parameters (Mannering and Bhat 2014). Spatial and temporal variables can address unobserved heterogeneity, and are commonly studied (Xu *et al.* 2014, Behnood and Mannering 2015, Chen *et al.* 2015, Xu and Huang 2015, Xu *et al.* 2017a). To explicitly address both spatial and temporal effects, a Bayesian hierarchical model with autoregressive priors is an effective approach (Chen *et al.* 2015, Mannering and Bhat 2014, Shaheed *et al*. 2016), as the designated error terms can simultaneously account for heterogeneity, spatial correlation and space-time interaction.

In this study, Bayesian hierarchical logistic models were established for Hong Kong taxi drivers and passengers, to estimate the possibility of them being killed or severely injured (KSI) in a taxi-involved crash. Environmental and demographic factors and traffic characteristics were collected from 2004 to 2013 and included as independent variables in the models, which were then tested for any unstructured random effect, a spatial correlation term, a temporal random effect and a space-time interaction. The model with the smallest deviance information criterion (DIC) value was selected as optimal, and the corresponding estimated posterior distributions of the parameters were discussed. Finally, the optimal models for taxi injuries and private car injuries were compared, with private cars as a benchmark for all vehicular classes.

2. Data

The datasets used in this study were obtained by integrating three comprehensive databases: the zoning system of Hong Kong, a traffic information system (TIS) database and a global positioning system (GPS) database. The primary information available and the corresponding variables extracted from each database are discussed below.

2.1. Introduction of databases

2.1.1. Zoning system

The Planning Department of Hong Kong established a zoning system with two levels, DB26 and PDZ454, based on the Territory Survey of 2011, which is commonly used for transport planning and modeling. On the DB26 level, the whole territory of Hong Kong is divided into 26 broad districts according to the land use and development features, and therefore similar traffic characteristics are expected within each district. We selected the DB26 level as the spatial panel when considering the spatial correlation and spatial heterogeneity of the occupants' injury severity.

The territory was further divided into 406 traffic analysis zones (TAZs) to enable detailed urban planning activities, consisting of 18 cross-boundary zones and 388 normal zones, which form the PDZ454 level zoning system (Meng *et al.* 2016). In the occupant-level injury severity models, the zonal average speeds and annual travel times of various vehicular classes in the 406 TAZs were used to represent the zonal traffic operation condition.

2.1.2. *Crash database*

The TIS database was established by the Transport Department of Hong Kong in collaboration with the Hong Kong Police Force (Wong *et al.* 2007). It records the vehicle attributes (vehicle class, license, age, etc.), environmental characteristics (time, location, lighting condition, weather, etc.) and casualty information (age, sex, seat occupied, etc.) of reported crashes. The TIS crash data from 2004 to 2013 was extracted, and the casualties divided into four categories: taxi driver casualties, taxi passenger casualties, private car driver casualties and private car passenger casualties. Over the studied period, 30,110 casualties in taxis were recorded (18,004 drivers and 12,106 passengers), and 37,220 casualties in private cars (21,202 drivers and 16,018 passengers). The distribution of the casualties, categorized by casualty role and severity for taxis and private cars, is shown in Table 1. To establish occupant-level injury severity models, each casualty's demographic information, the attributes of the vehicle carrying the casualty, and environmental characteristics of the crash were extracted and used as explanatory factors (see Tables 2 to 5 for the list of factors for each category of casualties).

Table 1 Distribution of driver and passenger casualties in taxis and private cars

Severity	Taxi			Private car		
	Driver	Passenger	**Total**	Driver	Passenger	**Total**
KSI	2,438 (8.1%)	1,764 (5.9%)	**4,202 (14.0%)**	2,358 (6.3%)	2,648 (7.1%)	**5,006 (13.4%)**
Slight injury	15,566 (51.7%)	10,342 (34.3%)	**25,908 (86.0%)**	18,844 (50.6%)	13,460 (36.2%)	**32,214 (86.6%)**
Total	**18,004 (59.8%)**	**12,106 (40.2%)**	**30,110 (100%)**	**21,202 (56.9%)**	**16,108 (43.3%)**	**37,220 (100%)**

Three occupant injury severity levels—killed, severely injured and slightly injured—were defined in the database. The first two levels were combined as KSI casualties, as the fatality rate of traffic crashes in Hong Kong is extremely low (Transport Department 2015). As only the individuals injured in a crash were recorded in the database, the lowest level of injury severity used in the study was “slight injury,” which was thus considered as the reference, and the dependent variable was defined as a dummy variable equaling 1 for “KSI” and 0 for “slight injury.” The frequencies of KSI for each casualty category are given in Tables 2 to 5.

2.1.3. GPS database

A GPS database was established using 460 probe taxis equipped with GPS modules in the Hong Kong road network in 2011. The time, GPS coordinates (in WGS84 format) and speed data were collected every 30 seconds. Following Pei *et al.* (2012) and Guo *et al.* (2017), a typical day was divided into six periods: 07:00-11:00 (morning), 11:00-15:00 (noon), 15:00-19:00 (afternoon), 19:00-23:00 (evening), 23:00-03:00 (midnight) and 03:00-07:00 (dawn). The GPS data were grouped according to these periods. The zonal average speed of each period in each of the 406 TAZs, described in Section 2.1.1, was calculated from the GPS database and used as an independent variable in the models. Annual zonal travel times for taxis, private cars and total traffic in each time period were also extracted from the GPS database using a modified linear data projection approach (see Meng *et al.* (2016) for details).

2.2. Descriptive statistics

From the integrated databases, four sets of dependent and independent variables were prepared for taxi driver, taxi passenger, private car driver and private car passenger casualties (see Tables 2 to 5), respectively. For the injured drivers, “driver age” and “driver sex” were included as demographic information; for the injured passengers, their age and gender and those of the drivers were considered as independent variables. The minimum and maximum values of all variables are provided, the mean values and standard deviations of the continuous variables listed, and the percentages and frequencies of observations by injury severity levels for the dummy variables shown in Tables 2, 3, 4 and 5.

Table 2 Descriptive statistics of variables for taxi drivers

Variable name	Description	Min	Max	Mean	Standard deviation	Frequency (Percentage)		
						Total	KSI	Slight injury
Dependent variable:								
Taxi driver injury severity	1 = KSI, 0 = slight injury	0	1			2,438	-	-
Continuous variables:								
Average speed	In km/h	10.0	128.0	27.0	18.6			
Zonal travel time of total traffic	In 10,000 hours	1.20	339.15	24.75	26.22			
Zonal taxi travel time	In 10,000 hours	0.0011	104.60	7.25	9.57			
Driver age		22.0	81.0	51.0	9.1			
Dummy variables:								
Time of day								
Morning	1 = 7:00-11:00, 0 = other	0	1			3,240	476(14.7%)	2,467(85.3%)
Noon	1 = 11:00-15:00, 0 = other	0	1			2,788	368(13.2%)	2,420(86.8%)
Afternoon	1 = 15:00-19:00, 0 = other	0	1			2,956	342(11.6%)	2,614(88.4%)
Evening	1 = 19:00-23:00, 0 = other	0	1			3,338	402(12.0%)	2,936(88.0%)
Midnight	1 = 23:00-3:00, 0 = other	0	1			3,450	464(13.4%)	2,986(86.6%)
Dawn (base)	1 = 3:00-7:00, 0 = other	-	-			2,232	-	-
Time of week								
Weekday	1 = weekday, 0 = weekend	0	1			12,480	1,632(13.1%)	10,848(86.9%)
No. of vehicles involved								
Single vehicle	Single-vehicle crash	0	1			1,460	332(22.7%)	1,128(77.3%)
Double vehicle	Double-vehicle crash	0	1			12,600	1,418(11.3%)	11,182(88.7%)
Multiple vehicle (base)	Multiple-vehicle crash	-	-			3,944	-	-
Weather								
Rain	1 = rain, 0 = other weather condition	0	1			3,418	450(13.2%)	2,968(86.8%)
Illuminating condition								

