Impact of data aggregation approaches on the relationships between operating speed and traffic safety

Rongjie Yu*, Ph.D.
School of Transportation Engineering, Tongji University
4800 Cao'an Road, 201804, Shanghai, China
Key Laboratory of Road and Traffic Engineering of Ministry of Education
4800 Cao'an Road, 201804, Shanghai, China
*Corresponding Author: yurongjie@tongji.edu.cn

Mohammed Quddus, Ph.D.
School of Architecture, Building and Civil Engineering,
Loughborough University, Loughborough LE11 3TU, United Kingdom

Xuesong Wang, Ph.D.
School of Transportation Engineering, Tongji University
4800 Cao'an Road, 201804, Shanghai, China
Key Laboratory of Road and Traffic Engineering of Ministry of Education
4800 Cao'an Road, 201804, Shanghai, China

Kui Yang, Ph.D. Student
School of Transportation Engineering, Tongji University
4800 Cao'an Road, 201804, Shanghai, China
Key Laboratory of Road and Traffic Engineering of Ministry of Education
4800 Cao'an Road, 201804, Shanghai, China

(*corresponding author)

June 2018

Revision submitted for possible publication in *Accident Analysis and Prevention*

Impact of data aggregation approaches on the relationships between operating speed and traffic safety

Abstract: The impact of operating speed on traffic crash occurrence has been a controversial topic in the traffic safety discipline as some studies reported a positive association whereas others indicated a negative relationship between speed and crashes. Two major issues thought to be accountable for such conflicting findings are the application of inappropriate statistical methods and the use of sample datasets with varying levels of aggregation. The main objective of this study is therefore to investigate the impacts of data aggregation schemes on the relationships between operating speed and traffic safety. A total of three aggregation approaches were examined: (1) a segment-based dataset in which crashes are grouped by roadway segment, (2) a scenario-based dataset where crashes are aggregated by traffic operating scenarios, and (3) a disaggregated crash-level dataset consisting of information from individual crashes. The first two aggregation approaches were used in examining the relationships between operating speed and crash frequency using Bayesian random-effects negative binomial models. The third disaggregated crash risk analysis was conducted utilizing Bayesian random-effects logistic regression models. From the modelling results, it has been concluded that the scenario-based approach shared similar findings with those of the disaggregated crash risk analysis approach in which a U-shaped relationship between operating speed and crash occurrence was identified. However, the commonly adopted segment-based aggregation approach revealed a monotonous negative relationship between speed and crash frequency. The implications of the different analyses results and the potential applications of the results on speed management systems have therefore been discussed.

Keywords: Speed and crashes relationship, Bayesian random-effects model, Urban expressway traffic safety, Crash aggregation approach.

Introduction

Speed management interventions are introduced to smooth traffic flow and enhance roadway capacity and safety. Such interventions primarily include fundamental speed limit settings (e.g. Fitzpatrick *et al*., 2016), Variable Speed Limits (VSL) in the Active Traffic Management Systems (e.g. Mirshahi *et al*., 2007) and safety improvement countermeasures such as traffic calming measures (e.g. Moreno and Garcia, 2013). However, both speed limit settings and countermeasure selections heavily rely on the in-depth understandings of the quantitative relationships between operating speeds and traffic safety. More specifically, studies were conducted to identify at which operating speed there is a high probability for crash occurrence and then countermeasures were further designed to alleviate or eliminate these conditions.

Given the importance of analyzing the relationships between operating speed and traffic safety, a few studies have established statistical models between operating speed and crash occurrence. However, since traffic crashes are random and sporadic events with low occurrence probabilities (AASHTO, 2010), spatio-temporal aggregations are needed when formulating the analysis datasets. During the aggregation, raw speed information captured by the traffic sensing detectors were also assembled; operating speed data prior to crash occurrence were mixed with operating speed data under normal conditions.

For instance, the widely adopted safety performance functions (SPFs) were developed using crash frequency by segment as the dependent variable (Abdel-Aty and Radwan, 2000); where raw speed data were processed to work out average speed for each segment over a certain period of time as an independent variable. Therefore, the identified relationships were basically an association between segment-level crash frequency and average operating speed in which the features of operating speeds prior to crash occurrence could not be analyzed.

Given the crash aggregation limitations, different analysis approaches have been utilized in order to unveil the effects of operating speed characteristics on crash occurrence. Table 1 has summarized a few studies with similar research objectives; comparisons were conducted from the aspects of crash data aggregation level, the nature of assembled speed information in the analyses datasets, and their primary findings.

Table 1 Literature that analyzed relationships between speed and crash occurrence

Authors & year	Crash aggregation level	Speed information assembled in the analysis	Key finding on the operating speed and crash occurrence
Taylor *et al.*, (2000)	Roadway segment	Average Speed	Excessive speed indicator is strongly and positively associated with crashes
Pei *et al.*, (2012)	Roadway segment	Average speed	The correlation between speed and crash risk is positive when distance exposure is considered, but negative when time exposure is used.
Quddus (2013)	Roadway segment	Average speed	Insignificant associations between crash rates and average speeds were identified
Yu *et al.*, (2013)	Roadway segment	Speed information prior to crash occurrence	Negative relationships between speed and crash occurrence
Elvik (2013)	Individual Crash	Speed information prior to crash occurrence	Exponential relationship between number of accidents and initial speeds
Pauw *et al.*, (2014)	Roadway segment	Speed limits	Reduced speed limits would lead to decreased crash rates
Ronshandel *et al.*, (2015)	Individual crash	Speed information prior to crash occurrence	Increasing values of speed are associated with reduced crash risk
Gargoum & El-Basyouny (2016)	Roadway segment	Average speed	Higher crash frequency is anticipated at roadway segments with higher average speeds
Imprialou *et al.*, (2016)	Traffic operating scenarios	Grouped average speed prior to crash occurrence	A quadratic relationship was revealed between operating speed and crash frequency

From Table 1 it can be seen that previous studies utilized crash aggregation at three levels: (1) segment-based, (2) scenario-based, and (3) individual-crash based. For the segment and scenario-based studies, crash frequency (or crash counts) was used as a dependent variable; while for the individual-crash based approach, the dichotomous crash and non-crash outcome was employed. Instead of using the average speed, several studies (Elvik, 2013; Yu *et al.,* 2013; Imprialou *et al.,* 2016) have tried to employ the operating speeds just prior to crash occurrence. However, the analyses conducted by using data at different levels of crash aggregation led to inconsistent results as shown in Table 1.

