Safety modeling of suburban arterials in Shanghai, China

Xuesong Wang\textsuperscript{a,b,*}, Yang Song\textsuperscript{a}, Rongjie Yu\textsuperscript{a}, Grant G. Schultz\textsuperscript{c}

\textsuperscript{a} School of Transportation Engineering, Tongji University, Shanghai 201804, China
\textsuperscript{b} Key Laboratory of Road and Traffic Engineering of the State Ministry of Education, Tongji University, Shanghai 201804, China
\textsuperscript{c} Department of Civil & Environmental Engineering, Brigham Young University, 368 Clyde Building, Provo, UT 84602, USA

\textbf{A R T I C L E   I N F O}

\textbf{Article history:}
Received 16 October 2013
Received in revised form 7 April 2014
Accepted 12 April 2014
Available online 4 May 2014

\textbf{Keywords:}
Suburban arterial
Safety analysis
Hierarchical Bayesian model
Bivariate Bayesian models
Risk factor

\textbf{A B S T R A C T}

As urbanization accelerates in Shanghai, land continues to develop along suburban arterials which results in more access points along the roadways and more congested suburban arterials; all these changes have led to deterioration in traffic safety. In-depth safety analysis is needed to understand the relationship between roadway geometric design, access features, traffic characteristics, and safety. This study examined 161 road segments (each between two adjacent signalized intersections) of eight suburban arterials in Shanghai. Information on signal spacing, geometric design, access features, traffic characteristics, and surrounding area types were collected. The effect of these factors on total crash occurrence was investigated. To account for the hierarchical data structure, hierarchical Bayesian models were developed for total crashes. To identify diverse effects on different crash injury severity, the total crashes were separated into minor injury and severe injury crashes. Bivariate hierarchical Bayesian models were developed for minor injury and severe injury to account for the correlation among different severity levels. The modeling results show that the density of signal spacing along arterials has a significant influence on minor injury, severe injury, and total crash frequencies. The non-uniform signal spacing has a significant impact on the occurrence of minor injury crashes. At the segment-level, higher frequencies of minor injury, severe injury, and total crashes tend to occur for the segments with curves, those with a higher density of access points, those with a higher percentage of heavy vehicles, and those in inner suburban areas. This study is useful for applications such as related engineering safety improvements and making access management policy.

© 2014 Elsevier Ltd. All rights reserved.


1. Introduction

In Shanghai, the urban area has been sprawling rapidly in recent years. In 2009, the area under construction within the city limits was 136 km\textsuperscript{2} (Bureau of Statistics of Shanghai, 2010). This construction area increased to 150 km\textsuperscript{2} in 2010, which consisted of 16.3% of the developed area in Shanghai (Bureau of Statistics of Shanghai, 2011). Most of the residential and commercial land development is occurring along suburban arterials. More and more people reside along these arterials and driveways are now directly linked to the arterials. The traffic composition on these suburban arterials is mixed and includes trucks, buses, cars, non-motorized vehicles, and pedestrians. The suburban arterial design originally followed highway standards to serve high speed and long distance traffic. However, these roadways are currently also used as busy city thoroughfares. The development in Shanghai’s suburban areas has been accompanied by the appearance of more and more cross traffic. This has had a negative effect on traffic safety on these roadways. More fatalities occurred in suburban areas than in the central urban area in Shanghai. This same situation is occurring in many rapidly developing Chinese cities.

The situation in China is not without precedence with regards to suburban safety problems. A similar problem arose in the United States (US) as its suburbs grew rapidly. In the US, however, research on arterial safety and access management has been ongoing for a long time. The US introduced access management guidelines several years ago to preserve arterial safety and maintain operational efficiency (Gluck et al., 1999). In China, there is still no specific guidance for signal spacing, access features, and other access management criteria. There is rarely careful consideration regarding safety during transportation planning and roadway design, especially in rapidly developing suburban areas. Inappropriate signal spacing exists due to a general lack of consideration of safety during the transportation planning process. Long and short (i.e., inconsistent) signal spacing currently exists, thus disturbing traffic flow and safety. The traffic composition on suburban arterials is also
complex, including heavy trucks, bicycles, and pedestrians. High truck volume has had a negative effect on pedestrian and bicyclist safety. Furthermore, there has been no in-depth analysis that can provide a basis for any such guidance in Shanghai.

The objective of this study is to analyze arterial safety for Chinese rapidly developing suburban areas with the Bayesian safety modeling approach. The relationship between safety and the significant risk factors at both arterial and segment-levels from the perspective of signal spacing, geometric design, access features, and traffic characteristics are investigated.

2. Literature review

Safety analysis for arterials is an important aspect of highway planning, design, and management. Substantial research has been conducted over the years on arterial safety. Abdel-Aty and Radwan (2000) analyzed the relationships among road geometric design, traffic characteristics, driver demographics and safety on arterials in Florida. Other studies have focused on impacts of the implementation of specific factors on traffic safety, including safety impacts of curbs (Baek and Hummer, 2008), horizontal curvature (Fitzpatrick et al., 2009), and median treatments and access density (Gattis et al., 2005; Schultz et al., 2011).