Daylight		0	1	4,080	558(13.7%)	3,522(86.3%)
Dim natural light	In dawn/dusk but out of street light hours	0	1	10,442	1,424(13.6%)	9,018(86.4%)
Street light (base)		-	-	3,482	-	-
Location in network						
Road section	1 = road section, 0 = intersection	0	1	12,550	1,912(15.2%)	10,638(84.8%)
Speed limit						
Low speed limit	1 = speed limit lower than 50, 0 = other	0	1	14,938	1,890(12.7%)	13,048(87.3%)
Taxi service area						
Urban taxi	1 = urban taxi, 0 = suburban taxi	0	1	14,126	1,812 (12.8%)	12,314(87.2%)
No. of seats						
Five-seat taxi	1 = five-seat taxi, 0 = four-seat taxi	0	1	17,260	2,348(13.6%)	14,912(86.4%)
Point of impact						
Front impact		0	1	7,892	1,338(17.0%)	6,554(83.0%)
Side impact		0	1	3,986	456(11.4%)	3,530(88.6%)
Back impact (base)		-	-	6,126	-	-
Vehicle age						
New vehicle	Vehicle age less than 5 years	0	1	2,008	252(12.5%)	1,756(87.5%)
Middle-age vehicle	Vehicle age between 5 and 10 years	0	1	8,820	1,244(14.1%)	7,576(85.9%)
Old vehicle (base)	Vehicle age larger than 5 years	-	-	7,176	-	-
Driver sex						
Male driver	1 = male, 0 = female	0	1	17,584	2,380(13.8%)	14,880(86.2%)
Seatbelt						
Seatbelt on	1 = wearing seatbelt when the crash took place, 0 = other	0	1	17,482	2,360(13.5%)	15,122(86.5%)

Table 3 Descriptive statistics of variables for taxi passengers

Variable name	Description	Min	Max	Mean	Standard deviation	Frequency(Percentage)		
						Total	KSI	Slight injury
Dependent variable:								
Taxi passenger injury severity	1 = KSI, 0 = slight injury	0	1			1,764	-	-
Continuous variables:								
Average speed	In km/h	10.0	128.0	26.6	17.3			
Zonal travel time of total traffic	In 10,000 hours	1.20	339.15	24.75	27.34			
Zonal taxi travel time	In 10,000 hours	0.0011	104.60	7.84	10.00			
Driver age		22.0	81.0	53.1	8.9			
Passenger age								
Dummy variables:								
Time of day								
Morning	1 = 7:00-11:00, 0 = other	0	1			2,206	360(16.3%)	1,846(83.7%)
Noon	1 = 11:00-15:00, 0 = other	0	1			2,092	292(14.0%)	1,800(86.0%)
Afternoon	1 = 15:00-19:00, 0 = other	0	1			1,974	254(12.9%)	1,720(87.1%)
Evening	1 = 19:00-23:00, 0 = other	0	1			2,622	340(13.0%)	2,282(87.0%)
Midnight	1 = 23:00-3:00, 0 = other	0	1			2,192	306(14.0%)	1,886(86.0%)
Dawn (base)	1 = 3:00-7:00, 0 = other	-	-			1,020	-	
Time of week								
Weekday	1 = weekday, 0 = weekend	0	1			8,094	1,096(13.5%)	6,998(86.5%)
No. of vehicles involved								
Single vehicle	Single-vehicle crash	0	1			1,954	272(13.9%)	1,682(86.1%)
Double vehicle	Double-vehicle crash	0	1			7,486	972(13.0%)	6,514(87.0%)

Multiple vehicle (base)	Multiple-vehicle crash	-	-	2,666	-	-
Weather						
Rain	1 = rain, 0 = other weather condition	0	1	2,516	404(16.1%)	2,112(83.9%)
Illuminating condition						
Daylight		0	1	3,040	460(15.1%)	2,580(84.9%)
Dim natural light	In dawn/dusk but out of street light hours	0	1	6,920	998(14.4%)	5,922(85.6%)
Street light (base)		-	-	2,146	-	-
Location in network						
Road section	1 = road section, 0 = intersection	0	1	8,498	1,366(16.1%)	7,132(83.9%)
Speed limit						
Low speed limit	1 = speed limit lower than 50, 0 = other	0	1	9,808	1,358(13.8%)	8,450(86.2%)
Taxi service area						
Urban taxi	1 = urban taxi, 0 = suburban taxi	0	1	10,362	1,424(13.7%)	8,938(86.3%)
No. of seats						
Five-seat taxi	1 = five-seat taxi, 0 = four-seat taxi	0	1	11,578	1,696(14.6%)	9,882(85.4%)
Point of impact						
Front impact		0	1	5,588	1,018(18.2%)	4,570(81.8%)
Side impact		0	1	1,858	286(15.4%)	1,572(84.6%)
Back impact (base)		-	-	4,660	-	-
Vehicle age						
New vehicle	Vehicle age less than 5 years	0	1	1,058	184(17.4%)	874(82.6%)
Middle-age vehicle	Vehicle age between 5 and 10 years	0	1	5,814	832(14.3%)	4,982(85.7%)
Old vehicle (base)	Vehicle age larger than 5 years	-	-	5,234	-	-
Driver sex						
Male driver	1 = male, 0 = female	0	1	11,884	1,734(14.6%)	10,150(85.4%)

Passenger sex								
Male passenger	1 = male, 0 = female	0	1			4,816	842(17.5%)	3,974(82.5%)
Seatbelt								
Seatbelt on	1 = wearing seatbelt when the crash took place, 0 = other	0	1			11,182	1,640(14.7%)	9,542(85.3%)

Table 4 Descriptive statistics of variables for private car drivers

Variable name	Description	Min	Max	Mean	Standard deviation	Frequency(Percentage)		
						Total	KSI	Slight injury
Dependent variable:								
Taxi driver injury severity	1 = KSI, 0 = slight injury	0	1			2,648	-	-
Continuous variables:								
Average speed	In km/h	10.0	128.0	29.2	18.6			
Zonal travel time of total traffic	In 10,000 hours	1.20	339.15	28.40	32.45			
Zonal taxi travel time	In 10,000 hours	0.0024	261.26	14.77	33.17			
Driver age		18.0	77.0	38.6	11.1			
Dummy variables:								
Time of day								
Morning	1 = 7:00-11:00, 0 = other	0	1			1,090	494(45.3%)	596(54.7%)
Noon	1 = 11:00-15:00, 0 = other	0	1			4,244	426(10.0%)	3,818(90.0%)
Afternoon	1 = 15:00-19:00, 0 = other	0	1			4,394	636(14.5%)	3,758(85.5%)
Evening	1 = 19:00-23:00, 0 = other	0	1			5,372	440(8.2%)	4,932(91.8%)
Midnight	1 = 23:00-3:00, 0 = other	0	1			3,700	430(11.6%)	3,270(88.4%)
Dawn (base)	1 = 3:00-7:00, 0 = other	-	-			2,402	-	-
Time of week								
Weekday	1 = weekday, 0 = weekend	0	1			15,088	1,798(11.9%)	13,290(88.1%)

