Accident Analysis and Prevention 104 (2017) 58–64

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Safety analytics for integrating crash frequency and real-time risk modeling for expressways

Ling Wang*, Mohamed Abdel-Aty, Jaeyoung Lee

Department of Civil, Environmental and Construction Engineering, University of Central Florida, Orlando, FL, United States

ARTICLE INFO

Keywords:
Crash frequency analysis
Real-time safety analysis
Integrated model
Expected crash frequency
Average daily standard deviation of speed

ABSTRACT

To find crash contributing factors, there have been numerous crash frequency and real-time safety studies, but such studies have been conducted independently. Until this point, no researcher has simultaneously analyzed crash frequency and real-time crash risk to test whether integrating them could better explain crash occurrence. Therefore, this study aims at integrating crash frequency and real-time safety analyses using expressway data. A Bayesian integrated model and a non-integrated model were built: the integrated model linked the crash frequency and the real-time models by adding the logarithm of the estimated expected crash frequency in the real-time model; the non-integrated model independently estimated the crash frequency and the real-time crash risk. The results showed that the integrated model outperformed the non-integrated model, as it provided much better model results for both the crash frequency and the real-time models. This result indicated that the added component, the logarithm of the expected crash frequency, successfully linked and provided useful information to the two models. This study uncovered few variables that are not typically included in the crash frequency analysis. For example, the average daily standard deviation of speed, which was aggregated based on speed at 1-min intervals, had a positive effect on crash frequency. In conclusion, this study suggested a methodology to improve the crash frequency and real-time models by integrating them, and it might inspire future researchers to understand crash mechanisms better.

1. Introduction

Traffic safety research includes wide-ranging areas, and the most prominent one is assessing the safety of roadway facilities, e.g., intersections, segments, and corridors. There are two common types of traffic safety analyses: (1) the crash frequency analysis and (2) the real-time crash risk analysis. The crash frequency analysis estimates crash frequencies with crash contributing factors, such as geometric design features (e.g., horizontal and vertical alignments) and traffic parameters (e.g., average daily traffic and truck percentage). The objective is to explore the safety performance of a roadway facility over a long time period, for example, one year, and the crash frequency analysis mainly utilize aggregated traffic data, such as average daily traffic (ADT). Based on crash frequency analyses, large numbers of research studies have been conducted to find the potential for safety improvements and to measure the safety impact of countermeasures (for example, Elyasi et al., 2016; Park and Abdel-Aty, 2015).

In the recent decade, many traffic safety researchers began to concentrate on the real-time safety analysis using dynamic traffic data. The objective of the real-time safety analysis is to investigate the crash likelihood of a roadway facility over a very short time period, for example, 5 min. Its assumption is that the occurrence of a crash is mainly because of short-term turbulence right before the crash, and the turbulence could be identified by crash precursors, such as real-time traffic and weather condition (Ahmed et al., 2014). The implementation of the real-time safety analysis is mainly in the field of active traffic management (ATM), such as, variable speed limit and ramp metering (Abdel-Aty et al., 2009; Lee et al., 2004).

Though the applications of the crash frequency analysis and the real-time safety analysis are various, both of them attempt to finding crash mechanisms and building relationship between safety and its contributing factors. Hence, it is supposed that they uncover similar crash mechanisms. However, previous studies carried out crash analyses based on crash frequency or real-time safety analyses separately, and no study incorporated these two safety analyses. Since these two types of safety studies have the same objectives – finding crash contributing factors – integrating crash frequency and real-time safety models might enhance the model performance of each model and might provide better explanation of the crash mechanisms.

This paper begins with a literature review of the crash frequency

* Corresponding author.
E-mail address: lingwang@knights.ucf.edu (L. Wang).

http://dx.doi.org/10.1016/j.aap.2017.04.009
Received 3 January 2017; Received in revised form 12 March 2017; Accepted 10 April 2017

and the real-time safety studies (Section 2). It moves to the methodologies about the integrated and non-integrated models (Section 3). Then, data preparation along with descriptive analyses is conducted (Section 4). The model results and interpretations are illustrated (Section 5). Finally, the paper concludes with a summary of the results and insights for future work (Section 6).

2. Literature review

In existing traffic safety studies, the research studies that evaluated the crash potentials of roadway facilities can be classified into two main categories: (1) the crash frequency analysis using highly aggregated traffic data, such as ADT; and (2) the real-time safety analysis using real-time traffic data, for example, traffic variables at 5-min intervals.

The crash frequency analysis focuses on identifying factors that contribute to crash counts. The subject examined in crash frequency models is crashes aggregated at a segment for a given long period of time (such as one year). These models are developed by using aggregated traffic data and roadway geometry. Researchers have conducted the crash frequency analysis on a variety of roadway facilities, such as intersections (Jonsson et al., 2007), expressways (Wang et al., 2015a), and interchanges (Bonneson et al., 2012). The statistical methods used in crash frequency analyses are Poisson, negative binominal, Poisson-lognormal, and other various count regression models (Lord and Mannering, 2010). The outcomes of these studies are used to propose effective countermeasures (such as improving roadway design) and to identify the potential for safety improvement (Elyasi et al., 2016). Meanwhile, these statistical models have been adopted to develop the safety performance functions (SPFs) and then are integrated to obtain the crash modification factors (CMFs) (Park and Abdel-Aty, 2016). However, crash frequency models are not able to dynamically identify short-term hazardous conditions. Thus, the aggregated crash models are not capable of providing guidance for the implementation of ATM.