2.1. Arterial selection and segmentation

Arterial selection is the first step for the safety analysis. Baek and Hummer (2008) selected 199 arterial segments randomly from the suburban highways throughout the state of North Carolina. Gattis et al. (2005) selected a list of rural and suburban four-lane highways throughout the state of Arkansas. Other studies selected arterials based on a variety of selection criteria (Ackaah and Salifu, 2011; Dinu and Veeraragavan, 2011; Fitzpatrick et al., 2009; Robert et al., 2006).

Arterial roadways are generally long, based on the function they provide. As a result, it is necessary to divide arterials into shorter segments for analysis. If segments become too short, they may end up with unstable crash frequencies since traffic crashes are low probability events. On the other hand, segments that are too long may include variability in cross-section that could lead to inaccurate rate estimations. Maou and Lum (1993) proposed two methods for segmentation, fixed-length segments (e.g., 0.1 mile or 1 mile), or homogeneous segments (i.e., sections are homogeneous in major geometric design and traffic characteristics). Most studies ensure homogeneous cross-sectional design on each segment (Abdel-Aty and Radwan, 2000; Gattis et al., 2005; Hauer et al., 2004). One special case of homogeneous segmentation is to divide arterials by signalized intersections and exclude the functional areas of signalized intersections (Mauga and Kaseko, 2010; Wang and Chen, 2012). In this study, the segments between two consecutive signalized intersections were selected as the study unit, ruling out the functional area, the same treatment as Jiang et al. (2014).

2.2. Statistical modeling

Crash rate (Gattis et al., 2005; Mauga and Kaseko, 2010; Schultz et al., 2010a) or crash frequency (Abdel-Aty and Radwan, 2000; Ackaah and Salifu, 2011; Dinu and Veeraragavan, 2011) is usually used as the dependent variable in models for safety analysis. In addition, different types of crash data (Hauer et al., 2004; Pande and Abdel-Aty, 2009) and different severity levels of crashes (Hauer et al., 2004; Mauga and Kaseko, 2010; Robert et al., 2006) can also be used as dependent variables. For example, Hauer et al. (2004) established models for property damage only (PDO), injury, and total crashes, respectively.

Different kinds of dependent variables are associated with different statistical models. Tobit models are recommended for use when the dependent variable is continuous data, like crash rate, which can address the problem that crash rate is usually left-censored at zero (Anastasopoulos et al., 2012a,b). If the occurring probability of crash of certain type or severity level is discussed, logistic models can be used (Anastasopoulos and Mannering, 2011; Milton et al., 2008). When the dependent variables are crash frequencies, negative binomial (NB) models are suitable, which are used in this study and discussed below.

Different arterial selection methods connect with different statistical modeling approaches. If the segments are randomly selected from a large area, this can ensure all observations are independent of each other, and a NB model can be employed. Baek and Hummer (2008) selected segments randomly from the suburban highways throughout the state of North Carolina and a NB model was used for analysis, which is suitable under this condition.

Random effect should be addressed when developing models. Many studies have addressed random effect from the temporal correlation and spatial correlation by incorporating the random effect into modeling, achieving more accurate estimation than fixed effect models (Caliendo et al., 2013; Jiang et al., 2014; Yu et al., 2013). In this study, if the segments are from a few arterials, those segments along the same arterials share similarities in geometric design and traffic characteristics, leading to the existence of within-group correlation and random effect. The independence assumption of NB models often does not hold true due to the hierarchical data structure. Hierarchical models offer the proper method to address such a hierarchical data structure (Huang and Abdel-Aty, 2010; Olsen et al., 2011; Wang and Chen, 2012; Xie et al., 2013; Yu et al., 2013). Hierarchical models have greater explanatory power because they contain varying intercepts or varying slopes that denote the random effects, which can incorporate variables at the specific levels where their effects occur (Gelman and Hill, 2007).

NB models come under the category of fixed parameter models, where the model parameters are assumed to be the same for any individual observation in the sample. One problem with NB models is that they do not account for the potential heterogeneity attributable to unobserved characteristics across observations, also leading to random effect. In several studies, random parameter (RP) models have been used to address this problem, which can be viewed as an extension of random effect models (Anastasopoulos and Mannering, 2009, 2011; Caliendo et al., 2013; Dinu and Veeraragavan, 2011; El-Basyouny and Sayed, 2009; Venkataraman et al., 2013). RP models are capable of allowing some or all estimated parameters to vary across not only individual observations but also homogeneous groups, and provide the necessary flexibility for handling any heterogeneity (Xie et al., 2013). The results attained by use of the RP models provide a good reason to believe that the different variables do not influence crash occurrence by the same amount on all segments (Anastasopoulos et al., 2012a; Garnkowski and Mannner, 2011; Milton et al., 2008).

When crash frequencies are modeled at different levels separately using univariate models, the potential correlation existing among crash frequencies at different severity levels for the unit may bias parameter estimates (Ma et al., 2008). For example, an increase in the number of crashes that were classified as incapacitating injury would also be associated with some change in the number of crashes that were classified by other injury types, which set up a correlation among the various injury-outcome crash frequency models (Lord and Mannering, 2010). Unobserved effects at the roadway-segment level are also likely to be shared across severities and bring in significant estimation error (Anastasopoulos et al., 2012b). To address this problem, Caliendo et al. (2013) developed a bivariate NB model to model the frequency of crash occurrence on motorway tunnels in Italy, jointly applied to non-severe crashes

216 X. Wang et al. / Accident Analysis and Prevention 70 (2014) 215–224

216 X. Wang et al. / Accident Analysis and Prevention 70 (2014) 215–224
and severe crashes. Ma et al. (2008) developed a multivariate Poisson–lognormal model to address this problem, which simultaneously modeled crash frequencies by injury severity, and the results showed statistically significant correlations between crash frequencies at different levels of injury severity.