No. of vehicles involved						
Single vehicle	Single-vehicle crash	0	1	3,036	534(17.6%)	2,502(82.4%)
Double vehicle	Double-vehicle crash	0	1	13,456	1,368(10.2%)	12,088(89.8%)
Multiple vehicle (base)	Multiple-vehicle crash	-	-	4,710	-	-
Weather						
Rain	1 = rain, 0 = other weather condition	0	1	3,938	486(12.3%)	3,452(87.7%)
Illuminating condition						
Daylight		0	1	6,138	672(10.9%)	5,466(89.1%)
Dim natural light	In dawn/dusk but out of street light hours	0	1	10,160	1,412(13.9%)	8,748(86.1%)
Street light (base)		-	-	4,904	-	-
Location in network						
Road section	1 = road section, 0 = intersection	0	1	16,592	2,256(13.6%)	14,336(86.4%)
Speed limit						
Low speed limit	1 = speed limit lower than 50, 0 = other	0	1	15,866	1,852(11.7%)	14,014(88.3%)
Point of impact						
Front impact		0	1	7,834	1,304(16.6%)	6,530(83.4%)
Side impact		0	1	4,856	544(11.2%)	4,312(88.8%)
Back impact (base)		-	-	8,512	-	-
Vehicle age						
New vehicle	Vehicle age less than 5 years	0	1	4,942	532(10.8%)	4,410(89.2%)
Middle-age vehicle	Vehicle age between 5 and 10 years	0	1	5,702	668(11.7%)	5,034(88.3%)
Old vehicle (base)	Vehicle age larger than 5 years	-	-	10,558	-	-
Driver sex						
Male driver	1 = male, 0 = female	0	1	16,372	2,238(13.7%)	14,134(86.3%)
Seatbelt						
Seatbelt on	1 = wearing seatbelt when the crash took place, 0 = other	0	1	20,576	2,552(12.4%)	18,024(87.6%)

Table 5 Descriptive statistics of variables for private car passengers

Variable name	Description	Min	Max	Mean	Standard deviation	Frequency(Percentage)		
						Total	KSI	Slight injury
Dependent variable:								
Taxi passenger injury severity	1 = KSI, 0 = slight injury	0	1			2,358	-	-
Continuous variables:								
Average speed	In km/h	10.0	128.0	29.4	18.7			
Zonal travel time of total traffic	In 10,000 hours	1.20	339.15	29.78	33.27			
Zonal taxi travel time	In 10,000 hours	0.0024	261.26	15.19	32.05			
Driver age		18.0	77.0	53.1	8.9			
Passenger age								
Dummy variables:								
Time of day								
Morning	1 = 7:00-11:00, 0 = other	0	1			2,336	288(12.3%)	2,048(87.7%)
Noon	1 = 11:00-15:00, 0 = other	0	1			3,196	454(14.2%)	2,742(85.8%)
Afternoon	1 = 15:00-19:00, 0 = other	0	1			4,164	540(13.0%)	3,624(87.0%)
Evening	1 = 19:00-23:00, 0 = other	0	1			3,600	444(12.3%)	3,156(87.7%)
Midnight	1 = 23:00-3:00, 0 = other	0	1			2,134	450(21.1%)	1,684(78.9%)
Dawn (base)	1 = 3:00-7:00, 0 = other	-	-			588	-	-
Time of week								
Weekday	1 = weekday, 0 = weekend	0	1			9,948	1,382(13.9%)	8,566(86.1%)
No. of vehicles involved								
Single vehicle	Single-vehicle crash	0	1			2,392	530(22.2%)	1,862(77.8%)
Double vehicle	Double-vehicle crash	0	1			9,752	1,192(12.2%)	8,560(87.8%)

Multiple vehicle (base)	Multiple-vehicle crash	-	-	3,874	-	-
Weather						
Rain	1 = rain, 0 = other weather condition	0	1	2,948	516(17.5%)	2,432(82.5%)
Illuminating condition						
Daylight		0	1	4,248	594(14.0%)	3,654(86.0%)
Dim natural light	In dawn/dusk but out of street light hours	0	1	8,522	1,340(15.7%)	7,182(84.3%)
Street light (base)		-	-	3,248	-	-
Location in network						
Road section	1 = road section, 0 = intersection	0	1	12,280	1,988(16.2%)	10,292(83.8%)
Speed limit						
Low speed limit	1 = speed limit lower than 50, 0 = other	0	1	11,826	1,628(13.8%)	10,198(86.2%)
Point of impact						
Front impact		0	1	5,948	1,176(19.8%)	4,772(80.2%)
Side impact		0	1	3,258	502(15.4%)	2,756(84.6%)
Back impact (base)		-	-	6,812	-	-
Vehicle age						
New vehicle	Vehicle age less than 5 years	0	1	3,536	422(11.9%)	3,114(88.1%)
Middle-age vehicle	Vehicle age between 5 and 10 years	0	1	4,308	622(14.4%)	3,686(85.6%)
Old vehicle (base)	Vehicle age larger than 5 years	-	-	8,174	-	-
Driver sex						
Male driver	1 = male, 0 = female	0	1	13,892	2,082(15.0%)	11,810(85.0%)
Passenger sex						
Male passenger	1 = male, 0 = female	0	1	5,668	944(16.7)	4,724(83.3%)
Seatbelt						
Seatbelt on	1 = wearing seatbelt when the crash took place, 0 = other	0	1	15,364	2,210(14.4%)	13,154(85.6%)

3. Methodology

To model the effects of various explanatory variables on the binary occupant injury severity outcome (i.e., KSI and slight injury), a logistic function was applied, following other studies regarding injury severities (Celik and Oktay 2014, Wu *et al.* 2014, Haleem and Gan 2015, Mitchell *et al.* 2015, Huang *et al.* 2016, Shaheed *et al.* 2016). Here, denote Y_{ijt} as the injury severity level of casualty i in district j in year t: $Y_{ijt} = 1$ for KSI and $Y_{ijt} = 0$ for slight injury. Denote the probability of $Y_{ijt} = 1$ as $\pi_{ijt} = \Pr(Y_{ijt} = 1)$, which is equal to a linear function of a set of independent variables with a log link according to the form of the logistic function:

$$\text{logit}(\pi_{ijt}) = \log\left(\frac{\pi_{ijt}}{1-\pi_{ijt}}\right) = \beta_0 + \sum_{p=1}^{P} \beta_p X_{ijtp}, \tag{1}$$

where X_{ijtp} is the pth independent variable, β_0 is the intercept of the model and β_p is the estimated parameter for X_{ijtp}.

To address spatial and temporal heterogeneity, a hierarchical Bayesian modeling approach with spatiotemporal effects was applied, by incorporating different combinations of three terms in the function: a structured spatial correlation effect (s_j), an unstructured spatial heterogeneity term (u_j) and a temporal heterogeneity effect (r_t). The temporal random effect within the studied period was accounted for by r_t in all three proposed models, and represented the random time trend, which varied in different years. The spatial correlation (s_j) assumed a structured correlation matrix among the 26 districts, and the spatial random effect (u_j) allowed a random normally distributed spatial trend to vary across the 26 districts. As other studies have established, the two spatial effects will not always exist at the same time (Chiou *et al.* 2014, Behnood and Mannering 2015, Chen *et al.* 2015, Xu and Huang 2015).Thus, three different combinations of s_j and u_j were incorporated into the study to test the optional forms of spatial heterogeneity in the occupant injury severity models. The three Bayesian hierarchical logistic functions were set as follows.

Model 1:

$$\log(\pi_{ijt}) = \log\left(\frac{\pi_{ijt}}{1-\pi_{ijt}}\right) = \beta_0 + \sum_{p=1}^{P} \beta_p X_{ijtp} + s_j + u_j + r_t. \tag{2}$$

Model 2:

$$\log(\pi_{ijt}) = \log\left(\frac{\pi_{ijt}}{1-\pi_{ijt}}\right) = \beta_0 + \sum_{p=1}^{P} \beta_p X_{ijtp} + s_j + r_t. \tag{3}$$

Model 3:

$$\log(\pi_{ijt}) = \log\left(\frac{\pi_{ijt}}{1-\pi_{ijt}}\right) = \beta_0 + \sum_{p=1}^{P} \beta_p X_{ijtp} + u_j + r_t. \tag{4}$$

In addition to the independent spatial and temporal terms space-time interaction is crucial, as the trend of spatial heterogeneity may change over time and vice versa (Aguero-Valverde and Jovanis 2006, Chiou and Fu 2015, Dong *et al.* 2016). DiMaggio (2015) established Bayesian hierarchical space-time models for pedestrian and cyclist injuries in New York, and found a space-time interaction effect in the injury severity analyses. A Bayesian model with unstructured and structured spatial terms, together with a space-time interaction term, was applied in this study.

