Accident Analysis and Prevention 120 (2018) 304–310

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

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

Rongjie Yu[a,b,*], Mohammed Quddus[c], Xuesong Wang[a,b], Kui Yang[a,b]

[a] School of Transportation Engineering, Tongji University, 4800 Cao'an Road, 201804, Shanghai, China
[b] Key Laboratory of Road and Traffic Engineering of Ministry of Education, 4800 Cao'an Road, 201804, Shanghai, China
[c] School of Architecture, Building and Civil Engineering, Loughborough University, Loughborough, LE11 3TU, United Kingdom

ARTICLE INFO

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

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 modeling 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.

1. 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 García, 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

* Corresponding author at: School of Transportation Engineering, Tongji University, 4800 Cao'an Road, 201804, Shanghai, China.
E-mail address: yurongjie@tongji.edu.cn (R. Yu).

https://doi.org/10.1016/j.aap.2018.06.007
Received 16 January 2018; Received in revised form 11 June 2018; Accepted 12 June 2018

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 and 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

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.

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 (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.

2. 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 m (compared to an average of around 800 m found in most studies (e.g. Xu et al., 2013; Abdel-Aty et al., 2005), 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.

2.1. 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 Fig. 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