2.3. Risk factor analysis

Geometric design includes number of lanes, curvature, median width, shoulder width, and lane width. The literature has been mixed with respect to safety. For example, Abdel-Aty and Radwan (2000) found that more lanes associate with more crashes. Hauer et al. (2004) found that more on-the-road crashes occurred on tangent segments than curves, while Abdel-Aty and Radwan (2000) found that the curvature was positively correlated with the frequency of crash occurrence. Additional research found that wider medians reduced the likelihood for crash involvement (Abdel-Aty and Radwan, 2000; Gattis et al., 2005). Narrow shoulder width has been found to increase the likelihood for crashes (Abdel-Aty and Radwan, 2000; Baek and Hummer, 2008; Dinu and Veeraragavan, 2011), and narrow lane width has also been found to increase the likelihood of crash occurrence (Abdel-Aty and Radwan, 2000).

Signal spacing includes primarily density of signals and uniformity of signal spacing. Higher signalized intersection density along an arterial was shown to have a negative effective on safety (Gluck et al., 1999; Mauga and Kaseko, 2010). Long and uniform signal spacing was reported as the best for arterial safety (Gluck et al., 1999). With respect to access features, the impacts of access density were consistent in the literature, indicating that higher access density leads to more crashes and injuries (Akaah and Salifu, 2011; Baek and Hummer, 2008; Gattis et al., 2005; Schultz et al., 2007, 2010b). El-Basyouny and Sayed (2009) used RP models and found that crash frequency is expected to increase with the density of unsignalized intersections on the majority of the corridor segments. Raised medians were consistently reported as the safest median type, while undivided segments were the least safe (Gluck et al., 1999), and sections with two-way left turn lane (TWLTL) were reported to be less safe than raised medians (Gattis et al., 2005; Gluck et al., 1999).

When considering traffic characteristics, researchers have noted that more traffic would increase the possibility of interaction among vehicles and thus more crashes would occur (Abdel-Aty and Radwan, 2000; Baek and Hummer, 2008). Dinu and Veeraragavan (2011) developed RP models and found that the impact for logarithm of hourly volume can be expected to vary, showing positive signs for most of the highway segments. They found that higher percentages of heavy vehicles would cause fewer night-time crashes, but Robert et al. (2006) found that the proportion of heavy vehicles had a negative effect on traffic safety. Nilsson (1982) found that crash rate increases when average speed increases, especially for severe crashes. However, Baruya (1998a,b) found that the crash frequency increased more with lower average speed.

Other risk factors mainly include speed limit, land use type, and area type. When considering speed limit, most studies found that lower speed limits were associated with higher crash frequencies (Abdel-Aty and Radwan, 2000; Mauga and Kaseko, 2010), but Qin et al. (2005) found that the states of California and Washington showed a positive relationship between speed limit and the number of crashes. With respect to land use type, Mauga and Kaseko (2010) found that the roadways with residential land uses have fewer crashes than the commercial ones. El-Basyouny and Sayed (2009) found that for most roadway segments, business land use was positively associated with crash frequencies. Abdel-Aty and Radwan (2000) found that more crashes occurred on urban road segments than rural road segments.

In summary, the evaluation of the safety effects of median width, shoulder width, lane width, access density, median treatments, density of signalized intersections, annual average daily traffic (AADT), and the type of land use for arterials are consistent. But different studies have found different safety impacts for the horizontal curve, number of lanes, the percentage of trucks, average running speed, and speed limit. In addition, the impacts of uniformity of signal spacing and non-motorized facilities (e.g., openings of bicycle lane separators) on safety were rarely considered in previous studies. China has its own peculiar traffic safety characteristic and different road design standards, so it is necessary to carry out safety analysis for Chinese suburban arterials.

3. Data preparation

Both arterial– and segment– level data were collected for the study. Arterial– level variables included density of signals along arterials (DOSP) and standard deviation of signal spacing (SDSP). Segment–level variables were divided into four categories: geometric design, access features, traffic characteristics, and area type. In the Chinese crash report, the crash severity level is recorded as PDO, slight injury, severe injury, and fatal. Effects of risk factors on total crash occurrence were investigated. To analyze the risk factors of different types of crashes, the total crashes were separated into minor injury and severe injury crashes. PDO and slight injury crashes were combined into “minor injury crashes,” while severe injury and fatal crashes were combined as “severe injury crashes.” Crash data in 2009 were collected for analysis. Frequencies of minor injury, severe injury, and total crashes per kilometer on each segment were used in the models.

3.1. Arterials selection and segmentation

In Shanghai, the suburbs are defined as the area outside the Outer Ring Road as shown in Fig. 1. The suburban areas are further subdivided into the inner suburb and the outer suburb by the Suburb Ring Road. The area between the Suburb Ring Road and the Outer Ring Road is considered as the inner suburb, while the area outside the Suburb Ring Road is considered as the outer suburb. The selected arterials were located in the suburban areas of Shanghai.