Model 4:

$$\log(\pi_{ijt}) = \log\left(\frac{\pi_{ijt}}{1-\pi_{ijt}}\right) = \beta_0 + \sum_{p=1}^{P} \beta_p X_{ijtp} + s_j + u_j + (\varphi + \delta_j)\tau_t, \tag{5}$$

where φ is the mean linear time trend over all districts, δ_j is the interaction between the time and spatial effects of district j and τ_t is the period t. Therefore, Model 4 allows for a linear time trend of each district j and thus time trends can vary from district to district.

A normal conditional prior was assigned to the structured spatial correlation term s_j, and space-time interaction δ_j, as recommended by Besag *et al.* (1991):

$$s_j \sim N\left(\frac{\sum_{k\neq j}\omega_{jk}s_k}{\sum_{k\neq j}\omega_{jk}}, \frac{{\sigma_s}^2}{\sum_{k\neq j}\omega_{jk}}\right),$$
$$\delta_j \sim N\left(\frac{\sum_{k\neq j}\omega_{jk}\delta_k}{\sum_{k\neq j}\omega_{jk}}, \frac{{\sigma_\delta}^2}{\sum_{k\neq j}\omega_{jk}}\right), \tag{6}$$

where ω_{jk} is the adjacency-based first order spatial proximity matrix: if districts j and k are adjacent, $\omega_{jk} = 1$; otherwise, $\omega_{jk} = 0$. ${\sigma_s}^2$ and ${\sigma_\delta}^2$ are variance terms and their priors are commonly assumed to be an inverse-Gamma (ϵ, ϵ) distribution (Wakefield *et al.* 2000, Quddus 2008). However, the inverse-Gamma distribution has been proven sensitive to the parameter ϵ if ${\sigma_s}^2$ and ${\sigma_\delta}^2$ are close to zero (Gelman 2006, Lee 2011). Therefore, a uniform (0, 10) prior was specified for σ_s and σ_δ in this case.

For the unstructured spatial heterogeneity u_j and temporal heterogeneity r_t, normal distributions were assumed:

$$u_j \sim \text{Normal}(0, {\sigma_u}^2)$$
$$r_t \sim \text{Normal}(0, {\sigma_r}^2) \tag{7}$$

where ${\sigma_u}^2$ and ${\sigma_r}^2$ are the variances for u_j and r_t, respectively.

As the maximum likelihood estimation method is not applicable in a Bayesian hierarchical model, an alternative two-chain Markov chain Monte Carlo (MCMC) approach was applied to construct the preceding models. Non-informative priors were assumed for the estimated parameters as follows:

$$\beta_p \sim \text{Normal}(0{,}1000). \tag{8}$$

To ensure convergence of all of the parameters, the first 5,000 iterations were removed as burn-ins, and the next 5,000 used to establish the model. MCMC chains, Gelman-Rubin plots and autocorrelation plots were chosen to monitor the MCMC chains and convergence of the parameters. All the models were established using WinBUGS14 software (Spiegelhalter *et al.* 2002). The Bayesian credible interval (BCI) was provided for each estimated parameter in all four models, to indicate the practical significance of the examined parameter (Gelman *et al.* 2003). A variable is considered to significantly affect the occupant's injury severity if the 95% BCI of its estimated mean does not cover 0, and vice versa (Chen *et al.* 2015, Shaheed *et al.* 2016).

For the purpose of model comparison, DIC was calculated for each of the models. Similar to the Akaike information criterion, the DIC permits comparisons between Bayesian hierarchical models with different numbers of estimated coefficients, and calculates the deviance at each iteration (Spiegelhalter *et al.* 2002). The DIC is defined as follows:

$$DIC = D(\bar{\theta}) + 2p_D = \bar{D} + p_D, \tag{9}$$

where $\bar{\theta}$ is the posterior means of parameters of interest, $D(\bar{\theta})$ is the deviance of $\bar{\theta}$, p_D is the effective number of parameters in the model and $\bar{D}$ is the posterior mean of the deviance statistics. The lower the DIC, the better the model performed (Spiegelhalter *et al.* 2002).

4. Results

As described in Section 2, the four data subsets of taxi driver, taxi passenger, private car driver and private car passenger casualties, were prepared for Bayesian logistic modeling. The focus of this study is to analyze occupants' injury severity in taxis, so the data for taxi driver and taxi passenger casualties were modeled first, and the comparison between the four model forms proposed in Section 3 was based on the taxi driver and passenger models only. After model comparison and selection, the optimal model form was then applied to private car driver and private car passenger injuries, to treat private car as a benchmark vehicle class and investigate the similarities and differences between occupant injury severity in taxis and in private cars.

Before modeling, correlation tests were performed for all the subsets of taxi drivers, taxi passengers, private car drivers and private car passengers. The results showed that all of the correlation coefficients between the independent variables were smaller than 0.5, meaning that none of the independent variables were highly correlated in the datasets. Based on the model forms of Equation (2), (3), (4) and (5) in Section 3, four different forms of Bayesian hierarchical logistic models with autoregressive priors were then established for taxi drivers and passengers, respectively. Table 6 shows the goodness-of-fit comparison across the proposed four model forms for taxi drivers and passengers based on DIC. For taxi drivers, the model with the lowest DIC was Model 4 (DIC = 6,751.77), indicating that the addition of unstructured spatial heterogeneity, structured spatial heterogeneity and space-time interaction explained the unobserved spatial and temporal effects for taxi driver injury severity better than the other proposed forms. Similar results were also concluded for taxi passengers. Model 4 had the lowest DIC of the three for taxi passengers (DIC = 4,550.39), so it was selected as the optimal model for both driver and passenger casualties, and was also used to model the injury severity for private car drivers and passengers. The detailed coefficient estimation results of Model 4 for taxi drivers and passengers are shown in Table 7, and the results for private car drivers and passengers are shown in Table 8.

In the comparison of the four heterogeneity terms estimated in Model 4, the space-time interaction standard deviation, σ_δ, typically has a wider range for passenger injuries than driver injuries. Thus, a relatively stronger spatiotemporal effect is found for passenger injuries as their unobserved characteristics, such as their travel patterns, are relatively more random than those of the drivers. The standard deviation of temporal heterogeneity, σ_u, and spatial correlation, σ_s, both exhibit significant effects in the four subcategories. As the unobserved spatial heterogeneity effects were thoroughly addressed by the model, the following discussion on the effects of other estimated parameters can be regarded as sufficiently explicit and accurate.

Table 6 Goodness-of-fit comparison across the four model forms

Category	Model number	Heterogeneity terms	$\bar{D}$	p_D	DIC
Taxi drivers	Model 1	s, u, r	6,722.24	52.44	6,774.60
	Model 2	s, r	6,718.17	49.19	6,767.36
	Model 3	u, r	6,700.22	58.04	6,758.26
	Model 4	$\boldsymbol{s, u, \theta, \delta, \tau}$	**6,622.69**	**64.54**	**6,751.77**
Taxi passengers	Model 1	s, u, r	4,619.06	54.46	4,673.52
	Model 2	s, r	4,624.37	52.81	4,677.18
	Model 3	u, r	4,615.13	49.33	4,664.46
	Model 4	$\boldsymbol{s, u, \theta, \delta, \tau}$	**4,405.93**	**72.23**	**4,550.39**

Table 7 Estimation results of occupant injury severity models for taxi drivers and passengers

