Accident Analysis and Prevention 131 (2019) 122–130

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Accounting for mediation in cyclist-vehicle crash models: A Bayesian mediation analysis approach

Mohamed Bayoumi Kamel[a], Tarek Sayed[a,*], Ahmed Osama[b]

[a] University of British Columbia, Dept. of Civil Engineering, 6250 Applied Science Lane, Vancouver, V6T 1Z4, Canada
[b] Faculty of Engineering, Ain Shams University, 1 El Sarayat St., ABBASSEYA, Al Waili, Cairo, 11535, Egypt

ARTICLE INFO

Keywords:
Cyclist safety models
Bayesian mediation analysis
Full bayes
Bike kilometers travelled

ABSTRACT

Cyclist safety is affected by many factors on the zonal level. Previous studies have found associations between cyclist-vehicle crashes and vehicle and bike exposures, network configuration, land use, road facility, and the built environment. In addition, the network configuration, land use, and road facility were found to affect bike exposure levels. The association of zonal characteristics with both exposure and crashes may bias the development of macro-level bike safety models. This paper aims to explain these associations simultaneously using a form of Structural Equation Modelling approach. The analysis assesses the mediated effects that some variables have on crashes through their effects on bike exposure (by setting bike exposure as a mediator). Data from 134 traffic analysis zones (TAZ's) in the City of Vancouver, Canada is used as a case study. The indirect effect of network configuration, land use, and road facility on cyclist-vehicle crashes was assessed through Bayesian mediation analysis. Mediation analysis is an approach used to estimate how one variable transmits its effects to another variable through a certain mediator. These effects could be direct only, indirect only (through a certain mediator), or both direct and indirect. The results showed that the bike kilometers travelled (BKT) was a mediator of the relationship between network configuration, land use, and road facility and cyclist-vehicle crashes. The mediation analysis showed that some variables have different direct and indirect effect on cyclist-vehicle crashes. This indicates that while some variables may have negative direct association with crashes, their total crash effect can be positive after accounting for their effect through exposure. For example, bike network coverage and recreational density have negative direct association with cyclist-vehicle crashes, and positive indirect association leading to positive total effect on cyclist-vehicle crashes.

1. Introduction

Cycling, as an active mode of transportation, has the potential to reduce traffic congestion and emissions, as well as promote a healthier lifestyle. Therefore, transportation agencies in many European and North American cities are prioritizing the promotion of cycling. However, cyclists are vulnerable road users who are usually subjected to an elevated level of injury/fatality risk. This could discourage many cyclists from using the bike network and raise the need for developing safe and efficient bike networks that can attract more road users to cycle. As such, developing safety models that can explain the relationship of various factors on cyclist-vehicle crashes is of significant importance.

Several previous studies have developed macro-level safety models for use in the proactive safety evaluation of bike networks. These models investigate the relationship between cyclist-vehicle crashes and various network characteristics and exposure levels (Harris et al., 2011; Chen et al., 2012; Teschke et al., 2012; Kaplan and Giacomo Prato, 2015). Exposure is a crucial variable in these models and was shown to be one of the most important factors affecting crash occurrence. As well, several studies have also shown strong associations between cycling activity levels (exposure) and bike network characteristics (Nelson and Allen, 1997; Dill and Carr, 2003; Parkin et al., 2008; Buehler and Pucher, 2012). The association of bike network characteristics with both exposure and crashes can cause exposure to be a potential mediator in cyclist-vehicle crash models leading to biased results. Mediation occurs when: 1) a predictor of interest (e.g. network configuration) and a mediator (exposure level) are associated with some outcome (crashes); 2) the predictor of interest and the mediator are associated; 3) the mediator is assumed a causal consequence of the predictor of interest (the mediator is in the middle of the causal chain). Mediation analysis yields three different effects on the dependent variable: direct, indirect,

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

https://doi.org/10.1016/j.aap.2019.06.009
Received 19 January 2019; Received in revised form 14 May 2019; Accepted 12 June 2019
Available online 25 June 2019

and total effects (Baron and Kenny, 1986). The direct effect is the part of the independent variable effect on the dependent variable that is not mediated by a given mediator. The indirect effect is the part of the independent variable effect on the dependent variable that is mediated by a given mediator. The total effect is the aggregate effect of the independent variable on the dependent variable. The interpretation of these relationships is important in mediation studies (MacKinnon, 2012). Compared to conventional frequentist mediation analysis, the Bayesian mediation analysis approach was shown to have several benefits (Yuan and MacKinnon, 2009). First, it allows incorporation of prior information into the mediation analysis process. Second, Bayesian mediation analysis interpretation is straightforward. Third, the Bayesian approach is simpler for multilevel mediation analysis.

This paper applies Bayesian mediation analysis to assess the effects that some variables have on cyclist-vehicle crashes accounting for their effects on bike exposure (by setting bike exposure as a mediator). The variables include network configuration, road facility, built environment, and land use. The full Bayesian models are developed using data for 134 traffic analysis zones (TAZs) in the city of Vancouver, Canada.