Crash data for 2009 was used for analysis, therefore the arterials under construction in 2009 were not included in the analysis to avoid construction impact on safety. Eight arterials were selected as shown in Fig. 1. The selected arterials comprised different numbers of lanes, median treatments, and area types.

Table 1 lists the area types, number of lanes, median treatment, and other key features of the selected arterials. The arterials comprise different area types, median treatments, and numbers of lanes. For those variables with more than one value on the same arterial (i.e., area types, median treatments, and the number of lanes), all possible values were listed.

The segments between two consecutive signalized intersections of any selected arterial were treated as a study unit. Each segment was selected to be homogeneous in the number of lanes and median treatments. In total, there were 161 segments identified from the eight selected suburban arterials. In order to focus the results on the roadway segments, the intersection functional areas were removed from the analysis. Because it has been shown that different intersections have different intersection functional areas, and considering that intersection functional area is not the primary focus of this study, the value of 76 m was utilized in this study (Wang et al., 2008).
3.2. Arterial-level variables

Arterial-level variables are focused on signal spacing (density and standard deviation). Signal spacing refers to the distance between every two consecutive signalized intersections. Signal spacing for the eight arterials was calculated based on the Geographic Information System (GIS) map and validated using Google Earth® (Google Inc., 2010). Based on the original information on signal spacing, two arterial-level variables were derived, DOSP and SDSP. DOSP is the density of signals along arterials. SDSP is the standard deviation of signal spacing which denotes the uniformity of signal spacing. The higher the SDSP, the less uniform are the signal spacings. The two arterial-level variables and their descriptive statistics are listed in Table 2.

3.3. Segment-level variables

Different sources were used to collect the data for geometric design, access features, traffic characteristics, and area type. The geometric design, access features, and area type were obtained from Google Earth® satellite images. Traffic characteristics were acquired from the traffic operational data on Shanghai’s highways. Any unclear information about median treatments, presence of an access, and traffic volume was confirmed by field surveys.

Geometric design variables included horizontal alignments, segment lengths, and the number of lanes in both directions. Horizontal alignments were divided into two categories: tangent segments (54.3%) and curved segments (45.7%). The index for identifying whether a segment was considered a tangent segment or a tangent segment.
curved segment was the steering angle of one segment, which corresponds to the acute angle intersected by the two tangents of one segment’s starting point and ending point. There were no segments whose angle ranged from 7 to 33 degrees. In this study, the curved segments were defined as the angle from 33 to 70 degrees; the tangent segments were less than 7 degrees. Segment length was calculated as the length between two consecutive signals minus the functional areas.

Access features included access density, median treatments, and the density of openings of bicycle lane separators (DOBLS). There are bicycle lanes on each of the eight arterials and turning bicycle causes conflicts with motorized traffic, which have negative effects on safety. The values of 10 openings/km and 20 openings/km were used as thresholds and DOBLS were divided into three levels. The value of 0–10 was labeled as low density; 10–20 was labeled as medium density; more than 20 was high density.

Traffic characteristics included average running speed, monthly average daily traffic (MADT), and daily volumes of cars, buses, minivans, medium trucks, trucks, and trailers separately. The running speed in this study was the average spot speed for day-time traffic (from 6:00 a.m. to 6:00 p.m.). High signal density and high access density lowered the average running speed. The unit of MADT is passenger car unit (pcu) and the unit of other traffic volume variables is vehicles per day (veh/day). Heavy vehicles used in the analysis included medium trucks, trucks, and trailers. The safety impacts of other treatments of traffic composition, such as percentage of certain traffic, logarithm of traffic volume, volume per lane, and logarithm of volume per lane, were also investigated.