Variables	Taxi drivers					Taxi passengers				
	Mean	S.D.	95% BCI		OR	Mean	S.D.	95% BCI		OR
			Lower	Upper				Lower	Upper	
Constant	-11.610	13.690	-48.290	-1.751	-	-3.038	0.586	-4.153	-1.897	-
Average speed	0.002	0.002	-0.002	0.006	1.002	0.002	0.003	-0.004	0.007	1.002
Weekday	-0.148*	0.069	-0.282	-0.012	0.863	-0.157	0.083	-0.322	0.004	0.855
Single vehicle	-0.026	0.121	-0.266	0.213	0.975	-0.699*	0.139	-0.975	-0.431	0.497
Double vehicle	-0.597*	0.081	-0.755	-0.438	0.550	-0.698*	0.099	-0.894	-0.505	0.497
Rain	-0.255*	0.085	-0.421	-0.089	0.775	0.017	0.095	-0.171	0.204	1.017
Daylight	0.027	0.101	-0.172	0.227	1.027	-0.056	0.125	-0.300	0.188	0.946
Dim natural light	0.083	0.111	-0.135	0.300	1.086	-0.014	0.143	-0.293	0.267	0.987
Road section	0.120	0.086	-0.052	0.290	1.127	0.043	0.103	-0.157	0.244	1.043
Urban taxi	0.210	0.119	-0.031	0.442	1.234	0.468*	0.167	0.140	0.796	1.596
Five-seat taxi	-0.210	0.177	-0.562	0.137	0.811	-0.207	0.203	-0.590	0.211	0.813
Front impact	0.478*	0.077	0.328	0.629	1.613	0.620*	0.094	0.438	0.806	1.858
Side impact	0.176	0.099	-0.018	0.369	1.193	0.640*	0.125	0.392	0.883	1.897

Low speed limit	0.090	0.091	-0.092	0.264	1.094	0.217	0.112	-0.002	0.436	1.242
New vehicle	-0.014	0.123	-0.246	0.235	0.986	0.213	0.148	-0.079	0.500	1.238
Middle-age vehicle	0.140	0.107	-0.047	0.360	1.150	-0.117	0.119	-0.348	0.115	0.889
Morning	-0.204	0.141	-0.471	0.091	0.815	-0.587*	0.182	-0.947	-0.232	0.556
Noon	-0.225	0.142	-0.499	0.062	0.798	-0.544*	0.182	-0.904	-0.193	0.580
Afternoon	-0.312*	0.139	-0.585	-0.036	0.732	-0.521*	0.178	-0.871	-0.181	0.594
Evening	-0.221	0.126	-0.465	0.028	0.802	-0.510*	0.156	-0.811	-0.203	0.600
Midnight	-0.093	0.154	-0.391	0.209	0.911	-0.364	0.195	-0.747	0.020	0.695
Seatbelt on	0.132	1.042	-0.315	0.526	1.141	0.142	0.160	-0.170	0.457	1.152
Total traffic travel time	-0.001	0.003	-0.006	0.005	0.999	-0.002	0.003	-0.009	0.004	0.998
Taxi travel time	-0.021*	0.007	-0.034	-0.008	0.979	-0.010	0.007	-0.023	0.002	0.990
Driver age	0.020*	0.004	0.013	0.028	1.020	0.010*	0.005	0.001	0.019	1.010
Male driver	-0.105	0.221	-0.519	0.335	0.901	0.333	0.318	-0.284	0.960	1.396
Passenger age	-	-	-	-	-	0.014*	0.002	0.009	0.018	1.014
Male passenger	-	-	-	-	-	0.377*	0.081	0.220	0.534	1.458
σ_s	2.616*	3.235	0.055	9.766	-	1.427*	0.731	0.201	2.926	-
σ_u	3.903*	4.505	0.061	9.996	-	1.000*	0.413	0.240	1.847	-
φ	0.023	0.037	-0.038	0.098	-	-0.053	0.041	-0.131	0.025	-
σ_δ	0.197*	0.102	0.026	0.414	-	0.585*	0.134	0.365	0.889	-
DIC	**6,751.77**					**4,550.390**				

Note: BCI is Bayesian credible interval; OR is odds ratio; *: zero is not included in the 95% BCI.

5. Discussion

5.1. Occupant injury severity models for taxi drivers and passengers

The following discussion is based on the occupant-level Bayesian hierarchical injury severity model with the best performance (Model 4) for taxi drivers and passengers, given in Table 7.

Three similarities can be identified when comparing the injury severity models for taxi drivers and passengers. First, impacts to the front of the taxis will significantly increase the KSI risk for taxi drivers and passengers, compared with back-impacts. Second, the older the driver is the higher the KSI risk for both driver and passenger. Third, double-vehicle crashes are safer for both drivers and passengers in terms of injury severity, compared with multiple-vehicle crashes.

Three major differences are also worth noting. First, as a unique operational attribute for taxis, the coefficient of "urban taxi" was significantly different from zero in the model for taxi passengers (mean = 0.468, OR = 1.596). Although the posterior distributions of the two parameters overlapped, the effect of "urban taxi" on taxi drivers' injury severity was mixed, as zero is included in the 95% BCI. The mean and the two boundaries of 95% BCI for the parameter in the taxi driver model were all smaller than those in the model for taxi passengers. According to the Transport Department (2014), the percentage of total taxis in 2014 classed as urban was 84%, which are designated to operate in most areas of Hong Kong with a relatively higher price rate, while rural taxi operation is limited to outlying areas such as the New Territories and Lantau Island. Passengers riding in taxis registered in Hong Kong urban areas are more likely to be severely injured or killed: the land in urban areas is highly developed; the density of taxi trip destinations is considerably greater; and searching for destinations can lead to risky taxi behavior such as sudden changes in speed and direction, overtaking and lane changing. This agrees with the findings of Horrey and Wickens (2003). Urban taxis may therefore cause more severe passenger injuries, but the effect is weaker for taxi drivers, who as professional drivers are proficient in controlling vehicles and may instinctively take self-protective action before a crash happens.

Second, crashes involving three or more vehicles were more likely to kill or severely injure taxi drivers and passengers than single- or double-vehicle crashes, a finding consistent with the study by Celik and Oktay (2014) on injury severity in Turkey. For taxi drivers, the coefficient of single-vehicle crash was insignificant and the range of 95% BCI was close to zero (from -0.266 to 0.213), indicating that the likelihood of taxi drivers being killed or severely injured in a single-taxi crash was similar to that in a multiple-vehicle crash (base). Double-vehicle crashes made the smallest contribution to KSI risk (mean = -0.597, OR = 0.550). Single-vehicle crashes mainly occur when a driver improperly manipulates the vehicle or loses control of it, and are particularly likely when driving at night in hazardous environmental conditions with little other traffic (Martensen and Dupont 2013), as the driver may not be able to take effective measures to avoid injury. Thus, single-vehicle crashes are dangerous and tend to cause severe injury to taxi drivers.

Third, compared with dawn (03:00-07:00), other time periods were significantly negative in taxi passenger injuries except the midnight period (23:00-3:00), according to the 95% BCI. The KSI risk to taxi passengers in the daytime between 7 a.m. and 11 p.m. was comparatively lower than at dawn. The pattern was relatively blurry for taxi drivers, as only the afternoon period was significantly negative (mean = -0.312, OR = 0.732). However, the mean values of the coefficients for all five time periods in the taxi drivers' injury severity model were negative indicating that out of the base categories, dawn also incurred the highest KSI risk for taxi drivers. A night-time elevated high KSI risk has been found in other studies (Lam 2004, Rifaat *et al.* 2011, Martensen and Dupont 2013, Lee and Li 2014). Rifaat *et al.* (2011) found that driving at night (from 12:01 a.m. to 6:30 a.m.) was more likely to cause more severe injury as the poor lighting conditions greatly influenced drivers' visibility.

Fatigued driving at night is also a significant factor, and results in more severe injury. As the night shift of Hong Kong taxis normally lasts from 5 p.m. to 5 a.m., fatigued driving behavior probably occurs from midnight to dawn, considerably increasing the possibility of drivers and passengers being severely injured or killed in a crash.

"Weekday" is another significant influential factor with a negative effect on injury severity for taxi drivers (mean = -0.148, OR = 0.863). For taxi passengers, the upper boundary of 95% BCI for weekday was very slightly positive (0.004). On weekdays, traffic is typically very heavy during peak hours (Transport Department 2014), which slows the driving speed to some extent and thus lowers the injury severity level of crashes. In addition, less route and destination searching may be required on weekdays, as the majority of trips are work-based, and taxi drivers are fairly familiar with the workplaces. By contrast, trip purposes and destinations tend to be more random and uncertain on weekends. Thus, route and destination searching behavior occurs more frequently, which incurs risky driving behavior and driving fatigue.