2. Previous work

2.1. Network association with cycling levels

Considerable research has been undertaken to investigate the association between bike network characteristics and cycling levels. Nelson and Allen (1997), Dill and Carr (2003), and Buehler and Pucher (2012) investigated the association between bike network length and the number of bike commuters. Using a dataset of 50 cities, Dill and Carr (2003) found that a 1% increase of bike commuting is associated with each added square mile to the bike facility per square mile of the city area. Buehler and Pucher (2012) associated a 3.1% increase in the bike commuters with 10% increase in supply of bike lanes.

Winters et al. (2016) found correlation (r = 0.52) between developed bike score and bike commute trips. The used bike score is comprised of three environmental components: a bike lane score, a hill score, and a destinations and connectivity score. Xing et al. (2010) used data from an online survey conducted in 2006 in six small cities in the western US to investigate factors affecting cycling for transportation compared to cycling for recreation. The results indicated that individual, social-environment, and physical-environment factors have considerable effect on commute and recreation cycling trips. Schoner and Levinson (2014) used linear regression to measure the impact of bike network quality on bicycle commuting after controlling for demographic variables and the size of the city. They found that network connectivity and directness are essential factors in predicting bicycle commuting. Osama et al. (2017) used a Full Bayes (FB) approach that incorporates spatial effects to investigate the association between bike various network indicators and Bike Kilometers Traveled (BKT). They found that bike network indicators, land use, and road facility are significantly associated with BKT.

2.2. Cyclist safety models

Several studies have developed models for cyclist safety on the aggregate (traffic zone) level. Wei and Lovegrove (2013) developed macro-level crash models showing association between cyclist-vehicle crashes and the total network length, bike network length, number of bus stops, traffic signals, and intersection density, among others. Chen (2015) explored the association between built environment factors and cyclist safety. Chen (2015) also suggested that TAZ-based bicycle crashes are spatially correlated. Zhang et al. (2012) developed geographically weighted regression model at the TAZ level to study the impact of network connectivity on pedestrian and cyclist safety. They found that pedestrian and cyclist crashes is negatively associated with network connectivity. On the other hand, Siddiqui et al. (2012); Strauss et al. (2013); Wei and Lovegrove (2013), and Osama and Sayed (2016) found a positive association between intersection density (network connectivity metric) and cyclist-vehicle crashes. For land use, commercial land use was found to be positively associated with cyclist-vehicle crashes (Narayanamoorthy et al., 2013; Vandenbulcke et al., 2014) while Strauss et al. (2013) found that commercial land use was not a significant predictor of cyclist-vehicle crashes. Osama and Sayed conducted several studies to develop bike crash models using Full Bayes (FB) approach that incorporate spatial effects. They considered zonal characteristics, network configuration, and traffic exposure to develop macro-level crash models, also they incorporated bike kilometers travelled as a metric for the bike activity levels (Osama and Sayed, 2017a, b).

2.3. Mediation analysis

Mediation analysis has been extensively employed in psychological research (Baron and Kenny, 1986). It was also employed in a number of transportation studies. Sümer (2003) proposed mediated a model to distinguish the distal (i.e. personality factors) and proximal (i.e. aberrant driving behaviors) factors in predicting traffic crash involvement. Gargoum and El-Basyouny (2016) used paths analysis to model the relationship between road average speed and vehicle crash frequency. Zhang et al. (2018) investigated a traffic climate scale relation to motorists' personality and dangerous driving behavior using mediation analysis. Liu and Khattak (2017) developed a framework using path analysis to explore the contributing factors to gate-violation behavior at highway-rail grade crossings. Huang et al. (2018) used the large taxi floating car data to evaluate how traffic congestion-related negative moods influenced motorists' speed choice, while evaluating the indirect effect of traffic delay on the cruising speed adjustment using a mediation analysis approach.

There are a few studies that employed Bayesian mediation analysis, mostly in social science. For example, Milfont and Sibley (2016) tested a Bayesian path model to examine the extent to which empathy and social dominance orientation predicted environmental values, and calculated effects that mediated the gender difference. In medical research, Detilleux et al. (2016) proposed a Bayesian path analysis framework to evaluate the direct and indirect relationships between bone mineral density and vertebral fracture.