There is a dearth of research in investigating the reasons for conflicting findings and identifying the optimal way of integrating crash and speed data. Therefore, the purpose of this research is to identify the impacts of crash data aggregation approaches on the relationships between operating speeds and traffic safety. More specifically, the abovementioned three crash aggregation levels were compared by using speed data prior to crash occurrence.

Data from Shanghai urban expressway systems were utilized here. Firstly, the segment-based and scenario-based approaches were compared with Bayesian random-effects negative binomial models. Then, disaggregate crash risk analyses were conducted for four subgroups of crashes separately using Bayesian random-effects logistic regression modeling technique, where crashes were classified by operating speeds prior to crash occurrence. Finally, the relationships between operating speed and traffic safety were concluded. In addition, the advantages and disadvantages of the adopted aggregation approaches were discussed along with the implications of their applications on safety improvement and management.

Data Preparation

Shanghai urban expressway system was selected as the study area due to the following two reasons: (1) Shanghai urban expressway systems have relatively high-dense inductive loop detectors as a traffic sensing system with an average spacing distance of 650 meters (compared to an average of around 800 meters found in most studies (e.g. Xu *et al.*, 2013; Abdel-Aty *et al.*, 2007), which could provide high quality traffic flow data for the analyses; (2) traffic crashes occurred on the urban expressway system hold accurate crash locations and occurrence time since the crash records were checked with the full-coverage video surveillance system. Therefore, speed data prior to crash occurrence could be obtained accurately.

A total of three datasets were utilized: (1) crash data of September, 2013; (2) roadway geometric characteristics; and (3) traffic data by road segment collected by loop detectors aggregated at 2-minute interval. Crashes occurred on Shanghai urban expressways were recorded by using a stake number as reference for their location description, where stake numbers are non-repetitive marked along the roadway network. Based on the stake numbers, upstream and downstream loop detectors corresponding to crashes could be matched. In addition, considering the geometric and traffic flow features of the expressway network, roadway segments in both directions were treated as independent to each other in this study.

In order to identify the impacts of crash data aggregations on the relationships between operating speeds and traffic safety, three different levels of data aggregation were formulated: two for analyzing crash frequency and the other is to examine individual crash risk. The datasets are briefly discussed below.

Datasets for the crash frequency analyses

The pre-crash traffic conditions data were then aggregated with two different approaches for the crash frequency analyses: (1) segment-based approach and (2) scenario-based approach. The pre-crash traffic conditions were represented by a 6-minute interval operating condition (average operating speed and traffic volume) prior to each crash occurrence; the 2-minute raw traffic condition data were aggregated into 6-minute intervals with the purpose of reducing data collection noises, which was also adopted by Ahmed and Abdel-Aty (2012).

For the segment-based approach, crashes were aggregated based on roadway segments. The Shanghai urban expressway system was split into 206 roadway segments using on-ramps and off-ramps as dividing points. For the roadway segments, there are 4 different types of ramp combinations (see Figure 1 for illustration). It was envisaged that a segment with on-ramp and on-ramp (Ramp type 1) may be different from a segment with on-ramp and off-ramp (Ramp type 2) due to the converging and diverging traffic operation characteristics. Therefore, this categorical variable was used in the segment-based analysis. Through the aggregation process, each roadway segment may result in zero crash, one crash, or multiple crashes; the operating speed and traffic volume information variables were then calculated using the following algorithm: (1) if no crash was occurred on a segment within the study period, average operating speed and traffic volume (from 6-minute intervals) for the segment were used; (2) if only one crash was reported on a segment, the corresponding pre-crash traffic status was then utilized; (3) if multiple crashes were happened on a segment, averaged pre-crash traffic conditions corresponding to these crashes were applied. In addition to these traffic variables, geometric characteristics of the roadway segments were obtained from online street-view map (Data©NavInfo) since no detailed design files were available; and the summary statistics of the segment based dataset are presented in Table 2.