In the study, taxi driver age was found to have similar effects on taxi drivers and passengers. In Hong Kong, the percentage of taxi drivers aged 50 or above was reported as being 67.7% by the Occupational and Health Council (2010), and 88.6% were male. Older drivers have been found to have a higher injury risk than younger ones (Hao and Daniel 2013, Wang and Delp 2014). Lam (2004) found that taxi drivers over 45 had the highest crash-related mortality and injury rate out of all age groups. The results of our study confirm previous findings and also extend them by concluding that older taxi drivers also presented a higher KSI risk for their passengers. The estimated coefficients for taxi passenger demographics indicate that older passengers are more fragile than younger ones in terms of injury severity in a crash, and that male passengers are 45.8% more likely to be killed or severely injured in a crash than females.

In terms of weather conditions, the estimated coefficient of “rain” was significantly negative when modeling taxi drivers’ injury severity (coefficient = -0.255, OR = 0.775), but became insignificantly different from zero in the model for taxi passengers. The negative effect of rain was consistent with the findings of Edwards (1998). As rain is the most common hazardous weather condition, professional taxi drivers are able to adapt to rainy conditions and control their vehicles well, by lowering their speeds and keeping a safe distance from the vehicle in front. Hence, taxi drivers’ awareness of hazards in rainy weather enabled them to protect themselves and maintain a lower KSI risk compared with other weather conditions (base).

The traffic condition was represented by the annual zonal travel time in the time period when the crash occurred. For taxi drivers, “taxi travel time” was significantly negative based on 95% BCI, indicating that the more taxis traveling in the network, the lower the KSI risk sustained. From an optimistic point of view, more taxis may actually lower the risk of death or severe injury for taxi drivers in a crash, but the perception of hazard among taxi drivers themselves may instead explain this result: they drive more cautiously if there are more taxis around, which actually signifies the hidden hazard that taxis possess.

For points of collision, a hit from the back (base) was found to be the safest for both taxi drivers and passengers based on the 95% BCI. For taxi drivers, an impact to the front of the car was the severest, as the collision point was close to the driver’s seat position (Lee and Li 2014). For taxi passengers, side and front impacts presented similar KSI risks. This corresponds with the location of the passenger seats in the rear of the car.

5.2. Comparison of taxis and private cars

To investigate the distinctiveness of taxi injury severity, private car was chosen as the benchmark vehicle type, as it engages the largest proportion of daily traffic and is the most representative of all vehicle classes. Two Bayesian hierarchical models with autoregressive priors were established for private car drivers and passengers, using the optimal form (Model 4) with space-time interaction considered (see Table 8), and the same sets of variables as the models for taxi injuries.

Compared with the injury severity patterns of taxi drivers, distinct factors with a significant effect on the injury severity of private car drivers were found. Two major differences were detected in this comparison: the effects of "road section" and "time periods of day." First, road sections were significantly positive for private car drivers based on 95% BCI (mean = 0.352, OR = 1.421). For private car drivers, crashes occurring on road sections are 42.1% more likely to lead to KSI than those at intersections, but this trend does not hold true for taxi drivers. Shaheed *et al.* (2016) found a 44% decrease in the probability of occupants getting injured at intersections than on road sections. Private car drivers are comparatively more obedient to traffic rules and more conservative while driving, so the segregation function of signalized intersections is effective and the severity of resulting injury from crashes is reduced. Conversely, truck drivers were reported to experience higher KSI risk at intersections than other highway locations (Khorashadi *et al.* 2005). Like truck drivers, taxi drivers are professionals and considered more confident in their driving skills than ordinary private car drivers, and thus may behave aggressively both in road sections and at intersections, which blurs the differences in the KSI risk between those two locations. Second, slight differences in the effects of various periods were found between the models for taxi and private car driver injury severity. Unlike taxi drivers, private car drivers exhibit a relatively consistent injury pattern between various time periods. The potential higher KSI risk at night has been proven (Rifaat *et al.* 2011 and Lam 2004), so the blurry pattern found for taxi drivers underscored the high risk they face during the daytime (except in the afternoon as the coefficient of “afternoon” is significantly negative).

In terms of injury severity of passengers, taxi and private cars exhibited substantial differences according to the modeling results, two of which were important and worth noting. First, zonal average speed was significant in the model for private car passengers, so the effect of driving speed on private car passenger injuries is crucial. Hao and Daniel (2013) concluded that vehicles at speeds of over 50 mph exhibited higher KSI risk at a highway-rail grade crossing than those at lower speeds, and that low speed limits significantly reduced the likelihood of fatality or severe injury. Second, whether a seat belt was worn proved to be a significant factor, and reduced the risk of a passenger in a private car being killed or severely injured in a crash. This result is intuitive and has been reported in other studies (Chen *et al.* 2016, Shaheed *et al.* 2016), but it did not apply to injuries in taxis. Seat belts only work in extremely hazardous situations, such as when sudden braking or collision stops the vehicle. Experienced taxi drivers may evade such dangerous situations by properly controlling their vehicles, which not only protects passengers from being severely injured but also results in a weakened effect of seat belts.

Table 8 Estimation results of occupant injury severity models for private car drivers and passengers

Variables	Private car drivers					Private car passengers				
	Mean	S.D.	95% BCI		OR	Mean	S.D.	95% BCI		OR
			Lower	Upper				Lower	Upper	
Constant	4.668	10.410	-2.588	32.680	-	6.849	11.160	-1.561	37.380	-
Average speed	-0.003	0.002	-0.007	0.001	0.997	0.004*	0.002	0.000	0.008	1.004
Weekday	-0.155*	0.066	-0.283	-0.025	0.857	-0.155*	0.069	-0.290	-0.018	0.857
Single vehicle	-0.288*	0.100	-0.484	-0.093	0.750	-0.209	0.113	-0.430	0.010	0.811
Double vehicle	-0.551*	0.077	-0.700	-0.401	0.576	-0.411*	0.087	-0.581	-0.242	0.663
Rain	-0.137	0.079	-0.294	0.018	0.872	0.177*	0.084	0.013	0.342	1.194
Daylight	-0.096	0.086	-0.266	0.074	0.908	0.055	0.103	-0.147	0.256	1.057
Dim natural light	-0.028	0.094	-0.213	0.156	0.972	0.083	0.110	-0.135	0.297	1.087
Road section	0.352*	0.089	0.176	0.525	1.421	0.270*	0.094	0.087	0.455	1.310
Front impact	0.587*	0.074	0.444	0.732	1.799	0.743*	0.080	0.586	0.902	2.103
Side impact	0.293*	0.089	0.118	0.466	1.341	0.600*	0.097	0.410	0.790	1.822
Low speed limit	-0.101	0.071	-0.241	0.040	0.904	-0.195*	0.080	-0.350	-0.040	0.823
New vehicle	-0.127	0.080	-0.283	0.029	0.881	-0.161	0.092	-0.344	0.017	0.851
Middle-age vehicle	-0.111	0.073	-0.256	0.030	0.895	0.030	0.081	-0.129	0.188	1.031
Morning	-0.520*	0.145	-0.803	-0.233	0.594	-0.901*	0.184	-1.263	-0.545	0.406
Noon	-0.712*	0.145	-0.996	-0.429	0.491	-0.678*	0.179	-1.027	-0.336	0.508
Afternoon	-0.492*	0.135	-0.755	-0.227	0.612	-0.773*	0.166	-1.102	-0.455	0.462
Evening	-0.509*	0.138	-0.781	-0.239	0.601	-0.961*	0.162	-1.283	-0.647	0.382
Midnight	-0.072	0.139	-0.345	0.199	0.931	-0.542*	0.161	-0.862	-0.226	0.582
Seatbelt on	-0.126	0.160	-0.422	0.195	0.882	-0.547*	0.147	-0.826	-0.256	0.579
Total traffic travel time	-0.001	0.001	-0.003	0.001	0.999	0.004*	0.002	0.001	0.008	1.004
Private car travel time	0.000	0.001	-0.002	0.002	1.000	-0.007*	0.003	-0.013	-0.002	0.993
Driver age	0.010*	0.003	0.004	0.015	1.010	-0.015*	0.003	-0.021	-0.008	0.986
Male driver	0.449*	0.084	0.286	0.613	1.567	0.037	0.105	-0.158	0.247	1.038
Passenger age	-	-	-	-	-	0.008*	0.002	0.004	0.012	1.008
Male passenger	-	-	-	-	-	0.083	0.073	-0.062	0.225	1.087
σ_S	1.219*	1.467	0.027	4.696	-	2.294*	2.142	0.141	7.638	-