Compared with the outer suburb, the inner suburb is more developed. More traffic volume, access points, and commercial land exist for the arterial segments within the inner suburb area. In addition, the signal spacing is shorter in the inner suburb than in the outer suburb. This necessitates different safety impacts and needs to be studied in the models. The 15 segment-level variables with brief descriptions and their descriptive statistics are listed in Table 2.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Description</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Arterial-level variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOSP</td>
<td>Number of signalized intersections per kilometer along arterial</td>
<td>0.892</td>
<td>1.857</td>
<td>1.020</td>
<td>0.300</td>
</tr>
<tr>
<td>SDSP (km)</td>
<td>Standard deviation of signal spacing on arterials</td>
<td>0.368</td>
<td>0.698</td>
<td>0.553</td>
<td>0.140</td>
</tr>
<tr>
<td><strong>Segment-level variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Segment length (km)</td>
<td>The length between two consecutive signals excluding the functional areas</td>
<td>0.248</td>
<td>2.061</td>
<td>0.795</td>
<td>0.558</td>
</tr>
<tr>
<td>Horizontal alignments</td>
<td>For curved segments, the angle ranged from 33 to 70 degrees in this study. For tangent segments, the angle was less than 7 degrees. 0 for curve, 1 for tangent</td>
<td>0</td>
<td>1</td>
<td>0.553</td>
<td>0.499</td>
</tr>
<tr>
<td>The number of lanes on both sides</td>
<td>Density of openings of bicycle lane separator. 0 for low density, 1 for medium density, 2 for high density</td>
<td>2</td>
<td>6</td>
<td>3.587</td>
<td>1.590</td>
</tr>
<tr>
<td>DOBLS</td>
<td>Density of openings of bicycle lane separator. 0 for low density, 1 for medium density, 2 for high density</td>
<td>0</td>
<td>36.424</td>
<td>1.158</td>
<td>7.63</td>
</tr>
<tr>
<td>Access density</td>
<td>Number of accesses per km along road segment on both sides</td>
<td>0</td>
<td>29.167</td>
<td>5.250</td>
<td>4.820</td>
</tr>
<tr>
<td>Median treatment</td>
<td>0 for no median, 1 for raised median</td>
<td>0</td>
<td>1</td>
<td>0.681</td>
<td>0.445</td>
</tr>
<tr>
<td>MADT (pcu)</td>
<td></td>
<td>6184</td>
<td>61,590</td>
<td>18,200</td>
<td>10,990</td>
</tr>
<tr>
<td>Daily volume of cars (veh/day)</td>
<td></td>
<td>1920</td>
<td>15,168</td>
<td>9230.600</td>
<td>7268.960</td>
</tr>
<tr>
<td>Daily volume of buses (veh/day)</td>
<td></td>
<td>0</td>
<td>933</td>
<td>306.500</td>
<td>289.500</td>
</tr>
<tr>
<td>Daily volume of minivans (veh/day)</td>
<td>Minivan includes small trucks</td>
<td>168</td>
<td>3744</td>
<td>1841</td>
<td>933</td>
</tr>
<tr>
<td>Daily volume of medium trucks (veh/day)</td>
<td></td>
<td>296</td>
<td>4032</td>
<td>923</td>
<td>363</td>
</tr>
<tr>
<td>Daily volume of trucks (veh/day)</td>
<td></td>
<td>192</td>
<td>3552</td>
<td>287</td>
<td>256</td>
</tr>
<tr>
<td>Daily volume of trailers (veh/day)</td>
<td></td>
<td>0</td>
<td>1728</td>
<td>533</td>
<td>285</td>
</tr>
<tr>
<td>Average running speed (km/h)</td>
<td>Spot speed from 6:00 a.m. to 6:00 p.m.</td>
<td>26</td>
<td>47</td>
<td>31.600</td>
<td>10.111</td>
</tr>
<tr>
<td>Area types</td>
<td>0 for outer suburb, 1 for inner suburb</td>
<td>0</td>
<td>1</td>
<td>0.832</td>
<td>0.469</td>
</tr>
</tbody>
</table>

### 3.4. Crash data

In China, the police reported crashes have two recording categories: simple procedure and normal procedure. The simple procedure applies for PDO and slight injury crashes. Crashes recorded by normal procedure are severe injury and fatal crashes. All police-recorded crash data in 2009 for the selected arterials were collected from the traffic police department. The geocoding procedure in ArcGIS® (Esri Inc., 2010) was used to locate the crashes on the GIS base map of the selected arterials according to the location description in each crash report. First, original crash information was imported into ArcGIS®. If the crash was recorded by the segment where the crash occurred, by the intersecting roadway, or by the distance from an adjacent signal, a program developed in ArcGIS was used to locate this crash automatically on the map. If the crash was recorded by house number, nearby building name, or bridge along the arterial, crashes were located on the GIS map manually by checking roadway information. Frequencies of crashes per kilometer on segments were calculated and then values were rounded up to integer values. Gattis et al. (2005) and Olsen et al. (2011) applied this method to treat crash rate for count data models.

### 4. Hierarchical Bayesian model

As shown in Fig. 2, the data used in this study can be considered as a two-level structure with Level 1 being the arterial–level, and Level 2 being the segment–level, where j is the number of arterials from one to eight, and nj is the number of the specific segment on the jth arterial. In the hierarchical model proposed in this study, parameters at the segment–level will be expressed by probability models at the arterial–level.

Where hierarchical data structure exists, hierarchical models are able to make more reliable estimations than traditional models (i.e., NB models) because hierarchical models can accommodate the heterogeneity among different groups (Huang and Abdel-Aty,
Hierarchical models are capable of including covariates at the segment and arterial-levels, and allow the effects of segment and arterial-level variables to be independently evaluated (Gelman and Hill, 2007).

Crash frequencies of different severities require a multivariate model because the response is multi-dimensional. Moreover, in practice, variables (such as access density and signal spacing) may simultaneously affect all crashes at different levels of severity for a particular roadway segment, thus introducing severity level correlation (Ma et al., 2008). Bivariate models were established simultaneously for minor injury crashes and severe injury crashes, and univariate models were modeled for total crashes.