3. Data collection

3.1. Data sources

The models developed in this study are based on 134 Traffic Analysis Zones (TAZ's), at the city of Vancouver. The data used in this study was extracted from several sources:

1 The Insurance Corporation of British Columbia (ICBC), a public automobile insurance company, provided the crash data for a five years period (2009–2013). Only cyclist-vehicle crashes are included in the analysis of this study. A five years period is selected to collect an adequate sample size. The sample included three severity levels, i.e. fatality, injury, and property damage only.
2 TransLink, the Metro Vancouver transportation authority, provided 2013 geocoded files of the bike network, road network, land use, and TAZ boundaries. In addition, TransLink provided the output of Emme2 transportation-planning model for the travel demand in Metro Vancouver in the year 2011. The 2011 household travel survey was used to calibrate the Emme2 model, and the 2011 cordon counts were used to validate the model assignments.
3 Acuere Analytics provided the Vancouver Cycling Data Model (VCDM). The VCDM uses bike counts for seven years from 2005 to 2011 to estimate the annual average daily bike traffic (AADB) in 2011 over the City of Vancouver bike network (El Esawey et al.,

2015). The available data covered more than 810,000 hourly volumes over 7 years. The model was efficient in estimating the AADB traffic on most links of the bike network (more than 70% of the network).
4 The open data catalogue of the City of Vancouver, provided traffic signals, bus stops, and contour map of the city of Vancouver.

3.2. Analysis variables

The response variable in the mediator model is the Bike kilometers travelled (BKT), while cyclist-vehicle crash frequency is the dependent variable of the crash models. Crashes are aggregated at the different TAZs according to their locations. To account for cyclist-vehicle crash exposure, Bike Kilometers Travelled and Vehicle Kilometers Travelled are used representing the bike and vehicle exposures, respectively. Most of the previous studies that developed macro-level bike crash models used proxies for the cyclist exposure due to data limitations (bike network length, etc.). Recently, actual cyclist exposure such as the BKT (bike kilometers travelled) was used in developing the safety models (Osama and Sayed, 2016, 2017b).

The bike network coverage is calculated as the ratio between the number of bike links and the number street links at each TAZ (Osama and Sayed, 2016). The bike network coverage was used to describe bike network connectivity in previous studies (Yigitcanlar and Dur, 2010). To measure bike network linearity, a hypothetical length (modified bike network length) should be calculated that represent the length of the bike network if all the links were straight, while maintaining the nodes location the same. The bike network linearity was calculated as the ratio between the modified bike network length and the original bike network length in the TAZ (Osama and Sayed, 2016). Average edge length is calculated as the ratio between the total length of the zonal bike network and the number of links in the corresponding TAZ (Kansky, 1963). For topography, the bike network slope is calculated. The bike network slope of the zonal bike network is calculated as the ratio between the total weighted slope and the length of bike network at each TAZ. Arterial, collector, and local roads length aggregated at each TAZ level as well as the arterial-collector roads proportion to the road network was calculated. The built environment variables included the signal density and bus stops density at each TAZ. Lastly, the land use category included recreational density and commercial density at the TAZ level. Socio-demographic variables (population density, and number of households) were extracted from the EMME2 model for each zone (Table 1).

As discussed in the previous work section, some of the above-mentioned variables are correlated with bike exposure (Bike Kilometers Travelled) and cyclist-vehicle crashes. For example, the bike network coverage may be associated with increase in the bike exposure, as the presence of bike lanes may encourage road users to have more cycling trips. Similarly, recreational density may be positively associated with BKT, since several studies showed that recreational density would motivate road users to initiate more biking trips (Daley and Rissel, 2011). However, an increase in recreational density may be associated with a decrease in the frequency of cyclist-vehicle crashes, since recreational areas usually provide off-street and continuous paths for active transportation commuters reducing the conflict risk between these vulnerable commuters and vehicles.

4. Method

4.1. Bayesian Mediation analysis

Bayesian path analysis using BKT as a mediator is employed to assess the mediated effects that some variables (e.g. network configuration, land use, and road facility) have on cyclist-vehicle crashes. The path analysis has several advantages. First, multiple relationships can be tested simultaneously, where some of these relationships can be mediated. Second, the path analysis model estimates are easily interpreted. Third, it provides flexibility through defining the error term for each model individually to suit the response behavior. Mediation analysis is used to estimate how a variable transmits its effects to another variable through a certain mediator. These effects could be direct only, indirect only (through a certain mediator), or both direct and indirect. Fully mediated variables are variables with indirect effect only. There are three main approaches to conduct statistical mediation analysis: (a) causal steps, (b) difference in coefficients, and (c) product of coefficients (MacKinnon, 2012). The most widely used method to assess mediation is the causal steps approach that is defined in many previous studies (Baron and Kenny, 1986 and Judd and Kenny, 1981a, 1981b), which is used in this study. Compared to conventional frequentist mediation analysis, the Bayesian mediation analysis approach was shown to have several benefits (Yuan and MacKinnon, 2009). First, it allows incorporation of prior information into the mediation analysis process. Second, Bayesian mediation analysis interpretation is straightforward. Third, the Bayesian approach is simpler for multilevel mediation analysis. Eq. (1) represents the effects of x (e.g. Network configuration, land use, road facility) on m (BKT). Eq. (2) represents the effect of m and x on y (cyclist-vehicle crashes). The indirect effect (τ) of x on y can then be estimated by the product of coefficient estimator (Yuan and MacKinnon, 2009; Hayes, 2013), as shown in Eq. (3).

$$m_i = \delta_0 + \delta_1 x_i + u_i^{'} \tag{1}$$

$$y_i = \gamma_0 + \gamma_1 m_i + \gamma_2 x_i + u_i^{''} + s_i \tag{2}$$

$$\tau = \delta_1 \times \gamma_1 \tag{3}$$

Where $u_i^{'}$ and $u_i^{''}$ are the unstructured heterogeneity, while s_i is the structured heterogeneity. A path diagram illustrating the assumed relationship is presented in Fig. 1. The total effect is calculated by aggregating the indirect and the direct effect. This study does not include the impact of socio-economic and built environment variables on bike exposure as they did not have a statistically significant impact on bike exposure (bike kilometers travelled).