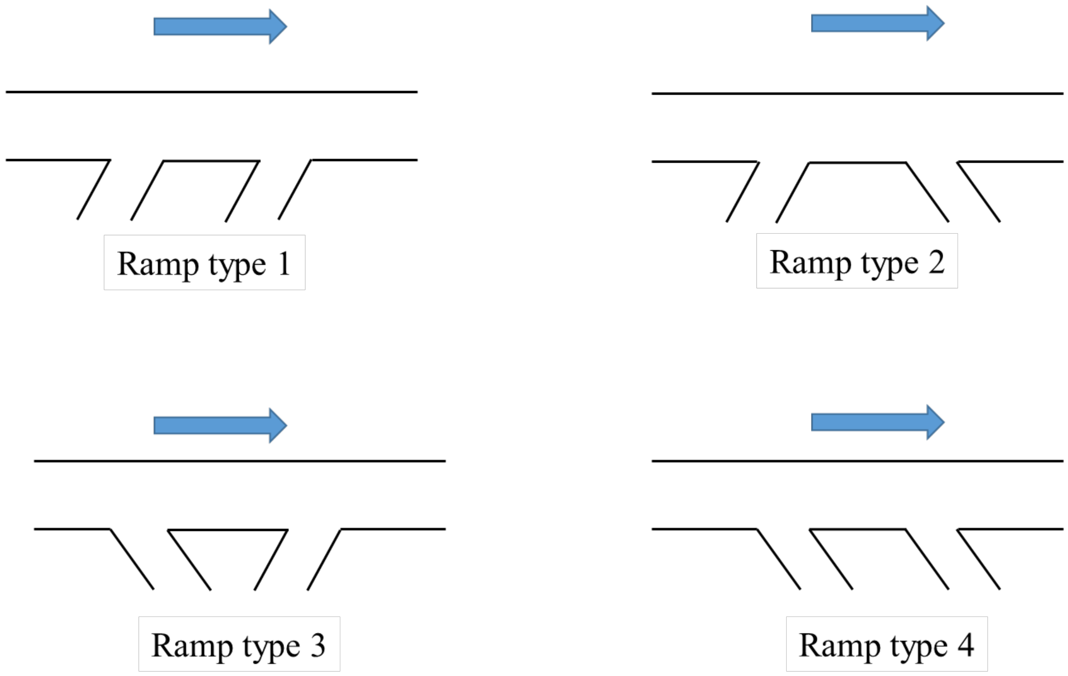

Figure 1 Ramp types

Table 2 Summary statistics of the segment-based dataset

Variable	Description	Summary Statistics
Length	Roadway segment length	Mean: 944.5 (m) Standard Deviation: 585.8 (m)
Lane Number	Number of lanes	2 lane: 59 (count) 3 lane: 59 more than 4 lanes: 88
Ramp Type	Ramp combination type: 1. On-ramp and On-ramp 2. On-ramp and Off-ramp 3. Off-ramp and On-ramp 4. Off-ramp and Off-ramp	 1: 79 (count) 2: 21 3: 71 4: 35
Average Speed	Average speed for the crashes that occurred on the same roadway segment	Mean: 44.2 (km/h) Standard Deviation: 19.2 (km/h)
Traffic Volume	Average traffic volume per lane for the crashes that occurred on the same roadway segment	Mean: 121.2 (pcupl per roadway segment) Standard Deviation: 51.2 (pcupl per roadway segment)
Auxiliary Lane Length		Mean: 177.7 (m) Standard Deviation: 178.9 (m)
Crash Frequency		Mean: 6.8 Standard Deviation: 9.2

For the scenario-based analysis, crashes were aggregated based on the combinations of similar traffic operating conditions and geometric characteristics as employed by Imprialou *et al.,* (2016). Four key variables were used as the control variable to define the potential crash scenarios: pre-crash operating speed, traffic volume, number of lanes, and ramp types. The traffic characteristics were first grouped into categories with the help of their cumulative distributions. For instance, pre-crash speed data were classified into 25 equal groups with a 4-percentile step. Similarly, traffic volume data were divided into 4 categories with a step of 25-percentile. Finally, a total of 1,200 crash occurrence scenarios were then created (i.e. 25 speed categories × 4 traffic volume categories × 3 lane numbers × 4 ramp types). For instance, one of the 1,200 observations is represented as speed is between the 20^{th} and 24^{th} percentile with the median value of 19 km/h, traffic volume is between 50^{th} and 75^{th} percentile with the median value of 154.6 veh/lane on a 3-lane expressway segment with a ramp type as on-ramp and off-ramp.

Crashes were then classified into the preset 1,200 scenarios according to their traffic conditions before crash occurrence and geometric characteristics of the crash locations. Then crashes grouped into the same scenario were aggregated to formulate the analysis dataset, and the median values of speed and traffic volume within each group were utilized to represent the traffic conditions corresponding to the calculated crash frequency. Table 3 presents the summary statistics of the scenario-based dataset.

Table 3 Summary statistics of the scenario-based dataset

Variable	Description	Summary Statistics
Lane	Number of lanes	# of lanes 2: 59 (count) # of lanes 3: 59 # of lane more than 4: 88
Ramp type	Ramp combination type: 1. On-ramp and On-ramp 2. On-ramp and Off-ramp 3. Off-ramp and On-ramp 4. Off-ramp and Off-ramp	 Type 1: 79 (count) Type 2: 21 Type 3: 71 Type 4: 35
Speed	Median speed for the preset crash occurrence scenario	Mean: 33.6 (km/h) Standard Deviation: 17.3 (km/h)
Traffic Volume	Median volume per lane for the preset crash occurrence scenario	Mean: 127.5 (pcupl per roadway segment) Standard Deviation: 46.6 (pcupl per roadway segment)
Crash Frequency		Mean: 3.8 Standard Deviation: 3.0

It is worth mentioning that since no prior assumptions used about the functional relationships between operating speed and crash frequency for the Shanghai expressway system, different functional forms should be tested. This includes: linear, logarithmic, and quadratic. In the final analysis results, only the significant variables and the best functional forms were kept.