In recent years, Bayesian methods applied in hierarchical models have attracted more and more attention. Several studies have shown the advantages of Bayesian methods over classical statistical methods in achieving valid results (Jiang et al., 2014; Olsen et al., 2011; Qin et al., 2005; Wang and Chen, 2012; Xie et al., 2013). Bayesian models view parameters as random variables that are characterized by a prior distribution. Instead of achieving maximum likelihood estimates (MLE) for the studied unknowns based solely on the sample data in MLE inference, the essential advantage of Bayesian methods is the explicit use of probability for quantifying uncertainty in inferences based on statistical data analysis (Spiegelhalter et al., 2003a). To accommodate the hierarchical data structure and within-group correlation, hierarchical negative binomial (HNB) models in a Bayesian framework were applied. The theoretical framework for Bayesian inference can be expressed as outlined in Eq. (1).

\[ \pi(\theta|y) = \frac{L(y|\theta) \pi(\theta)}{\int L(y|\theta) \pi(\theta) d\theta} \]

where \( y \) is the vector of observed data, \( \theta \) is the vector of parameters required for the likelihood function, \( L(y|\theta) \) is the likelihood function, \( \pi(\theta) \) is the prior distribution of \( \theta \), \( \int L(y|\theta) \pi(\theta) d\theta \) is the marginal distribution of observed data, and \( \pi(\theta|y) \) is the posterior distribution of \( \theta \) given \( y \).

A univariate hierarchical negative binomial (UHNB) model has been proposed in the Bayesian framework for total crashes. The specific structure of hierarchical model can be expressed in Eq. (2) for segment-level models and in Eqs. (3) through (7) for arterial-level models.

**Segment-level model**: \( y_{i,j} \sim \text{Negbin}(\lambda_{i,j}, k) \) (2)

**Arterial-level model**: \( \log \{\lambda_{i,j}\} = \beta_{0,i,j} + \sum_{p=1}^{7} \beta_{p,i,j} X_{p,i,j} \) (3)

\[ \beta_{0,i,j} = \gamma_{0,0} + \sum_{q=1}^{2} \gamma_{0,q} W_{q,i,j} + \epsilon_{0,i,j} \]  

\[ \beta_{1,i,j} = \gamma_{1,0} \]  

\[ \beta_{2,i,j} = \gamma_{2,0} \]  

\[ \beta_{p,i,j} = \gamma_{p,0} \] (7)

where \( i \) is the serial number of segments on one arterial and start from 1, \( j \) is the serial number of arterials and from 1 to 8, \( p \) is the serial number of segment-level independent variables and from 1 to 7, \( q \) is the serial number of arterial-level independent variables and from 1 to 2, \( y_{ij} \) is the crash frequencies on a certain segment, \( \lambda_{ij} \) is the expectation of \( y_{ij} \), \( k \) is the over-dispersion coefficient, \( X_{p,i,j} \) is the segment-level variable, \( \beta_{0,i,j} \) is the model intercept of the \( j \)th arterial, \( \beta_{p,i,j} \) is the coefficient of the \( p \)th segment-level independent variable of the \( j \)th arterial, \( \gamma_{0,0} \) is the model intercept of \( \beta_{0,i,j} \) for certain severity level, \( \gamma_{p,0} \) is the regression coefficient of the \( p \)th segment-level independent variable for certain severity level, \( \gamma_{q,0} \) is the regression coefficient of the \( q \)th arterial-level independent variable for certain severity level.
for certain severity level, the random effect at arterial-level where \( \epsilon_{ij} \sim N(0, \sigma^2) \), and \( \sigma^2 \) is the variance of random effect.

The Deviance Information Criterion (DIC), defined in Eq. (14) can be used to compare the complex models. It offers a Bayesian measure of model fitting and complexity (Spiegelhalter et al., 2003a). Models with smaller DIC are preferred.

\[
DIC = D(\hat{\theta}) + p_D
\]

where \( D(\hat{\theta}) \) is the Bayesian deviance of the estimated parameter, \( D(\hat{\theta}) \) is the posterior mean of \( D(\hat{\theta}) \), viewed as a measure of model fitting, and \( p_D \) is the effective number of parameters, indicating complexity of models.

The Bayesian method is usually implemented using a Markov Chain Monte Carlo (MCMC) algorithm. Open source software WinBUGS® (Spiegelhalter et al., 2003b) was used to provide a computing approach for calibration of Bayesian models using MCMC. In the process of developing models, because of convergence and time of updating, two MCMC chains of 20,000 iterations were run, and the first 2000 samples were discarded as burn-in.

5. Modeling results

A UHNB model was developed for total crashes, while minor injury and severe injury crashes were analyzed using BHNB models.

5.1. Total crash modeling results

Upon completion of the model runs, those variables that were significant at least at the 90% level were retained in the models. The summary of UHNB model fitting for total crashes is reported in Table 3. Correlation analyses showed that the correlation coefficients among the independent variables in Table 3 were less than 0.38. The condition of identifying the significant variables was that the 95% Bayesian credible interval (95% BCI) does not include 0 (Spiegelhalter et al., 2003b). If a variable meets the above condition, then it can be viewed as significant.

The variable var(int) denotes random effect of \( \epsilon_{ij} \) in Eq. (4). It is significant and affirms the presence of within-group correlation. The estimated dispersion value \( k \) provides strong evidence of overdispersion.

5.2. Minor injury and severe injury crash modeling results

The summary of BHNB model fitting for minor injury, severe injury crashes is reported in Table 4. Those variables that were significant at least at the 90% level were retained in the models. Correlation analyses showed that the correlation coefficients among the independent variables in Table 4 were less than 0.34. The variable var(int) denotes random effect of \( \epsilon_{ij} \) in Eq. (4). It is significant and affirms the presence of within-group correlation. The estimated dispersion value \( k \) provides strong evidence of overdispersion. The correlation coefficient \( \rho \) was 0.423 in the BHNB, indicating the existence of severity level correlation. The DIC value for bivariate NB models of minor injury crashes and severe injury crashes is 1074.3. The DIC values for UHNB models are 826.64 for minor injury crashes and 287.85 for severe injury crashes. The BHNB models perform better than the corresponding bivariate NB models and UHNB.