4.2. Model development

The full Bayesian analysis has the advantage of accounting for uncertainty and a more flexible structure that can be modified to suit the modeled process (El-Basyouny and Sayed, 2009; Sacchi and Sayed, 2015). More importantly, the full Bayes (FB) approach is more suitable for spatial modelling because of its ability to accommodate complex correlation structures (Aguero-Valverde and Jovanis, 2008).

The bike exposure (BKT) FB model was established with lognormally distributed random error. The BKT distribution is right skewed and therefore, the lognormal distribution error showed a good fit. A simple technique to represent a lognormally distributed model is to apply natural logarithm on the BKT. In this way the error term $u_i^{'}$, can be described as shown in Eq. (4). The model final form is shown in Eq. (5).

$$u_i^{'} \sim Normal\ (0, \sigma_{u'}^2) \tag{4}$$

$$\ln(BKT_i) = b_0 + \Sigma_m b_m X_{mi} + u_i^{'} \tag{5}$$

Where, $\sigma_{u'}^2$ is the dispersion parameter, b_0 is the intercept, m is the number of variables, X_{mi} is the considered covariates, and b_m is the model parameters. Network configuration, land use, and road facility variables were tested in a forward stepwise manner. Since, some of the variables were highly correlated, to avoid collinearity, highly correlated variables were not included in the same model.

For the crash models, Poisson lognormal models that account for spatial effects are employed to handle the over-dispersion in count data, and to account for both the unstructured and structured (spatially correlated) heterogeneities. Spatial correlation might exist since

Table 1
Presents the set of variables included in the analysis. These variables were used in several previous studies (e.g., Osama and Sayed, 2016, 2017b).

Data summary statistics				
Variable (Description)	Mean	Standard Deviation	Minimum	Maximum
Crashes				
Cyclist-vehicle crashes	12.72	13.49	0.00	78.00
Exposure				
Vehicle Kilometer Travelled (VKT) in thousands of kilometers	4.29	3.33	0.19	22.29
Bike Kilometer Travelled (BKT) in thousands of kilometers	1.05	2.11	0.00	21.46
Network Configuration				
Bike network coverage	0.34	0.19	0.00	1.01
Bike network linearity	0.68	0.27	0.00	1.00
Average edge length	0.13	0.05	0.00	0.57
Bike network slope	2.53	0.90	0.64	6.66
Road Facility				
Arterial-collector roads proportion (of all road links in the TAZ, by length)	0.35	0.21	0.12	1.00
Total road length (Arterial, Collector, and Local Roads Length aggregation)	11,382.56	7912.09	954.34	38484.37
Built Environment				
Bus stops density (number of bus stops/zone area in km^2)	24.28	23.62	0.00	162.24
Signal density (number of signals/zone area in km^2)	0.92	1.13	0.00	7.16
Land use				
Recreational density (recreational areas/zone area)	0.10	0.13	0.00	0.91
Commercial density (commercial areas/zone area)	0.08	0.11	0.00	0.58
Socio-Economic				
Population density (population/zone area in km2)	8391.82	6995.85	0.00	33658.90
Number of households	2058.95	1217.41	0.00	6163.00

neighboring zones typically have similar environmental and geographic characteristics and thereby form a cluster that has similar crash occurrence (Mountain et al., 1998; Shankar et al., 1998). This form of extra variation (structured heterogeneities) implies that there exists spatial autocorrelation between spatial units (Traffic Analysis Zones). The development of FB models in this study followed the procedure described by El-Basyouny and Sayed (2009). Y_i is assumed to be the number of cyclist-vehicle crashes at zones, and it is assumed to follow a Poisson distribution with a parameter λ_i as shown in Eq. (6). λ_i is considered itself a random variable and modeled according to Eq. (7).

$$Y_i \tilde{}\ Poisson\ (\lambda_i) \tag{6}$$

$$\ln(\lambda_i) = a_o + a_1 \ln(VKT_i) + a_2 \ln(BKT_i) + \Sigma_m d_m X_{mi} + u_i^{''} + s_i \tag{7}$$

