Accident Analysis and Prevention 113 (2018) 38–46

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

A cross-comparison of different techniques for modeling macro-level cyclist crashes

Yanyong Guo, Ahmed Osama, Tarek Sayed*

Department of Civil Engineering, The University of British Columbia, 6250 Applied Science Lane, Vancouver, BC, V6T 1Z4, Canada

ARTICLE INFO

Keywords:
Cyclist crashes
Macro-level crash models
Spatial effects
Random parameters model
Full Bayesian estimation

ABSTRACT

Despite the recognized benefits of cycling as a sustainable mode of transportation, cyclists are considered vulnerable road users and there are concerns about their safety. Therefore, it is essential to investigate the factors affecting cyclist safety. The goal of this study is to evaluate and compare different approaches of modeling macro-level cyclist safety as well as investigating factors that contribute to cyclist crashes using a comprehensive list of covariates. Data from 134 traffic analysis zones (TAZs) in the City of Vancouver were used to develop macro-level crash models (CM) incorporating variables related to actual traffic exposure, socio-economics, land use, built environment, and bike network. Four types of CMs were developed under a full Bayesian framework: Poisson lognormal model (PLN), random intercepts PLN model (RIPLN), random parameters PLN model (RPPLN), and spatial PLN model (SPLN). The SPLN model had the best goodness of fit, and the results highlighted the significant effects of spatial correlation. The models showed that the cyclist crashes were positively associated with bike and vehicle exposure measures, households, commercial area density, and signal density. On the other hand, negative associations were found between cyclist crashes and some bike network indicators such as average edge length, average zonal slope, and off-street bike links.

1. Introduction

Cycling has been promoted worldwide as a sustainable mode of transportation and is identified as a leading driver for a healthy, livable and resource-efficient environment. In particular, cycling has benefits related to traffic congestion, energy consumption, and air pollution mitigation. Moreover, cycling is preferred by many commuters since it is the fastest mode for trips that are less than 5 km (Dejister and Schollaert, 1999). However, vulnerable road users such as cyclists suffer from an elevated risk of crash involvement. As such, concerns about cyclist crashes can discourage the public from using bikes, which raises the need to improve cyclist safety through studying the underlying factors that may lead to crashes.

Proactive road safety management using macro-level crash models (CMs) has been advocated as an efficient safety analysis approach due to their benefit of addressing safety issues before crashes emerge (de Leur and Sayed, 2003). These models have many applications such as evaluating the effectiveness of safety improvement measures, detecting and ranking of crash-prone locations, and estimating the safety of various transportation and land use planning options (Sawalha and Sayed, 2001).

However, one important issue that requires consideration while developing the CMs is the effect of unobserved heterogeneity on the estimated model parameters. Overlooking the unobserved heterogeneity can lead to a biased, inefficient, and erroneous inferences Washington et al. (2010). To address this issue, different full Bayesian (FB) modeling approaches were applied in previous studies such as spatial effects, random intercepts, and random parameter approaches among others (El-Basyouny and Sayed, 2009a,b; Anastasopoulos and Mannering, 2009; Siddiqui et al., 2012; Chen and Tarko, 2014).

The objective of this study is twofold: a) investigating the best FB modeling approach for tracking the unobserved heterogeneity within the macro-level cyclist CMs; and b) evaluating the contributory factors that are associated with cyclist crashes using a comprehensive and original dataset. Four different types of macro-level CMs, i.e. Poisson lognormal model (PLN), random intercepts PLN model (RIPLN), random parameters PLN model (RPPLN), and spatial PLN model (SPLN), are developed for 134 traffic analysis zones (TAZs) in the city of Vancouver, British Columbia. The CMs incorporate several area-wide variables including socio-economics, land use, built environment, and cycling network indicators in order to investigate the impacts of those variables on cyclist safety and improve the CM fit. Additionally, a cyclist exposure measure (bike kilometers travelled) and a vehicle exposure measure (vehicle kilometers travelled) are incorporated in the

* Corresponding author.
E-mail addresses: yanyong.guo@ubc.ca (Y. Guo), ahmed.osama@eng.asu.edu.eg (A. Osama), tsayed@civil.ubc.ca (T. Sayed).

https://doi.org/10.1016/j.aap.2018.01.015
Received 28 September 2017; Received in revised form 12 January 2018; Accepted 13 January 2018
0001-4575/

CMs. The results of this study should provide new insights into FB macro-level crash modeling techniques and a better understanding of the factors associated with cyclist safety.

2. Literature review

2.1. Cyclist safety covariates

A variety of exposure variables were investigated in the cyclist CMs in the previous studies (e.g. Osama and Sayed, 2016; Osama and Sayed, 2017), including bike traffic volume (Miranda-Moreno et al., 2011), vehicle traffic volume (Hamann and Peek-Asa, 2013), bike lane length (Wei and Lovegrove, 2013), and length of roads (Siddiqui et al., 2012). Results from the previous studies showed positive association between exposure and cyclists crashes. Also, socioeconomic variables such as population, employment, and household income were reported to be positively associated with the frequency of cyclist crashes (Siddiqui et al., 2012). Moreover, land use was found to have significant impacts on cyclist crashes, where the increase in commercial land use, residential area percentage, industrial area percentage, and land use balance mix were positively associated with cyclist crashes (Narayanamoorthy et al., 2013; Vandenbulcke et al., 2014). As for the built environment, factors such as bus stop density (Wei and Lovegrove, 2013; Strauss et al., 2013), parking sign density (Chen, 2015), traffic signal density (Chen, 2015; Wei and Lovegrove, 2013), and street lighting (Siddiqui et al., 2012), were found positively associated with cyclist crashes.