6. Variable interpretation

Two arterial-level variables (DOSP and SDS) and four segment-level variables (access density, horizontal curve, percentage of heavy vehicles, and area types) are significant in at least one model. The safety impacts of these variables are investigated in the following sections based on the best performing models.

6.1. Arterial-level variables

Signalized intersection density was found to be positively correlated with crash occurrence on the suburban arterials in three models. This finding is consistent with previous studies (Gluck et al., 1999; Mauga and Kaseko, 2010). It can be explained that because more intersections are correlated with an increase in the number of vehicles turning and changing lanes on the segments which will interfere with traffic flow, conflicts among vehicles would also increase which lead to increase of crash occurrence probability. For the UHNB and BHNB models, the coefficients of DOSP are 1.109, 0.597, and 1.002 (openings/km) for minor, severe, and total crashes, respectively. When other factors are held constant, the number of crashes per km will predict an increase in expected crash frequency of about 20% (e^{1.109} - 1), 81.7% (e^{0.597} - 1), and 172% (e^{1.002} - 1) for minor, severe, and total crashes, respectively.

The total crashes rate (crashes per 10^4 vehicle kilometers traveled per lane) of the arterial segments is presented in Fig. 3, ordered by DOSP. The segments in the inner suburb and outer suburb are presented in the two figures, respectively. Fig. 3 shows that the total crash rate increases with the increase of DOSP for most arterial segments.

In the model of minor injury crashes, SDS is a significant variable. This variable demonstrates the irregularity of signal spacing, which had a negative impact on safety. Gluck et al. (1999) introduced that long and uniform signal spacing can achieve efficient traffic operation and improve safety. However, it has rarely been proven in previous studies. Irregular signal spacing will disrupt traffic operation and make drivers frequently accelerate and decelerate. The subsequent increase in speed variation decreases safety. Signal spacing should be given more attention in future applications of improving traffic safety.

6.2. Segment-level variables

According to the UHNB and the BHNB models, access density was found to have an increasing effect on minor crashes, severe crashes, and total crashes. Unsurprisingly, there are many studies that have analyzed the safety effect of access density and similar conclusions have been reached. Papayannoullis et al. (1999) found that crashes per million vehicle miles traveled will increase by 40% when access density increases from 6 openings/km to 13 openings/km. Residents along suburban arterials set up private accesses for their own convenience. Vehicles can enter or drive out of these
Table 4
Posterior summary of BHNB models.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Minor injury crashes</th>
<th>Severe injury crashes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>95% BCI</td>
</tr>
<tr>
<td>Arterial-level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOSP</td>
<td>1.109 (0.587)</td>
<td>(0.497, 1.85)</td>
</tr>
<tr>
<td>SDSP</td>
<td>0.177 (0.158)</td>
<td>(0.012, 0.345)</td>
</tr>
<tr>
<td>Segment-level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access density</td>
<td>0.0678 (0.0442)</td>
<td>(0.0212, 0.124)</td>
</tr>
<tr>
<td>Tangent vs. curve</td>
<td>-0.297 (0.163)</td>
<td>(-0.554, -0.068)</td>
</tr>
<tr>
<td>Percentage of heavy vehicles</td>
<td>4.46 (3.26)</td>
<td>(1.18, 6.93)</td>
</tr>
<tr>
<td>Inner suburb vs. outer suburb</td>
<td>0.609 (0.387)</td>
<td>(0.197, 0.996)</td>
</tr>
<tr>
<td>(k)</td>
<td>2.325 (0.749)</td>
<td>(1.23, 3.14)</td>
</tr>
<tr>
<td>(\rho) var(int)</td>
<td>0.423</td>
<td>(0.0976, 0.387)</td>
</tr>
<tr>
<td>DIC</td>
<td>1027.5</td>
<td></td>
</tr>
</tbody>
</table>

Note: For statistically significant variables, their mean and SD are in bold.

accesses freely. More accesses result in more conflict points and deteriorated safety.

Compared with tangent segments, the model results indicate that more crashes occurred on curved segments. The average length of tangent segments is relatively short at 0.649 km, which would tend to eliminate crashes caused by visual fatigue. There are more and more buildings along suburban arterials because of land development in the suburban area. When drivers are turning, sideswipe or rear-end crashes may occur due to sight distance restrictions. Dinu and Veeraragavan (2011) found that the horizontal curvature of the roadway was positively correlated with day-time crash frequency.

Fig. 3. Relationship of total crash rates and DOSP for selected arterial segments in the (a) inner suburb and (b) outer suburb.
Higher percentages of heavy vehicles were also found to be correlated with more crashes. Previous studies (Abdel-Aty and Radwan, 2000) have found that truck proportion was positively correlated with crash occurrence, which is consistent with this study. Traffic composition has an impact on safety. Robert et al. (2006) studied the impact of mixed nature of traffic in India on crash occurrence, and they found truck percentage was positively correlated with crashes. Vehicles’ frequent lane changing and turning movements would increase conflicts due to large speed differences between cars and trucks. Drivers’ sight would be obstructed by trucks because of trucks’ large size. Besides, trucks need longer stopping distance due to trucks’ heavy weight, which would also damage safety levels on segments.