Datasets for the crash risk analysis

In order to conduct the individual crash level analysis, a 30-minute period traffic data prior to crash occurrence were first identified. This means that five 6-minute intervals of traffic data were obtained during the data preparation process. For example, if a crash occurred on September 13, 2013 at 8:40 p.m., traffic data from 8:10 p.m. to 8:40 p.m. (i.e. a 30-minute window) were then extracted and named as time-slices 1, 2, 3, 4, and 5, with slice 1 being the 0-6 minutes interval just before the reported crash time. Meanwhile, traffic flow characteristics (e.g. average speed, total volume, standard deviation of speed and volume, coefficient of variance for volume and speed) were calculated from 6-minute intervals. In addition, instead of only utilizing traffic related variables from the crash current segments (C), data from both upstream (U) and downstream (D) segments were incorporated. The spatial relationship between the roadway segments is shown in Figure 2. As a result, a total of 90 variables (i.e. 6 traffic flow variables × 3 detector stations × 5 time slices) were generated and used in the latter model estimation procedure.

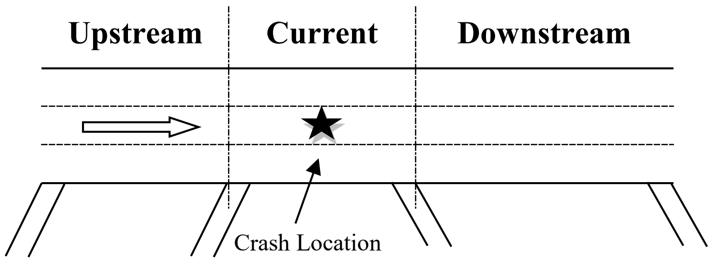

Figure 2 Arrangement of roadway sections

Since the primary idea of this part analysis is to compare normal traffic conditions with those of pre-crash conditions, traffic data from non-crash cases were also extracted. For each crash, four non-crash cases were extracted by following the matched-case control data structure as employed in existing students (e.g. Ahmed and Abdel-Aty, 2012), which was also tested in the previous sensitivity analysis. Non-crash traffic conditions were collected when no crash was observed within a 2-hour window, given the same time of day, day of week, and roadway section. For example, if a crash occurs on a segment with NN0312 (stake number) on September 13, 2013 at 8:40 p.m., traffic data for the same roadway section and time on August 31 and September 6 (i.e. two observations before the crash event) and September 20 and

September 27 (i.e. two observations after the crash event) were collected as non-crash cases only if there is no crash at the time period from7:40 p.m. to 9:40 p.m. on these dates. Through matching, the final dataset has 1,387 matched strata with 1,387 crashes and 3,811 non-crashes (in a few cases, the non-exact 1:4 crash and non-crash ratio is due to the traffic data availability issue).

Methodology

In order to quantify the impacts of aggregation levels on the relationships between operating speed and traffic safety, two types of models have been employed in this study: random-effects negative binomial models were used for crash frequency analyses while random-effects logistic regression models were adopted for crash risk analyses. These models were estimated by employing the Bayesian inference technique. This section introduces the model structure and the relevant inference settings.

Random-effects negative binomial model

Crash frequency data aggregated by roadway segments or by operating scenarios were assumed to follow the negative binomial distribution suitable for accounting for the over-dispersion inherent in count data (e.g. Lord and Mannering, 2010). As suggested by the previous studies (e.g. Yu *et al.,* 2013), a random-effect term was added to account for the unobserved heterogeneity. The random effects negative binomial model can be setup as follows (Ntzoufras, 2009):

$$Y_i \sim Negative\ Binomial(p_i, r)$$

$$p_i = \mathrm{r}/(\mathrm{r} + \lambda_i)$$

$$ln\lambda_i = \ offset_i + \textstyle\sum_{j=1}^{k} X_{ij}\beta_j + \ u_i \qquad (1)$$

$$and \ \ i = 1, 2, \ldots, n; \qquad j = 1,2,\ldots,k$$

where Y_i is the crash count for a roadway segment *i* or the crash count for a scenario i, r is the dispersion parameter, p_i and λ_i are the negative binomial distribution parameters, X_{ij} represent the set of explanatory variables and β_j is the corresponding regression parameters to be estimated, *k* is the number of explanatory variables and *n* is the total number of observations. Segment length denoted as $\ln(SegmentLenth_i)$ can be used as the *offset* variable in the segment-based analysis while average vehicle-hours spent per scenario denoted as $\ln(AverageVehicleHours)$ can be used as the *offset* variable in the scenario-based model as suggested by Imprialou *et al.* (2016). u_i is the segment/scenario specific random effect which set to follow the normal distribution with $u_i \sim N\ (0, {}^{1}\!/_{\tau})$, where τ was specified a gamma prior as $\tau \sim$ Gamma (0.001, 0.001).

Random-effects logistic regression model

In the crash risk analysis, the target variable is a binary category with 1 being crash cases and 0 represents non-crash cases. Suppose observation Y_i has the outcomes of crash and non-crash with corresponding probabilities being p_i and $1-\mathrm{p}_i$ respectively. The random effects logistic regression model can be set up as follows:

$$Y_i \sim Binomial\ (p_i)$$

$$logit\ (p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \sum_{j=1}^{m} Z_{ij}\beta_j + \varepsilon_t \qquad (2)$$

$$and \quad i = 1, 2, \ldots, N; \qquad j = 1,2,..,m$$

where β_0 is the intercept and Z_{ij} is the set of explanatory variables, β_j is the corresponding regression coefficients to be estimated, *m* is the number of explanatory variables, *N* is the number of observations ε_t is the random effects term:

$$\varepsilon_t \sim N\left(0, {}^{1}\!/_{\alpha}\right)$$

$$\alpha \sim \text{ Gamma (0.001, 0.001)}$$

where t stands for the crash unit index (crash observation and their matched non-crash cases). The random effects term can take into account any potential unobserved heterogeneity arising from omitted geometric characteristics not considered in the set of explanatory variables such as auxiliary lane length.