σ_u	3.750*	4.426	0.030	9.992	-	4.413*	4.478	0.086	9.994	-
φ	0.041	0.024	-0.007	0.088	-	0.091*	0.028	0.037	0.147	-
σ_δ	0.125*	0.106	0.013	0.352	-	0.363*	0.114	0.178	0.619	-
DIC	**7715.44**					**6267.720**				

Note: BCI is Bayesian credible interval; OR is odds ratio; *: zero is not included in the 95% BCI.

6. Conclusion

In this study, a comprehensive Bayesian hierarchical logistic modeling approach with conditional autoregressive priors was adopted to analyze taxi occupants' injury severity in Hong Kong. As unobserved spatial and temporal effects had proved to be influential in previous injury severity analyses, four model forms were proposed to incorporate different combinations of an unstructured spatial effect, a structured spatial correlation, an unstructured temporal effect and a space-time interaction effect. Independent variables were extracted from several comprehensive databases and included in the models, to investigate the potential factors that had significant effects on the risk of taxi drivers and passengers being killed or severely injured. Given that the effects on taxi drivers and passengers of a crash are distinct, two separate Bayesian logistic models were established for taxi drivers and passengers. The model with space-time interaction (Model 4 in Section 3) was found to outperform the other three model forms, as it always had the lowest DIC value. The space-time interaction could thus better address the unobserved heterogeneity in the database. The optimal model form was then applied to model the injury severity for private car drivers and passengers, respectively, as a benchmark vehicle class.

Several noteworthy results were produced by the injury severity models for taxi drivers and passengers. Urban taxis, single-vehicle crashes and various time periods were the three dominant factors significantly influencing taxi passengers' KSI risk. However, rain significantly decreased the risk of mortality and severe injury of taxi drivers compared with other weather conditions, but seldom had a significant effect on the severity of taxi passenger injuries. The different characteristics of taxi drivers and passengers could explain these results. As taxi drivers control the vehicles, they are often aware of hazards earlier than passengers and take action to avoid hazardous conditions. Therefore, although taxi drivers are exposed to more dangerous conditions most of the time, they are still able to protect themselves intuitively before a crash happens. In addition, this study found that as professional drivers, taxi drivers were able to adapt to hazardous conditions well, and that fatigued driving was one of the main hidden hazards that resulted in death or severe injury in a crash.

The differences in injury severity of private car drivers compared with taxi drivers were explained mainly by “road section,” which tended to increase injury severity. The pattern of effects among various time periods were also more explicit for private car drivers than taxi drivers. The aggressive driving attitudes and the specific working hours of taxi drivers resulting from different work shifts were able to explain these differences. For passengers, driving speed significantly influenced the level of injuries in private cars, but speed did not have a significant effect on taxi passenger injuries. Whether a seat belt was worn was identified as a factor that reduced the risk of private car passengers being killed or severely injured and was also a major difference between taxi and private car passengers.

Based on the modeling results, several policies in taxi safety related management are suggested. First, specific policies should be implemented for “urban taxis” registered in Hong Kong, as the KSI risk for passengers was found to substantially increase for urban taxis. This high KSI risk must therefore be compensated by adjusting the working hours of the drivers, or distributing more urban taxis to weaken the fatigued effects resulting from serving passengers in urban areas. Second, the problem of aging taxi drivers in Hong Kong, and the significant positive effect of “driver age” in the model for taxi drivers, can be regarded as warning signal for policy makers and transport planners in Hong Kong. A review of the age limit in the taxi driver registry is a possible measure to alleviate this situation. Third, the two-shift driving schedule of Hong Kong taxi drivers needs modification, as long shifts may result in fatigued driving and lead to more single-taxi crashes, in which the KSI risk of taxi passengers is extremely high.

Further studies can compare the injury severity of taxi drivers and other professional drivers, such as truck and bus drivers, to investigate the similarities and differences between influential factors. Information regarding at-fault taxi drivers can also be incorporated into the models, as different mechanisms apply to hitting and being hit in a crash. Establishing a bivariate model would also be beneficial, to account for the dependence between injury severities of taxi drivers and passengers in the same crash/vehicles.

Acknowledgments

This study was supported by a Research Postgraduate Studentship, and grants from the University Research Committee of the University of Hong Kong (201511159015), and the Joint Research Scheme of the National Natural Science Foundation of China/Research Grants Council of Hong Kong (project Nos. 71561167001 and N_HKU707/15). We would like to express our special thanks to the Transport Department of the Hong Kong Special Administrative Region for providing the TIS and ATC data, and to Concord Pacific Satellite Technologies Limited and Motion Power Media Limited for providing the GPS taxi data.

References

Aguero-Valverde, J., Jovanis, P.P., 2006. Spatial analysis of fatal and injury crashes in Pennsylvania. Accident Analysis & Prevention 38 (3), 618-625.

Baker, S.P., Wong, J., Baron, R.D., 1976. Professional drivers: Protection needed for a high-risk occupation. American Journal of Public Health 66 (7), 649-654.

Behnood, A., Mannering, F.L., 2015. The temporal stability of factors affecting driver-injury severities in single-vehicle crashes: Some empirical evidence. Analytic Methods in Accident Research 8, 7-32.

Besag, J., York, J., Mollie, A., 1991. Bayesian image restoration, with two applications in spatial statistics. Annals of the Institute of Statistical Mathematics 43 (1), 1-20.

Burns, P.C., Wilde, G.J.S., 1995. Risk taking in male taxi drivers: Relationships among personality, observational data and driver records. Personality and Individual Differences 18 (2), 267-278.

Celik, A.K., Oktay, E., 2014. A multinomial logit analysis of risk factors influencing road traffic injury severities in the Erzurum and Kars provinces of Turkey. Accident Analysis & Prevention 72, 66-77.

Chen, C., Zhang, G., Qian, Z., Tarefder, R.A., Tian, Z., 2016. Investigating driver injury severity patterns in rollover crashes using support vector machine models. Accident Analysis & Prevention 90, 128-39.

Chen, C., Zhang, G., Tian, Z., Bogus, S.M., Yang, Y., 2015. Hierarchical Bayesian random intercept model-based cross-level interaction decomposition for truck driver injury severity investigations. Accident Analysis & Prevention 85, 186-198.

Chin, H., Huang, H., 2009. Safety assessment of taxi drivers in Singapore. Transportation Research Record: Journal of the Transportation Research Board 2114, 47-56.

Chiou, Y.C., Fu, C., 2015. Modeling crash frequency and severity with spatiotemporal dependence. Analytic Methods in Accident Research 5-6, 43-58.

Chiou, Y.C., Fu, C., Chih-Wei, H., 2014. Incorporating spatial dependence in simultaneously modeling crash frequency and severity. Analytic Methods in Accident Research 2, 1-11.

Dalziel, J.R., Job, R.F.S., 1997. Motor vehicle accidents, fatigue and optimism bias in taxi drivers. Accident Analysis & Prevention 29 (4), 489-494.

