Accident Analysis and Prevention 83 (2015) 57–66

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Safety modeling of urban arterials in Shanghai, China

Xuesong Wang[a,b,*], Tianxiang Fan[a], Ming Chen[a], Bing Deng[c], Bing Wu[a], Paul Tremont[a]

[a] School of Transportation Engineering, Tongji University, Shanghai 201804, China
[b] The Key Laboratory of Road and Traffic Engineering, Ministry of Education, China
[c] General Motors Safety Electronics & Innovation, Warren, MI 48090, USA

ARTICLE INFO

Article history:
Received 26 April 2014
Received in revised form 14 April 2015
Accepted 3 July 2015
Available online xxx

Keywords:
Urban arterial
Safety analysis
Floating Car Data
Poisson-lognormal models
Bayesian inference
Hierarchical structure

ABSTRACT

Traffic safety on urban arterials is influenced by several key variables including geometric design features, land use, traffic volume, and travel speeds. This paper is an exploratory study of the relationship of these variables to safety. It uses a comparatively new method of measuring speeds by extracting GPS data from taxis operating on Shanghai's urban network. This GPS derived speed data, hereafter called Floating Car Data (FCD) was used to calculate average speeds during peak and off-peak hours, and was acquired from samples of 15,000+ taxis traveling on 176 segments over 18 major arterials in central Shanghai. Geometric design features of these arterials and surrounding land use characteristics were obtained by field investigation, and crash data was obtained from police reports. Bayesian inference using four different models, Poisson-lognormal (PLN), PLN with Maximum Likelihood priors (PLN-ML), hierarchical PLN (HPLN), and HPLN with Maximum Likelihood priors (HPLN-ML), was used to estimate crash frequencies. Results showed the HPLN-ML models had the best goodness-of-fit and efficiency, and models with ML priors yielded estimates with the lowest standard errors. Crash frequencies increased with increases in traffic volume. Higher average speeds were associated with higher crash frequencies during peak periods, but not during off-peak periods. Several geometric design features including average segment length of arterial, number of lanes, presence of non-motorized lanes, number of access points, and commercial land use, were positively related to crash frequencies.

1. Introduction

Arterials carry most traffic on urban road networks, yet these are the least safe roads. In China's cities, the problem is especially severe with many urban arterials having unusual characteristics including short signal spacing, diverse cross-section configurations, mixed traffic flows, and high volumes (Xie et al., 2013). Previous studies showed roadway geometric and cross-section properties (e.g., segment length) and traffic flow characteristics (e.g., average annual daily traffic) to be important predictors of urban arterial safety (Greibe, 2003; El-Basyouny and Sayed, 2009; Gomes, 2013). However, these studies did not have an effective method to capture travel speeds throughout their study areas. The present study acquires average travel speeds on selected roadway segments throughout the network in addition to roadway geometric design features, land use, and traffic volumes. This broad spectrum of variables is expected to allow for a more complete analysis of the urban arterial safety problem.

In Shanghai, travel speeds on the arterial network were acquired using Global Positional System (GPS) data extracted from Shanghai's 50,000+ GPS equipped taxis. These taxis account for about 20% of the traffic on its urban arterials, and this data acquisition approach is hereafter referred to as the Floating Car Data (FCD) method. Compared to fixed-based methods such as radars or loop detectors typically used in speed-safety studies (Garber and Gadiraju, 1989; Baruya, 1998; Taylor et al., 2000), the FCD method is capable of capturing continuous speed data throughout the entire network over any time period of interest. The sample size of more than 15,000 GPS equipped taxis used in this study enabled the calculation of average speeds at an acceptable level of accuracy (Chen and Chien, 2000; Cheu et al., 2002). A total of 176 road segments from 18 arterials in Shanghai were selected for study. Bayesian (rather than classical) models were developed to investigate the relationship between geometric design features, land use, traffic flow, travel speeds, and crash frequency during both peak and off-peak periods.

The classical statistical approach relies on large random samples for accurate estimates, however, in much of transportation research, especially that which considers geometric or land use variables, limited data are available. It is therefore rarely

---

* Corresponding author at: School of Transportation Engineering, Tongji University, Shanghai 201804, China.
E-mail address: wangxs@tongji.edu.cn (X. Wang).

http://dx.doi.org/10.1016/j.aap.2015.07.004
0001-4575/

possible to conduct repeated studies under the same conditions, rendering the classical approach less than ideal. In contrast, another approach that can manage the small sample size problem is available. This approach starts by considering the prevailing conditions prior to the collection of sample data, and defines parameters that reflect these conditions to develop a model (that is, a set of predictive equations). After observing the sample, following Bayes' theorem, (Bayes and Price, 1763), provides a way to calculate probability using prior information or conditions. This Bayesian approach, obtains a posterior distribution for each the unknown parameters by considering both the prior distributions and the current conditions as reflected in the sample. When using the Bayesian approach, the prior distribution is fundamental and represents probabilities for each possible value of the unknown. Two types of prior distributions are used in Bayesian analysis: non-informative prior distributions that do not favor a particular value of the parameter, and are characterized by large variances (Lord and Miranda-Moreno, 2008), and informative prior distributions that assign higher probabilities to values of the parameter that are believed to be more likely to occur (Ntzoufras, 2011). These beliefs are often derived from previous experiments about the issues of interest. An important advantage of the Bayesian approach is its use of iterations to calculate posterior distributions. This has the effect of allowing the Bayesian approach to yield robust predictions in the absence of large samples.