Bayesian Inference

Full Bayesian inference was employed in this study with non-informative priors. For each model, three chains of 20,000 iterations were set up in WinBUGS (Lunn *et al.,* 2000) with the thin set equal to 3; the first 5,000 stored iterations were used as burn-in samples and the rest was used to estimate the poster distribution. Convergences of the developed models were checked by monitoring the MCMC (Markov chain Monte Carlo) trace plots for the parameters and the model convergence issue was further checked through calculating BGR statistics (Gelman and Rubin, 1992) and conducting the Geweke diagnostic through R package - *boa* (Smith, 2007).

Modeling Results

Segment-based Analysis

Table 4 shows the posterior estimations of the Bayesian random-effects negative binomial model for the segment-based dataset. Five explanatory variables became statistically significant based on their 95% posterior credible levels. For the operating speed, Av_Spd is significant with a negative coefficient, which indicates that as the operating speed increase, crash frequency would be reduced. Similar results have also been concluded in the previous study (Yu *et al.,* 2013), which can be understood as that crashes are more prone to happen at congested segments.

Besides, traffic volume - Ln(Vol per lane) holds a positive estimate; indicating that the larger traffic exposure, the larger crash frequency. For the variable representing lane numbers, Lane_3 was treated as the reference group; Lane_2 shows a positive association with the crash frequency whereas Lane_4 has a negative coefficient, which indicates that as segments with high number of lanes are associated with lower crash counts. Aux_length was found to have a significant impact on crash frequency. More specifically, longer auxiliary length within the roadway segment would substantially reduce crash frequency. For the ramp types, Ramp_1 was identified as no substantial difference when compared to Ramp_2, while Ramp_3 and Ramp_4 were proved to provide lower crash hazardous.

Table 4 Coefficient estimates for segment-based analysis

Variable	Mean	S.D.	2.5%	97.5%
Intercept	-3.7	1.36	-6.28	-1.12
Lane_2	0.68	0.24	0.21	1.17
Lane_4	-0.46	0.20	-0.86	-0.06
Lane_3 (reference)	0	-	-	-
Aux_Length	-0.002	0.0007	-0.003	-0.0007
Ramp_1	0.067	0.26	-0.45	0.59
Ramp_2 (reference)	0	-	-	-
Ramp_3	-1.08	0.31	-1.71	-0.46
Ramp_4	-0.48	0.26	-1.02	-0.02
Av_Spd	-0.03	0.006	-0.04	-0.02
Ln(Vol per lane)	1.32	0.26	0.82	1.83
Offset variable	1	ln(Total link length)		
Tau	2.54	1.62	1.2	7.11
# of observations		206		
DIC		913.6		

Scenario-based Analysis

Table 5 shows the estimation results for the scenario-based analysis. Both operating speed and its quadratic parameter became significant. The speed parameter holds a negative coefficient and speed quadratic parameter shows a positive impact; the relationship between operating speed and crash occurrence can therefore be regarded as a U-shaped curve. This means that crash frequency decreases as operating speed increases before a critical speed is reached. After the critical speed, crash frequency increases with the operating speed. From the estimated coefficients (see Table 4), this critical speed is predicted to be 25 km/h for the sample data from the Shanghai Urban Expressway system. This reveals that the impact of operating speed on crashes reaches to a minimum level when the mean operating speed is about 25 km/h.

In addition, for the geometric characteristic parameters, lane numbers and ramp types were also statistically significant. Consistent results have been concluded for number of lanes with the segment-based approach, where segments with more lanes are related to reduced crash occurrences. While for ramp types, Ramp_2 was identified to be the most hazardous one, the combination of on-ramp and off-ramp would pose large needs of traffic weavings; which is inconsistent with the segment-based analysis. Furthermore, the estimation result for traffic volume (Vol per lane) is consistent with the segment-based analysis, whereas the increase of volume would increase the crash occurrence exposure.

Table 5 Coefficient estimates for scenario-based analysis

Variable	Mean	S.D.	2.5%	97.5%
Intercept	1.13	0.17	0.80	1.46
Lane_2	0.61	0.16	0.28	0.93
Lane_4	-0.14	0.09	-0.22	-0.03
Lane_3 (reference)	0	-	-	-
Ramp_1	-0.36	0.11	-0.60	-0.16
Ramp_2 (reference)	0	-	-	-
Ramp_3	-0.37	0.09	-0.55	-0.18
Ramp_4	-0.58	0.18	-0.93	-0.22
Speed	-0.025	0.009	-0.045	-0.008
Speed*Speed	0.0004	0.00008	0.0003	0.0006
Vol per lane	0.0038	0.0008	0.0023	0.0053
Offset variable	1	ln(Average vehicle-hours spent per scenario)		
Tau	2.56	0.41	1.84	3.49
# of observations		974		
DIC		4252.68		

Crash risk analysis model

In this section, disaggregate crash risk analyses were conducted to identify the relationships between operating speed and individual crash occurrence probability. Since it was claimed in the previous studies that crash risk analysis varies by different operating conditions (Abdel-Aty *et al.*, 2005), four speed categories were classified in this study according to the operation conditions at Shanghai urban expressway system: low speed (less than 20 km/h), medium speed (between 20 km/h and 40 km/h), high speed (40 km/h to 60 km/h), and free-flow speed (above 60 km/h).

Table 6 shows the modeling results for the crash risk analysis that considers different operating speed conditions. For each model, three significant variables were achieved. For low speed conditions, average speed at crash segment at time slice 1 (ASC1) poses a negative relationship with crash risk, which refers to congested flow would have higher crash likelihood. Traffic volume at crash segment at time slice 2 (TFC2) has a positive coefficient, which indicates that the increase of traffic volume would lead to larger crash hazardous. In addition, speed standard deviation of downstream segment time slice 1 (SSD1) has a positive coefficient, which can be understood as larger speed variation at downstream would enhance the crash risk.