Where a_o, a_1, a_2, and d_m are model parameters, VKT_i is the vehicle exposure variable, BKT_i is the bike exposure variable, X_{mi} represents the covariates, $u_i^{''}$ accounts for the unstructured heterogeneity among the zones, and s_i accounts for the spatially correlated heterogeneity among the zones. The followed statistical methodology to add explanatory variables into a Crash model is a forward stepwise procedure, after including variables representing exposure. The unstructured heterogeneity follows lognormal distribution, as implied from Eq. (7), and Eq. (8). The spatial effect is structured by Gaussian Conditional Autoregressive Regressive (CAR) techniques and calculated by Eq. (9).

$$u_i^{''} \tilde{}\ Normal\ (0, \sigma_{u''}^2) \tag{8}$$

$$S_i | S_{-i} \tilde{}\ Normal \left(\bar{s}_i, \frac{\sigma_s^2}{n_i} \right),\ where\ \bar{s}_i = \sum_{j \epsilon C(i)} \frac{s_j}{n_i} \tag{9}$$

Where $\sigma_{u''}^2$ is the unstructured heterogeneity variation, σ_s^2 is the spatial variation, n_i is the number of neighbors of zone i, $C(i)$ is the set of

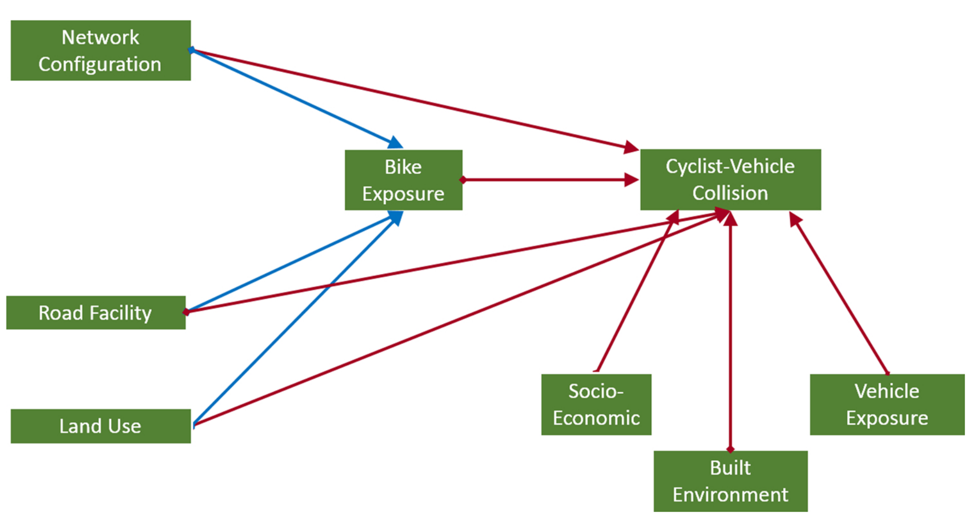

Fig. 1. Proposed Path Model.