Dimaggio, C., 2015. Small-area spatiotemporal analysis of pedestrian and bicyclist injuries in new york city. Epidemiology 26 (2), 247-54.

Dong, N., Huang, H., Lee, J., Gao, M., Abdel-Aty, M., 2016. Macroscopic hotspots identification: A Bayesian spatio-temporal interaction approach. Accident Analysis & Prevention 92, 256-264.

Edwards, J.B., 1998. The relationship between road accident severity and recorded weather. Journal of Safety Research 29 (4), 249-262.

Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models. Bayesian Analysis 1, 515-533.

Gelman, A., Carlin, J.B., Stern, H.S., 2003. Bayesian data analysis (3rd edition) CRC Press.

Guo, Q., Xu, P., Pei, X., Wong, S.C., Yao, D., 2017. The effect of road network patterns on pedestrian safety: A zone-based Bayesian spatial modeling approach. Accident Analysis and Prevention 99, 114-124.

Haleem, K., Gan, A., 2015. Contributing factors of crash injury severity at public highway-railroad grade crossings in the U.S. Journal of Safety Research 53, 23-9.

Hao, W., Daniel, J., 2013. Severity of injuries to motor vehicle drivers at highway-rail grade crossings in the United States. Transportation Research Record: Journal of the Transportation Research Board 2384, 102-108.

Horrey, W.J., Wickens, C.D., 2003. Multiple resource modeling of task interference in vehicle control, hazard awareness and in-vehicle task performance. In: Proceedings of the Second International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design. Park City, UT.

Huang, H., Li, C., Zeng, Q., 2016. Crash protectiveness to occupant injury and vehicle damage: An investigation on major car brands. Accident Analysis & Prevention 86, 129-36.

Johnson, N.J., Sorlie, P.D., Backlund, E., 1999. The impact of specific occupation on mortality in the US national longitudinal mortality study. Demography 36 (3), 355-367.

Khorashadi, A., Niemeier, D., Shankar, V., Mannering, F., 2005. Differences in rural and urban driver-injury severities in accidents involving large-trucks: An exploratory analysis. Accident Analysis & Prevention 37 (5), 910-21.

Klassen, J., El-Basyouny, K., Islam, M.T., 2014. Analyzing the severity of bicycle-motor vehicle collision using spatial mixed logit models: A city of Edmonton case study. Safety Science 62, 295-304.

La, Q.N., Lee, A.H., Meuleners, L.B., Van Duong, D., 2013. Prevalence and factors associated with road traffic crash among taxi drivers in Hanoi, Vietnam. Accident Analysis & Prevention 50, 451-455.

Lam, L.T., 2004. Environmental factors associated with crash-related mortality and injury among taxi drivers in New South Wales, Australia. Accident Analysis & Prevention 36 (5), 905-908.

Lee, C., Li, X., 2014. Analysis of injury severity of drivers involved in single- and two-vehicle crashes on highways in Ontario. Accident Analysis & Prevention 71, 286-295.

Lee, D., 2011. A comparison of conditional autoregressive models used in Bayesian disease mapping. Spatial and Spatio-temporal Epidemiology 2 (2), 79-89.

Machin, M.A., De Souza, J.M.D., 2004. Predicting health outcomes and safety behaviour in taxi drivers. Transportation Research Part F: Traffic Psychology and Behaviour 7 (4-5), 257-270.

Mannering, F.L., Bhat, C.R., 2014. Analytic methods in accident research: Methodological frontier and future directions. Analytic Methods in Accident Research 1, 1-22.

Martensen, H., Dupont, E., 2013. Comparing single vehicle and multivehicle fatal road crashes: A joint analysis of road conditions, time variables and driver characteristics. Accident Analysis & Prevention 60, 466-71.

Meng, F., Wong, S.C., Wong, W., Li, Y.C., 2016. Estimation of scaling factors for traffic counts based on stationary and mobile sources of data. International Journal of Intelligent Transportation Systems Research.

Mitchell, R.J., Bambach, M.R., Toson, B., 2015. Injury risk for matched front and rear seat car passengers by injury severity and crash type: An exploratory study. Accident Analysis & Prevention 82, 171-9.

Occupational and Health Council, 2010. Report on safety condition of occupational drivers survey (part I). Green Cross, Hong Kong.

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

Quddus, M.A., 2008. Modelling area-wide count outcomes with spatial correlation and heterogeneity: An analysis of London crash data. Accident Analysis & Prevention 40 (4), 1486-1497.

Rifaat, S.M., Tay, R., De Barros, A., 2011. Effect of street pattern on the severity of crashes involving vulnerable road users. Accident Analysis & Prevention 43 (1), 276-83.

Rosenbloom, T., Shahar, A., 2007. Differences between taxi and nonprofessional male drivers in attitudes towards traffic-violation penalties. Transportation Research Part F: Traffic Psychology and Behaviour 10 (5), 428-435.

Routley, V., Ozanne-Smith, J., Qin, Y., Wu, M., 2009. Taxi driver seat belt wearing in Nanjing, China. Journal of Safety Research 40 (6), 449-454.

Shaheed, M.S., Gkritza, K., Garriquiry, A.L., Hallmark, S.L., 2016. Analysis of occupant injury severity in winter weather crashes: A fully Bayesian multivariate approach. Analytic Methods in Accident Research 11, 33-47.

Shams, M., Shojaeizadeh, D., Majdzadeh, R., Rashidian, A., Montazeri, A., 2011. Taxi drivers’ views on risky driving behavior in Tehran: A qualitative study using a social marketing approach. Accident Analysis & Prevention 43 (3), 646-51.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64 (4), 583-639.

Sumner, S.A., Pallangyo, A.J., Reddy, E.A., Maro, V., Pence, B.W., Lynch, C., Turner, E.L., Egger, J.R., Thielman, N.M., 2014. Effect of free distribution of safety equipment on usage among motorcycle-taxi drivers in Tanzania—A cluster randomised controlled trial. Injury 45 (11), 1681-6.

Transport Department, 2014. Travel characteristics survey 2011 final report. Transport Department, HKSAR, Hong Kong.

Transport Department, 2015. Road traffic accident statistics. Transport Department, HKSAR, Hong Kong.

Wakefield, J.C., Best, N.G., Waller, L., 2000. Bayesian approaches to disease mapping. In: Spatial epidemiology: Methods and applications, Oxford University Press.

Wang, P.C., Delp, L., 2014. Health status, job stress and work-related injury among Los Angeles taxi drivers. Work 49 (4), 705-712.

Wei, Z., Wang, X., Zhang, D., 2017. Truck crash severity in New York city: An investigation of the spatial and the time of day effects. Accident Analysis & Prevention 99, 249-261.

Wong, S.C., Sze, N.N., Li, Y.C., 2007. Contributory factors to traffic crashes at signalized intersections in Hong Kong. Accident Analysis & Prevention 39 (6), 1107-1113.

Wu, Q., Chen, F., Zhang, G., Liu, X.C., Wang, H., Bogus, S.M., 2014. Mixed logit model-based driver injury severity investigations in single- and multi-vehicle crashes on rural two-lane highways. Accident Analysis & Prevention 72, 105-15.

Xu, P., Huang, H., 2015. Modeling crash spatial heterogeneity: Random parameter versus geographically weighting. Accident Analysis & Prevention 75, 16-25.

Xu, P., Huang, H., Dong, N., Abdel-Aty, M., 2014. Sensitivity analysis in the context of regional safety modeling: Identifying and assessing the modifiable areal unit problem. Accident Analysis & Prevention 70, 110-120.

Xu, P., Huang, H., Dong, N., Wong, S.C., 2017a. Revisiting crash spatial heterogeneity: A Bayesian spatially varying coefficients approach. Accident Analysis & Prevention 98, 330-337.

Xu, X., Xie, S., Wong, S.C., Xu, P., Huang, H., Pei, X., 2017b. Severity of pedestrian injuries due to traffic crashes at signalized intersections in Hong Kong: A Bayesian spatial logit model. Journal of Advanced Transportation, Accepted.