While for moderate speed conditions, ASC1 again holds a negative coefficient and the speed standard deviation of crash segment at time slice 1 (SSC1) has a positive coefficient, which can be illustrated as smoother and more homogenous traffic would lead to reduced crash probability. Besides, upstream traffic volume standard deviation at time slice 1 (SFU1) has a positive coefficient, which means that the variation of upstream flow would enhance the crash occurrence likelihood.

In addition, an interesting finding is that instead of ASC1, the average speed at downstream segment time slice 1 (ASD1) was found to provide more substantial impacts on crash occurrence likelihood, while average operating speed at crash locations does not have substantial correlations with crash occurrence. Furthermore, standard deviation of traffic volume at crash segment time slice 1 (SFC1) and SSC1 both have positive coefficients, which means turbulence traffic would lead to larger crash hazardous.

Furthermore, for the free-flow conditions, SSC1 and SFC1 hold consistently estimated coefficients. However, the ASC1 has a positive sign, which indicates that as the increase of operating speed, the crash risk would be also increased. This is a contradictory of the results identified in the low speed condition and moderate speed condition.

Table 6 Coefficient estimates for crash risk analysis by speed conditions

Variable	Definition	Low Speed Condition		Moderate Speed Condition		High speed condition		Free-flow speed condition	
		Mean (S.D.)	95% C.I.	Mean (S.D.)	95% C.I.	Mean (S.D.)	95% C.I.	Mean (S.D.)	95% C.I.
Intercept		1.52 (0.27)	(1.0 , 2.01)	0.92 (0.18)	(0.56, 1.29)	-1.42 (0.22)	（-1.87, -0.99）	-4.43 (0.94)	(-6.32, -2.65)
TFC2	Traffic volume at the crash segment at time slice 2	0.002 (0.0007)	(0.0005, 0.0033)	-	-	-	-	-	-
ASC1	Average speed at crash segment at time slice 1	-0.16 (0.017)	(-0.20, -0.13)	-0.08 (0.005)	(-0.09, -0.07)	-	-	0.038 (0.012)	(0.014, 0.064)
ASD1	Average speed at downstream segment at time slice 1	-	-	-	-	-0.02 (0.003)	(-0.03, -0.01)	-	-
SSD1	Speed standard deviation of downstream segment at time slice 1	0.07 (0.02)	(0.02, 0.11)	-	-	-	-	-	-
SSC1	Speed standard deviation of crash segment at time slice 1	-	-	0.18 (0.02)	(0.15, 0.22)	0.25 (0.024)	（0.20, 0.29）	0.15 (0.05)	(0.005, 0.26)
SFU1	Upstream traffic volume standard deviation at time slice 1	-	-	0.027 (0.0064)	(0.015, 0.04)	-	-	-	-
SFC1	Traffic volume standard deviation at crash segment at time slice 1	-	-	-	-	0.038 (0.007)	（0.024, 0.052）	0.03 (0.01)	(0.006, 0.05)
Tau		87.11 (220.8)	(2.84, 708.9)	242.3 (416.1)	(10.58, 1427)	374.6 (496.8)	(16.64, 1748)	368.5 (521.3)	(16.32, 1865)
# of observations		1496		1997		1245		460	
AUC		0.84		0.81		0.78		0.63	

Therefore, the relationships between operating speed and traffic safety at crash individual aggregation level is concluded as: operating speed has negative impacts on crash occurrence risk under low and moderate speed conditions, at high speed conditions the impacts of speed on crash occurrence is vague, while at free-flow conditions speed holds positive impacts. The modeling results indicate that the relationship between operating speed and traffic safety do not hold a linear line, it varies at different operation conditions.

Discussions and Conclusions

Emerging active safety management systems, such as Variable Speed Limits System or in-vehicle speed advisory system under Connected Vehicle (CV) scenario, require deep understandings of the relationships between operating speed and crash occurrence. As alluded earlier that most previous studies however used spatio-temporal average speed instead of speed information prior to crash occurrence in their analyses due to the data aggregation issue. As a result, there were no consistent findings being obtained as the over-aggregated data might fail to reveal the true association between the two.

In this study, the impacts of aggregation approaches on the relationship analyses were investigated based on the advanced traffic sensing data of Shanghai urban expressway systems. Crash frequency analyses with segment-based approach and scenario-based approach were firstly being conducted, and then crash risk analyses were developed at individual crash level. The segment-based crash frequency analysis revealed a negative relationship between the two. On the other hand, as shown in Figure 3, the results from the scenario-based crash frequency analysis, average crashes per kilometer are relatively high at both low speed traffic conditions and high speed conditions; the relationships between operating speed and crash occurrence were therefore concluded as a U-shape curve.

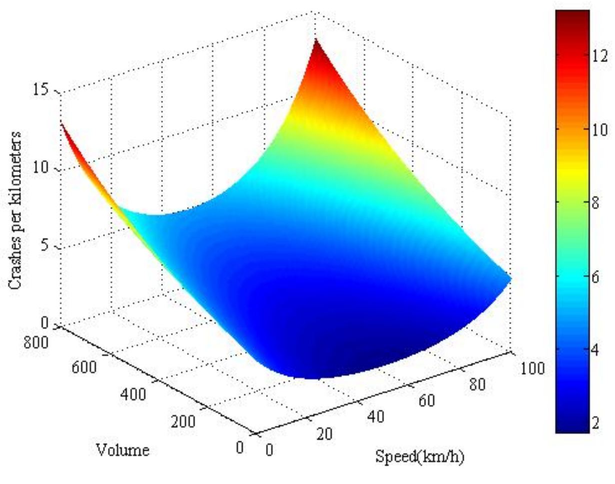

Figure 3 Relationships between speed, volume and crash rates from scenario-based crash frequency analysis

Given the inconsistent results obtained from the crash frequency analyses, disaggregate crash risk analyses were further conducted. Figure 4 shows the box plot of the estimated coefficients for the operating speed parameter (ASC1) and Table 7 shows the estimated marginal effects of parameter ASC1, where the coefficient of ASCI indicates the crash occurrence likelihood and operating speed. It can be concluded that during the congestion period (i.e. low and moderate speed conditions), the increase of operating speed would reduce a crash likelihood; for medium operating speed the changes of operating speed do not have substantial effects on crash occurrence probability; while for free-flow, the increase of operating speed would further enhance crash hazardous.