Compared with the outer suburb, crash frequency is higher for the arterial segments in the inner suburb, which is related with the land use types and developing intensity of land use. Some researchers found a relationship between crash rate and developing intensity of land use: the probability of crashes was higher in more developed land, such as commercial land (Ackaah and Salifu, 2011; Gattis et al., 2005; Schultz et al., 2010a). In the inner suburban area, the land was developed more intensively, and more people reside along the arterials. Traffic safety is negatively affected by the existence of accesses, and more access points exist in the inner suburb compared with the outer suburb. The probability of crashes is therefore higher in the inner suburb. In addition, greater traffic volumes in the inner suburb provide more opportunities for exposure to conflicts and result in more crashes.

7. Conclusions and discussion

The purpose of this study was to provide a safety analysis for Chinese suburban arterials and thorough understanding of the relationships of arterial design to traffic safety. First, a suitable modeling approach was proposed. Next, the impact of risk factors at both the segment and arterial-levels in signal spacing, geometric design, access feature, traffic characteristic, and area type were identified. A total of 161 segments from eight suburban arterials in Shanghai were selected for study. Frequencies of minor injury, severe injury, and total crashes per km were selected as response variables. The effect of risk factors on total crash occurrence was investigated first. In order to capture the hierarchical data structure, UHNB models were developed for total crashes. For the purpose of identifying diverse effects on different crash injury severity, the total crashes were separated into minor injury and severe injury crashes. To account for the correlation among different severity levels, BHN model for minor injury and severe injury crashes was developed. In total, five variables at the segment-level and two variables at the arterial-level were chosen in the models. The values of DIC showed that the BHN models performed better than the bivariate NB models and UHNB models. A total of four variables at the segment-level and two variables at the arterial-level were chosen in the models.

According to the modeling results, four variables at the segment-level (access density, horizontal alignment, percentage of heavy vehicles, and area types) and two variables at the arterial-level (DOSP and SDSP) have statistically significant impacts on minor injury crash occurrence. For severe injury and total crashes, four variables at the segment-level (access density, horizontal alignment, percentage of heavy vehicles, and area types) and one variable at the arterial-level (DOSP) were significant.

DOSP and SDSP were chosen to represent the density and uniformity of signal spacing along the arterial. SDSP was rarely considered before. Long and uniform signal spacing has been determined as the best for traffic safety and operation (Gluck et al., 1999). According to our modeling results, higher density of signals and non-uniform signal spacing would lead to more crash occurrence. Long and uniform signal spacing on an arterial was beneficial for arterials safety according to research by Gluck et al. (1999). Therefore, signal spacing should be given more attention during the transportation planning process.

Many studies have shown the safety impact of access points (Ackaah and Salifu, 2011; Dihu and Veeraragavan, 2011; Fitzpatrick et al., 2009; Gattis et al., 2005; Mauga and Kaseko, 2010). There is a need to make access design standards and laws about opening accesses along suburban arterials in the future to solve this problem. Traffic safety managers should focus more on inner suburb curve segments, because more crashes occurred on these segments. The results of this study indicate that a larger percentage of heavy vehicles will also cause more crashes, which is consistent with previous studies (Hauer et al., 2004; Pande and Abdel-Aty, 2009). DOBLS represented the relationship between non-motorized traffic and safety, which was found to have a positive correlation with crashes, though only significant at the 90% level in the model of severe crashes with a positive coefficient. Risk of non-motorized vehicles is a common safety issue in developing countries, including China. More attention should be given to the management of non-motorized vehicles and heavy vehicles.

In the light of the above results, the models developed for Shanghai suburban arterials in this paper appear to be useful with many applications such as setting access spacing and signal spacing standards for engineering safety improvements, impacting access management policy for Shanghai. However, there are several factors that were not considered in the analysis, such as lighting, weather, and above all, human behavior, which are also heavily influential on crash occurrence. The impacts of DOBLS and signal spacing were discussed in statistical models, but should be deeply analyzed in the future. In addition, the segments without intersection functional areas were discussed in this study, but access points located within intersection functional areas should also be investigated in further studies, as they also affect safety in China.

Considering the distinct characteristics of travel behavior and mixed traffic composition, identifying significant risk factors for suburban arterial roads in China and other fast growing Asian cities is of great significance. This study investigated suburban arterials’ risk factors under the circumstance of rapid urban development in Shanghai. For many rapidly developing Chinese cities, there are still no studies analyzing land development and safety along suburban arterials. There could be some similar risk factors among these cities, but the magnitude and safety impact of these factors in different cities should be investigated, respectively, due to different traffic flow, driving behavior, access management policy, local standards and so on. Therefore the model developed in this study would not be directly transferred to other cities. Researchers in other Chinese and Asian cities should take local characteristics into account and conduct in-depth safety analysis to understand problems and ultimately improve safety.

Acknowledgments

This study was jointly sponsored by the Chinese National Science Foundation (51008230), the program for New Century Excellent Talents in the University (NCET-11-0387), and the Fundamental Research Funds for the Central Universities.

References