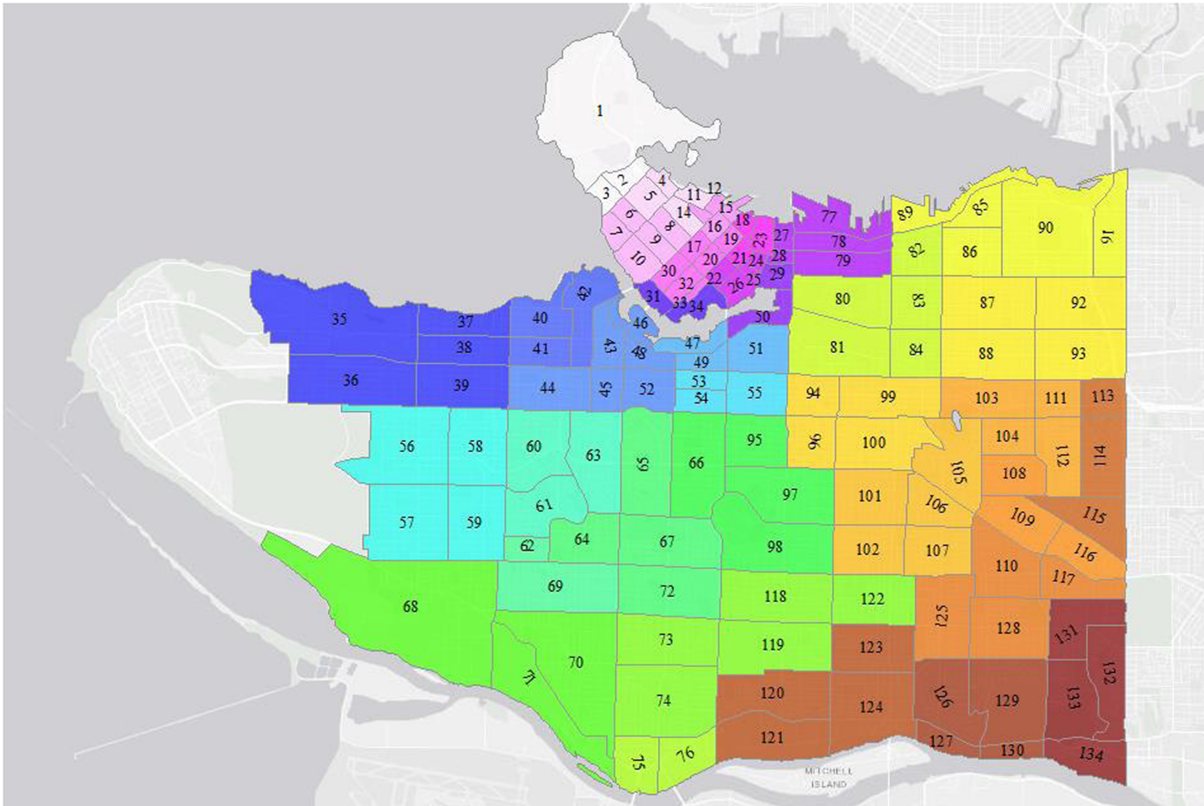

Fig. 2. City of Vancouver Spatial Structure.

neighbors of zone i, S_i accounts for the spatially correlated (structured) heterogeneity among zones, and S_{-i} is the set of all spatial effects except S_i. The spatial component S_i suggests that zones that are closer to each other are likely to have common features affecting their crash occurrence. Fig. 2 shows the spatial structure considered in this study, where 134 traffic analysis zones are allocated to 35 neighborhoods. Each color represents a neighborhood, for example zones 1, 2, and 3 are in a neighborhood. Spatial correlation is considered significant if the spatial variation is found to be greater than 0.5 (Aguero-Valverde and Jovanis, 2008). The spatial variation is assessed according to Eq. (10). Inclusion of spatial effect term usually enhance the model goodness of fit, however, sometimes it affects parameters level of significance. Although the variables might be significant before adding the spatial term, some parameters may lose their significance after the addition of the spatial term (Karim et al., 2013).

$$\psi_s = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_u^2} \tag{10}$$

Markov chain Monte Carlo (MCMC) is applied using WinBUGS tool to sample the posterior distribution and estimate the FB model parameters. MCMC methods use sampling to generate sequences of random points, the distribution of which converges to the target posterior distributions. A subsample is used for monitoring convergence, and then excluded as a burn-in sample. Parameter estimation, performance evaluation, and inference are obtained by the following iterations. Obtaining FB estimates requires the specification of prior distributions for the parameters reflecting the prior knowledge about the considered parameters. The prior is either informative or non-informative (vague), which depends on the availability of prior information. The most commonly used prior is a diffuse prior that has normal distribution with a zero mean and a large variance, which is considered vague prior (El-Basyouny and Sayed, 2009). For the dispersion parameter, σ_u^2, and σ_u^2, gamma distribution is usually used as a prior with parameters (ε, ε), where ε is a small number, e.g. 0.001 (Karim et al., 2013). Furthermore, for the spatial models developed in this study, the σ_s^2 prior distribution is assumed to follow gamma distribution with parameters of $(1 + \Sigma l_i/2,\ 1 + n/2)$, where l_i represent each zone contribution and is calculated as in Eq. (11).

$$l_i = n_i s_i (s_i - \bar{s}_i) \tag{11}$$

WinBUGS software was used to sample the posterior distributions and estimate the model parameters. A subsample with 20,000 iterations is used to avoid out of order start up points, and then excluded as a burn in sample. Two chains were used with different initial points, to be able to check convergence. Convergence was checked by three methods. First, the model is considered converged if Brooks-Gelman-Rubin statistics is less than 1.1 (Brooks and Gelman, 1998). Second, convergence was checked by calculating the ratio between Monte Carlo errors and the respective standard deviations. For each estimate, as a rule of thumb, convergence occurs when the ratio is less than 0.05. Finally, convergence was checked visually by inspecting the MCMC trace plots of the model estimates. After reaching convergence, 20,000 iterations were performed for the two chains to estimate the model parameters. The Deviance Information Criteria (DIC) was used to judge the FB model goodness of fit (Spiegelhalter et al., 2003). According to Spiegelhalter et al. (2003), DIC differences between 5 and 10 are considered substantial, and differences higher than 10 might definitely rule out the model with the higher DIC.

5. Results

5.1. Bike kilometers travelled model

Table 2 presents the results of the bike exposure models. Models 1 and 2 are developed to include several variables from the different investigated categories. Model 1 and 2 show that bike network slope is negatively associated with BKT since steep upgrade slopes may work as

Table 2
Bike kilometers travelled FB models.