The crash risk analyses have been an important topic in the traffic safety analysis discipline in which different study area and research objectives have been investigated. The earlier studies were mostly conducted based on total crashes and have identified that the coefficient of variation of speed was the crash occurrence contributing factor (Lee *et al.*, 2003, Abdel-Aty *et al.,* 2004), which could be understood as lower operating speed and large speed variation would lead to more crashes. Recently, a few studies investigated the effect of different operating conditions on safety. For instance, Pande and Abdel-Aty (2006) investigated the rear-end crash occurrence influencing factors, and the crashes were separated into low speed and high speed conditions. Their findings are consistent with this current study where speed is positively associated with traffic crashes for high operating speed conditions; while in the low speed conditions, larger coefficient of variation of speed would lead to increased crash risk. However, instead of split crashes by operating conditions, majority crash risk analyses divide crashes by crash types (Christoforou *et al.*, 2011), weather conditions (Xu *et al.*, 2013), and crash injury severity (Yu and Abdel-Aty 2014). But inconsistent findings reappear which may be due to the heterogeneity effect resulting from different operating conditions. For instance, Oh and Kim (2010) identified a positive correlation between speed and crash for rear-end crashes while Christoforou *et al.,* (2011) found a negative association. Therefore, based on the current findings, it is advisable that further crash risk analyses shall consider the heterogeneity effects of operating speed on traffic safety.

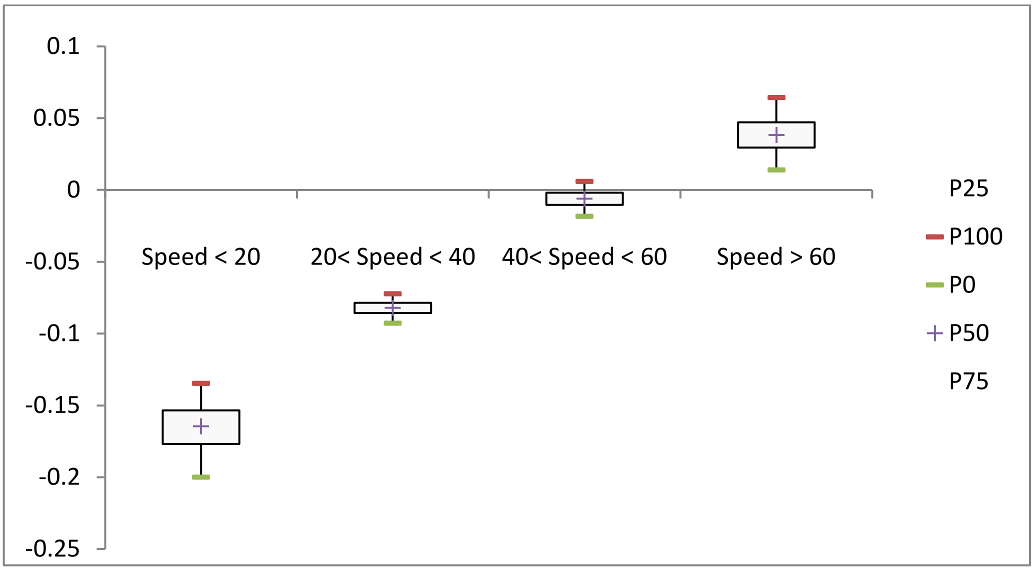

Figure 4 Box-plot for estimated coefficients of ASC1 in the crash risk analyses

Table 7 Coefficient marginal effects for ASC1

Speed Conditions	Marginal Effects for ASC1
Low Speed Condition	-0.01649
Moderate Speed Condition	-0.00345
High Speed Condition	-0.000521*
Free-flow Speed Condition	0.00399

* Insignificant marginal effect at 95% level

Through comparisons, results of the crash risk analyses are consistent with the scenario-based approach crash frequency analysis. A U-shape curve relationship may be a better illustration between the operating speed and traffic safety. The linear relationship exits in the segment-based approach may be attributed to the data aggregation process; during the aggregation, crashes with high speed would be averaged by medium or low speed crash-prone speed, which leads to a monotonous relationship between speed and safety. Therefore, the scenario-based aggregation approach and crash risk analysis by speed categories are more plausible and preferred for future studies with similar objectives.

In addition, through the crash risk analyses, typical crash occurrence scenarios can be speculated with the significant contributing factors. For low speed conditions, crashes are mostly likely to happen within congested segments, where traffic flow dissipates at its downstream segment. At moderate speed conditions, crashes occurred at turbulence flow segment while its upstream has a large traffic flow. While at high speed conditions, crashes are more likely to occur at the end of shockwave propagation segment where its downstream segments were congested. In addition, for crashes occurred under free-flow conditions, the crash causations are mostly related to the unexpected traffic turbulence. With these profound understandings of crash mechanisms, targeted ATMS could be designed to improve traffic safety for the urban expressway system.