On the other hand, the real-time safety study analyzes real-time crash likelihoods with short-term conditions. These studies distinguish some traffic and weather parameters as “crash precursors”, because these parameters were significantly related to crash occurrence. The data used in such analyses are dynamic traffic and weather data, which are usually specified to 5-min intervals (Fang et al., 2016; Kwak and Kho, 2016). Dissimilar to the crash frequency analysis whose study subject is wide-ranging, the real-time safety study are only carried out for traffic detector equipped and uninterrupted-flow freeway or expressway facilities. The main statistical method used in the real-time safety analysis is binary logistic regressions (Abdel-Aty et al., 2004; Wang et al., 2015b). Additionally, in order to exclude the impact of segment geometry, matched-case-control logistic regression models were widely adopted (Abdel-Aty et al., 2004; Kwak and Kho, 2016). Once the crash precursors are quantitatively identified by the real-time safety studies, researchers began to implement real-time crash prediction in ATM. When the hazardous conditions were detected using crash precursors, variable speed limit (Allaby et al., 2007) or ramp metering was applied to improve traffic safety (Lee et al., 2006).

There are significant differences between the crash frequency analysis and the real-time safety analysis. The implementations are dissimilar. The implementation of real-time safety studies is in a proactive way by allowing ATM to be active early in an attempt to prevent crashes (Abdel-Aty and Pande, 2007). On the other hand, current crash frequency analyses are mainly used in reactive process, such as “hotspot” identification after the observed crash count exceed a predefined threshold (Chung et al., 2011). Meanwhile, the target subjects are different. Real-time safety studies focus on the status of a 5-min event (outcome is crash and non-crash status), but crash frequency analyses take a long period of a segment as the whole and concentrate on the total crash count. However, these two safety analyses also share common aspects. First, both of them are focusing on the safety of roadway facilities. Second, both of them have explored the crash contributing factors and found similar patterns, for example, both studies have found that a wider shoulder provides safer traffic (Park and Abdel-Aty, 2015; Shi et al., 2016). Third, the traffic data used in these two safety analyses are related: a segment with a higher ADT also has a high potential to experience a higher 5-min traffic count than a segment with a lower ADT.

Because of the common aspects of these two safety analyses, crash frequency analyses and real-time safety analyses might support each other. Hence, integrating them might help researchers in better finding crash mechanisms and could provide better estimation results. However, no study has integrated these two safety studies.

3. Methodology

This study includes two types of safety analyses: crash frequency and real-time. For the crash frequency analysis, a Poisson-lognormal model was used. Poisson, Poisson-gamma, and Poisson-lognormal regression models are commonly used in crash frequency studies (Lord and Mannering, 2010). The difference between these three models is the error term, which is added in the Poisson-gamma and Poisson-lognormal regression models to account for the over-dispersion of the Poisson model. To be more specific, the exponent of the error term of the Poisson-gamma and Poisson-lognormal models follow gamma and lognormal distributions, respectively. In this study, the Poisson-lognormal was used, as the Poisson-lognormal model is potentially more flexible than the Poisson-gamma regression model (Lord and Mannering, 2010).

As for the real-time safety analysis, a logistic regression model was used. The logistic regression model has been commonly used in real-time safety studies, because it is capable of handling the categorical target variable (Washington et al., 2010) and also because it can deal with case–control design by providing valid estimations of the coefficients of the independent variables (Hosmer et al., 2013). In the real-time safety analysis, the number of the case events (crash) is much less than that of the control events (non-crash). For example, in one year, a segment has ten crash events but has about one million non-crash events. Handling such huge non-crash data might not be practical, hence, in the real-time safety analysis, the case-control design has been broadly applied (Abdel-Aty and Pemmanaboina, 2006; Kwak and Kho, 2016). This study chose 1:5 as the case-to-control ratio: for each crash event, a sample of five non-crash events were randomly selected. The 1:5 ratio has been used in previous studies, and the selection of a ratio did not have a significant impact on model results (Abdel-Aty et al., 2004; Zheng et al., 2010).

Both the Poisson-lognormal regression and the logistic regression models were estimated under the Bayesian inference framework. One of the advantages of Bayesian models over traditional statistical models is the value of the coefficients of independent variables (Reis and Stedinger, 2005). Traditional models treat the coefficients as fixed values, but the Bayesian inference sets the coefficients to follow specific distribution(s). Hence, for the model results, the Bayesian inference can provide full distributions of the coefficients; on the contrary, the traditional model only gives point estimations of the coefficients and adopts an asymptotic normality assumption to describe the uncertainties of these coefficients.

Two models were estimated, one is non-integrated model and the other one is integrated model. The non-integrated model independently estimated the crash frequency model and the real-time model. For the crash frequency model, the crash frequency on segment j (Y_j) follows a Poisson distribution, whose mean was the expected crash frequency (λ_j). And λ_j was determined by the crash frequency level parameters, for example, ADT and geometric variables.

$$Y_j \sim Poisson(\lambda_j) \tag{1}$$