Variable	Model 1				Model 2			
	Estimate	SD	Credible Interval		Estimate	SD	Credible Interval	
			2.50%	97.50%			2.50%	97.50%
Intercept	−0.503	0.407	−1.297	0.312	−0.831	0.37	−1.572	−0.129
Bike network coverage	1.92	0.635	0.689	3.164				
Bike network slope	−0.202*	0.115	−0.43	0.022	−0.281	0.108	−0.492	−0.067
Arterial-collector roads proportion	−0.945*	0.556	−2.034	0.139				
Total road length					0.042	0.015	0.013	0.071
Recreational density					2.063	0.746	0.601	3.523
Commercial density					1.831*	1.009	−0.164	3.809
DIC	235.04				231.66			

* Significantly different from zero at 10%, all other variables were significantly different from zero at 5%.

a deterrent to cyclists. This is consistent with previous studies (Dill and Carr, 2003; Winters et al., 2016). Model 1 shows that bike network coverage is positively associated with BKT, which is intuitive and consistent with a previous study by Schoner and Levinson (2014), in which they showed that connectivity is positively associated with bike commuting rates. Model 1 shows that arterial-collector roads proportion is found to have a negative association with BKT. This may be due to that arterial and collector roads are usually perceived by cyclists as less safe and less friendly (Marshall and Garrick, 2011). Model 2 shows that recreational density is positively associated with BKT. This result is plausible, as several studies showed that recreational density would motivate commuters to initiate more cycling trips (e.g., Daley and Rissel, 2011). Model 2 shows that commercial density is positively associated with the BKT level. This is likely due to commercial areas attracting more cycling trips than residential areas, which is in agreement with results from Daley and Rissel (2011). Model 2 shows that total road length at each zone was found positively associated with BKT, which is intuitive as it represents cyclists' using the available road network.

5.2. Impact on cyclist-vehicle crashes

Tables 3 and 4 show four crash models (models A, B, C, and D) that were developed to capture the direct effect on cyclist-vehicle crash frequency, where all the variables are statistically significant at the 95% level unless otherwise mentioned. Spatial effects are significant in all developed models (i.e. $\psi_s \gg 0.5$). This highlights the importance of accounting for spatial autocorrelation in the macro-level crash models. Both bike and vehicle exposures were found to be positively associated with cyclist-vehicle crashes. The results are intuitive and consistent with previous research (Miranda-Moreno et al., 2011; Strauss et al., 2013; Hamann and Peek-Asa, 2013; Kaplan and Giacomo Prato, 2015). The exposure estimate is less than one, which suggests a decrease in cyclist-vehicle crash risk with increasing exposure and is consistent with safety in numbers hypothesis (Jacobsen, 2003).

Table 5 shows the indirect effect on the cyclist-vehicle crashes. None of the investigated variables was fully mediated through the BKT. A summary of direct, indirect, and total effect on cyclist-vehicle crashes is presented in Table 6. The Bike network slope Indirect effect is calculated from model 2 as it has higher statistical significance compared to the same estimate from model 1.

Model D shows that the bike network slope has direct negative association with cyclist-vehicle crashes. This association is consistent with the previous research done by Chen (2015). This may be attributed to cyclists reducing their speed at upgrades, which would lower cyclist-vehicle crash risk. In addition, the bike network slope has indirect and total negative association with cyclist-vehicle crashes. This implies that an increase in the bike network slope is associated with decrease in bike ridership and decrease in cyclist-vehicle crashes directly and indirectly. Although the bike network coverage has negative direct association with cyclist-vehicle crashes (as shown in Model B), its indirect positive association outweighs its direct effect resulting in a positive total effect. This might explain the difference in studies investigating the association of network connectivity with cyclist-vehicle crashes (Siddiqui et al., 2012; Strauss et al., 2013; Wei and Lovegrove, 2013; Zhang et al., 2012). The network connectivity association sign may depend on the quality of the bike exposure in the developed model. To illustrate, including reliable bike exposure may result negative association between network connectivity and cyclist-vehicle crashes, as the bike exposure would be able to mediate the positive effect on crashes. The change in the direction of the bike network coverage impact on cyclist-vehicle crashes can be explained as the increase in bike network coverage is

Table 3
Crash models A &B with spatial effects.

Variable	Model A				Model B			
	Estimate	SD	Credible Interval		Estimate	SD	Credible Interval	
			2.50%	97.50%			2.50%	97.50%
Intercept	1.896	0.318	1.270	2.512	2.454	0.319	1.822	3.087
VKT	0.347	0.098	0.167	0.552	0.340	0.100	0.136	0.533
BKT	0.491	0.061	0.377	0.612	0.513	0.066	0.383	0.642
Bike network coverage					−0.811*	0.420	−1.642	0.032
Bike network linearity	0.395*	0.190	−0.015	0.802				
Average edge length	−2.108*	1.042	−4.034	0.083	−3.613	1.434	−6.478	−0.7737
Arterial-collector roads proportion					0.905	0.446	0.020	1.805
Signal density	0.244	0.066	0.103	0.373				
Recreational density	−1.010*	0.501	−2.150	0.008				
DIC	736.75				743.63			
ψs	0.896	0.078	0.705	0.972	0.850	0.081	0.662	0.974

* Significantly different from zero at 10%, all other variables were significantly different from zero at 5%.

Table 4
Crash models C &D with spatial effects.