Moreover, findings from this study should be carefully interpreted as the detailed design data were not obtained for the studied area, and some roadway geometry variables (e.g. degree of curvature, gradient) were not included. Additionally, it would also be interesting to analyze the impacts of statistical modeling approach on the relationships. For instance, applying models such as random-parameter negative binomial model, finite-mixture models rather than a random-effect negative binomial model employed in this study. Last but not the least, another important factor that needs an attention is the impact of speed variation (Pei *et al.*, 2012).

Acknowledgement

This study was jointly sponsored by the Chinese National Natural Science Foundation (NSFC 71771174 and 71531011) and the 111 Project (B17032).

Reference

Aarts, L., Schagen, I., 2006. Driving speed and the risk of road crashes: A review. Accident Analysis and Prevention, Vol. 38, pp. 215-224.

Abdel-Aty, M., and Radwan, E., 2000. Modeling traffic accident occurrence and involvement. Accident Analysis & Prevention, Vol. 32, No. 5, pp. 633-642.

Abdel-Aty, M., Abdalla, F., 2004. Linking roadway geometrics and real-time traffic characteristics to model daytime freeway crashes: Generalized estimating equations for correlated data. Transportation Research Record: Journal of the Transportation Research Board, 1897(1), 106-115.

Abdel-Aty, M., Uddin. N., Pande, A., 2005. Split models for predicting multivehicle crashes during high-speed and low-speed operating conditions on freeways. Transportation research record: journal of the transportation research board, No. 1908, pp. 51-58.

Ahmed, M, and Abdel-Aty, M., 2012. The Viability of Using Automatic Vehicle Identification Data for Real-Time Crash Prediction. IEEE Transactions on Intelligent Transportation Systems, Vol. 13, No. 2, pp. 459-468.

American Association of State Highway and Transportation Officials (AASHTO), 2010. Highway Safety Manual.

Christoforou, Z., Cohen, S., Karlaftis, M., 2011. Identifying crash type propensity using real-time traffic data on freeways. Journal of Safety Research, 42(1), 43-50.

Elvik, R., 2013. A re-parameterisation of the Power Model of the relationship between the speed of traffic and the number of accidents and accident victims. Accident Analysis & Prevention, Vol. 50, pp. 854-860.

Elvik, R., Vaa, T., 2004. The Handbook of Road Safety Measures. Elsevier, Oxford.

Fitzpatrick, C., McKinnon. I., Tainter. F., Knodler, M., 2016. The application of continuous speed data for setting rational speed limits and improving roadway safety. Safety science, Vol. 85, pp. 171-178.

Gelman, A., Rubin, D., 1992. Inference from Iterative Simulation Using Multiple Sequences. Statistical Science, Vol. 7, pp. 457-511.

Imprialou, M., Quddus, M., Pitfield, D., Lord, D., 2016. Re-visiting crash–speed relationships: A new perspective in crash modelling. Accident Analysis & Prevention, Vol. 86, pp. 173-185.

Kloeden, C, McLean, A., Glonek, G., 2002. Reanalysis of travelling speed and the rate of crash involvement in Adelaide South Australia. Report No. CR 207. Australian Transport Safety Bureau ATSB, Civic Square, ACT

Lee, C., Hellinga, B., Saccomanno, F., 2003. Real-time crash prediction model for application to crash prevention in freeway traffic. Transportation Research Record: Journal of the Transportation Research Board, 1840 (1), 67-77.

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

Lunn, D., Thomas, A., Best, N., Spiegelhalter, D., 2000. WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing, Vol.10, No.4, pp.325-337.

Mirshahi, M., Obenberger, J., Fuhs, C., Howard, C., Krammes, R., Kuhn, B., Mayhew, R., Moore, M., Sahebjam, K., Stone, C., 2007. Active traffic management: the next step in congestion management.

Moreno, A., García, A., 2013. Use of speed profile as surrogate measure: Effect of traffic calming devices on crosstown road safety performance. Accident Analysis & Prevention, Vol. 61, pp. 23-32.

Ntzoufras, I., 2009. Bayesian modeling using WinBUGS. WILEY.

Pande, A., Abdel-Aty, M., 2006. Comprehensive analysis of the relationship between real-time traffic surveillance data and rear-end crashes on freeways. Transportation Research Record: Journal of the Transportation Research Board, 1953 (1), 31-40.

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

Quimby, A., Maycock, G., Palmer, C., Buttress, S., 1999. The factors that influence a driver's choice of speed: a questionnaire study. TRL Report No. 325. Transport Research Laboratory TRL, Crowthorne, Berkshire.

Roshandel S., Zheng Z., Washington S., 2015. Impact of real-time traffic characteristics on freeway crash occurrence: Systematic review and meta-analysis. Accident Analysis & Prevention, Vol.79, pp. 198-211.

Smith, B., 2007. Boa: An r package for MCMC output convergence assessment and posterior inference. Journal of Statistical Software, Vol.21, pp.1-37.

Taylor, M., Lynam, D., Baruya, A., 2000. The effects of drivers' speed on the frequency of road accidents. TRL REPORT 421.

Xu, C., Tarko, A., Wang, W., Liu, P., 2013. Predicting crash likelihood and severity on freeways with real-time loop detector data. Accident Analysis & Prevention, 57, 30-39.

Yu, R., Abdel-Aty, M., Ahmed, M., 2013. Bayesian random effect models incorporating real-time weather and traffic data to investigate mountainous freeway hazardous factors. Accident Analysis & Prevention, Vol. 50, pp.371-376.

Yu, R., Abdel-Aty, M., 2014. Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data. Safety Science, 63, 50-56.