2. Literature review

2.1. Previous urban arterial safety models

Previous safety research has improved our understanding of variables linked to crash rates on urban arterials. In much of this research the analytical models relied mainly on traffic volume, although variables such as segment length, number of intersections, number of access points, pedestrian traffic volumes, road width, lane width, and speed limits have sometimes been considered.

Mountain et al. (1996) developed crash prediction models for roads that included the number of minor junctions (between A-/B-roads and C-/unclassified roads), road type, carriageway type (single or dual) and speed limit. They found crash frequencies to be a positively accelerating function of exposure (e.g., annual traffic flow on segment length), and also to be positively associated with the number of minor junctions per kilometer. Bonneson and McCoy (1997) found Average Daily Traffic (ADT), driveway density, unsignalized public- street-approach density, and business land use were all positively correlated with crash frequencies, while raised-curb median treatments were correlated with reduced crash frequencies.

Greibe (2003) developed Generalized Linear Models (GLM) to predict crashes based on data from 142 km of roads located in urban areas in Denmark. The author divided these roads into 314 homogeneous segments averaging 450 m in length. Average Annual Daily Traffic (AADT) was the most powerful predictor of crashes, followed by speed limit, land use, the number of minor side roads per km and parking. Similarly, Turner et al. (2003) found that AADT was strongly associated with crashes for urban road sections in New Zealand.

EI-Basyouny and Sayed (2009) collected data from 58 arterials in Vancouver, British Columbia and found segment length, AADT, crosswalk density, business land use, unsignalized intersection density and the number of lanes were all positively associated with crash frequency. In a recent study, Gomes (2013) developed a crash prediction model for urban roads located in Lisbon, Portugal. He used a Negative Binomial (NB) model to examine crashes in relation to vehicle and pedestrian traffic flow, and highway geometric design features. The crash prediction model identified AADT, segment length, and lane-count greater than 4, as key variables associated with crash frequencies on urban roads.

Yanmaz-Tuzel and Ozbay (2010) conducted a before–after analysis on 4 urban arterials in New Jersey in which they investigated the effects of increases in lane width, installation of median barriers and guide rails, and improvements to vertical and horizontal road alignments. Using a full Bayesian approach, they observed reductions in crash rate (ranging from 23.1% to 28.6%) after adjusting the vertical and horizontal alignments (e.g., curve and grade). Increases in lane width, median barrier installations, and guide rail installations were associated with crash rate reductions of 28.1%, 14.3%, and 15%, respectively.

Das and Abdel-Aty (2011) examined the association of roadway and traffic characteristics including pavement, surface, shoulder, daylight, and street light conditions and surface width, shoulder width, speed limits, visual obstructions, presence of parking, median settings, vertical alignment, and ADT on the frequency and severity of rear-end crashes in Florida. They found higher ADTs were associated with increased crashes, and the absence of on-street parking was associated with reduced injury severity in crashes.

Dixit et al. (2011) investigated the relationship between arterial operations and crashes in the Orlando Metro using a two-fluid model. This model assumes traffic flow consists of running and stopped vehicles, and uses maximum speed and fraction of running vehicles (ratio of running time per mile to travel time per mile) to describe running speed. Running speed and travel time data were measured using a chase-car method in which drivers followed randomly selected vehicles for about a mile or until they left the arterial. They found higher speeds were positively correlated with both rear-end and angle crash rates.

The research to date has shown the most important variable predictive of crashes is traffic volume, while some of the less important variables are segment length, number of lanes, and number of access points. The relationship of speed to arterial crashes has not received much attention in this prior work. The recent availability of FCD in Shanghai offers a unique opportunity to examine travel speed in combination with other variables that affect urban arterial safety.

2.2. Statistical modeling methods

In past decades, Generalized Linear Models (GLM) such as Poisson-lognormal (PLN) and Negative Binomial (NB) have been commonly used to predict crash frequency (Shankar et al., 1995; Vogt and Bared, 1998; Abdel-Aty and Radwan, 2000). These traditional safety models assume study units are independent. This condition, however, might not be met in the current study because segments within single arterials share similar characteristics such as geometric design, signal control, and traffic flow.

Bayesian spatial models can incorporate hierarchical data structures to correct for spatial correlations (Guo et al., 2010; Huang and Abdel-Aty, 2010), and hierarchical structures can be implemented using random effect terms in Bayesian models (Huang and Abdel-Aty, 2010). EI-Basyouny and Sayed (2009) showed that model predictions can be improved by using spatial Poisson-lognormal models to incorporate random corridor effects. Yanmaz-Tuzel and Ozbay (2010) also showed that a hierarchical structure with random effects improves the predictability of PLN models because of its ability to adjust for spatial and temporal correlations.

Some attention has been directed at the effects of informative priors on model results when adopting the Bayesian approach. Yu and Abdel-Aty (2013) showed the benefits of different informative priors including the two-stage Bayesian priors, the