Recently, the impact of road network features on cyclist crashes was investigated by several researchers. For example, Chen et al. (2012) showed that the installation of bike lanes did not lead to additional crashes, but a possible increase in the number of cyclists instead. Vehicle lanes and intersection density were found positively associated with cyclist crashes (Siddiqui et al., 2012; Wei and Lovegrove, 2013). Also, high-speed streets (Siddiqui et al., 2012), and on-road bike lane (Reynolds et al., 2009; Hamann and Peek-Asa, 2013) were found positively associated with cyclist crashes (Osama and Sayed, 2017).

2.2. Crash modeling

The most common technique for modeling crashes is the Poisson-Gamma hierarchy that leads to the Negative Binomial (NB) regression model (Sawalha and Sayed, 2001; Guo et al., 2016), Effects of parallelogram-shaped pavement markings on vehicle speed and safety of pedestrian crosswalks on urban roads in China, 2016). However, several researchers have also proposed using the Poisson-Lognormal (PLN) model as an alternative to the Poisson-Gamma model as it offered more flexibility in handling over-dispersion (e.g. Aguero-Valverde and Jovanis, 2008; Karim et al., 2013). Moreover, the FB Poisson-lognormal models were recommended over FB Poisson-gamma models when assuming vague priors and whenever crash data was characterized by low sample mean values (Lord and Miranda-Moreno, 2008).

The two models described above are considered global models, as the variables from these models are forced to have the same effect on all units or zones (Amoh-Gyimah et al., 2016). Such models may not be efficient in accounting for the unobserved heterogeneity. Therefore, other modeling techniques that incorporated spatial effects, or random parameters, or random intercepts have been recommended in the literature to account for the unobserved heterogeneity (El-Basyouny, 2011).

As for spatial effects crash models, various studies have discussed the CMs that incorporated spatial and temporal correlations (Zeng and Huang, 2014; Huang et al., 2016; Zeng et al., 2017a, b). Abdel-Aty and Wang (2006) were among the first to account for the temporal and spatial correlation among the data while modeling rear-end crash frequencies at signalized intersections. They applied generalized estimating equation (GEE) approach for analyzing the longitudinal data for 208 signalized intersections over 3 years and the spatially correlated data for 476 signalized intersections which were located along different corridors that were collected in the state of Florida. Also, Aguero-Valverde and Jovanis (2010) explored spatial correlation in multilevel collision frequency models for different types of urban and rural road segments. More recently, several studies attempted to incorporate spatial effects in several crash analysis applications (e.g. Karim et al., 2013; Prato et al., 2016). The previous studies showed that the models incorporating spatial effects performed better than those that did not since the spatial models can address the spatial heterogeneity among zones.

Alternatively, for random parameters crash models, Milton et al. (2008) were among the first to adopt the random parameters modeling approach in the traffic safety analysis. They used highway-injury data from Washington State to estimate a mixed (random parameters) logit model. The results indicated that volume-related variables such as average daily traffic per lane, average daily truck traffic, truck percentage, interchanges per mile and weather effects such as snowfall were best modeled as random-parameters, while roadway characteristics such as the number of horizontal curves, number of grade breaks per mile and pavement friction were best modeled as fixed parameters. Also, Ukkusuri et al. (2011) used data from New York City to study the factors that influence the frequency of pedestrian crashes. They used a random parameter, negative binomial model that was developed for predicting pedestrian crash frequencies at the census tract level. Since then, there has been a growing body of research that focused on random parameters approach (e.g. Anastasopoulos et al., 2012). The previous studies suggested that the random parameters models outperformed the fixed parameters models due to their ability to effectively account for the unobserved heterogeneity across individual observations or observation clusters.

More specifically, random intercepts model can be looked at as a simpler form of the random parameters model as the variation is allowed in the intercept only while the other regression coefficients are fixed. Huang et al. (2008) applied such model form when they developed a Bayesian hierarchical binomial logistic model to identify the significant factors affecting the severity level of driver injury and vehicle damage in traffic crashes at signalized intersections. Also, El-Basyouny and Sayed (2009a,b) applied the random intercepts model, by corridor, as they clustered 392 segments into 58 corridors while analyzing a dataset composed of urban arterials in Vancouver, British Columbia.

In general, few studies in the literature cross-compared the aforementioned modeling approaches on the macro-level. Xu and Huang (2015) utilized crash data for 738 TAZs in the county of Hillsborough, Florida to develop spatial effects and random parameters models. The analysis demonstrated that the conditional autoregressive (CAR) spatial effects model outperformed the random parameter negative binomial (RPNB) model for severe crashes. Moreover, they found out that the semi-parametric geographically weighted Poisson regression model, which was capable of accounting for the spatial correlation in crash data, performed best with the lowest mean absolute deviance and Akaike information criterion measures. Truong et al. (2016) also employed spatial effects model as well as random parameters model based on crash data for 63 provinces of Vietnam. The results showed that spatiotemporal model with conditional autoregressive prior model outperformed the random intercepts negative binomial (RINB) model and the RPNB model. They did not find significant difference between the performance of the RINB models and the RPNB models. Amoh-Gyimah et al. (2016) presented different estimation methods to model pedestrian and cyclist crashes. They compared the results of a RPNB models with the results of a non-spatial negative binomial (NB) model and a Poisson-Gamma CAR model. They found out that the RPNB model performed best with the lowest mean absolute deviation, mean squared predicted error and Akaiki information criterion measures.