Variable	Model C				Model D			
	Estimate	SD	Credible Interval		Estimate	SD	Credible Interval	
			2.50%	97.50%			2.50%	97.50%
Intercept	1.251	0.245	0.767	1.739	1.358	1.378	0.932	2.004
VKT	0.275	0.113	0.059	0.502	0.344	0.353	0.117	0.518
BKT	0.500	0.063	0.379	0.625	0.504	0.164	0.369	0.636
Bike network slope					−0.053*	0.085	−0.201	0.104
Total road length	0.028	0.013	0.001	0.054				
Bus stop density					0.014	0.028	0.004	0.018
Commercial density	2.771	0.828	1.118	4.312				
Population density	$3.4^{*}10^{-5}$	$1.2^{*}10^{-5}$	$1.1^{*}10^{-5}$	$5.9^{*}10^{-5}$				
Number of households					$2.1^{*}10^{-4}$	$0.8\text{-}10^{-4}$	$1.1^{*}10^{-4}$	$3.3^{*}10^{-4}$
DIC	741.70				739.97			
ψs	0.855	0.086	0.657	0.971	0.882	0.081	0.687	0.972

* Significantly different from zero at 10%, all other variables were significantly different from zero at 5%.

Table 5
Indirect Effects on Cyclist-vehicle crashes through BKT.

Variable	Estimate	SD	Credible Interval	
			2.50%	97.50%
Bike network coverage	0.985	0.355	0.338	1.729
Arterial-collector roads proportion	−0.485*	0.296	−1.092	0.071
Bike network slope	−0.142	0.057	−0.255	−0.029
Recreational density	1.012	0.397	0.292	1.853
Commercial density	0.916*	0.522	−0.064	1.994
Total road length	0.021	0.008	0.006	0.037

* Significantly different from zero at 10%, all other variables were significantly different from zero at 5%.

Table 6
A summary of direct, indirect, and total effect on cyclist-vehicle crashes.

Variable	Direct	Indirect	Total effect
Bike network coverage	−0.811	0.985	0.174
Arterial-collector roads proportion	0.905	−0.485	0.420
Bike network slope	−0.053	−0.142	−0.195
Recreational density	−1.010	1.012	0.002
Commercial density	2.771	0.916	3.687
Total road length	0.028	0.021	0.049
Bike network linearity	0.395*	0.000	0.395*
Average edge length	−2.108*	0.000	−2.108*
Signal density	0.244	0.000	0.244
Total road length	0.028	0.000	0.028
Bus stop density	0.014	0.000	0.014
Population density	$3.4^{*}10^{-5}$	0.000	$3.4^{*}10^{-5}$
Number of households	$2.1^{*}10^{-4}$	0.000	$2.1^{*}10^{-4}$

* Significantly different from zero at 10%, all other variables were significantly different from zero at 5%.

associated with an increase in BKT and a decrease in the cyclist-vehicle crashes. On the other hand, the increase of the BKT is associated with an increase in the cyclist-vehicle crashes, and therefore, the increase in the bike network coverage indirectly increases the cyclist-vehicle crashes. The bike network coverage total effects reveal that the bike network coverage impact on cyclist-vehicle crashes through the BKT (indirect effect) outweighs the bike network coverage impact on cyclist-vehicle crashes (direct effect).

The bike network coverage direct impact on cyclist-vehicle crash may be interpreted as the impact of the bike network coverage on cyclist-vehicle crash risk. Conversely, the bike network coverage total effect may be interpreted as the total impact of the bike network coverage on cyclist-vehicle crash frequency. Therefore, planners may use the direct effect to measure the bike network coverage impact on cyclist-vehicle crash risk to quantify the risk of each cyclist. Additionally, planners may use the total effect to measure the bike network coverage impact on the total number of cyclist-vehicle crashes at each zone.

Model A shows that the increase in recreational density is directly associated with a decrease in the frequency of cyclist-vehicle crashes. However, recreational density is indirectly positively associated with cyclist-vehicle crashes through BKT. To illustrate, recreational density encourages road users to cycle (that increase the BKT), simultaneously the increase in recreational density provides off-street and continuous paths (that decrease cyclist-vehicle crashes). Recreational density total effect on cyclist-vehicle crashes was very small (0.002), which indicates that the recreational density indirect effect (through BKT) nullifies its direct effect on crashes.

As shown in Model B, Arterial-collector roads proportion was found to be directly positively associated with cyclist-vehicle crashes. This is intuitive as arterial and collector roads usually have a higher speed, and have more diverse road users, which would increase the cyclist-vehicle crash risk. These results agree with previous studies conducted by Chen (2015) and Siddiqui et al. (2012). Although, it has a negative indirect effect on cyclist-vehicle crashes, its total effect on cyclist vehicle crashes is positive. The commercial density showed positive association directly (as shown in Model C) and indirectly with cyclist-vehicle crashes. The association between the commercial density and cyclist-vehicle crashes is consistent with previous studies (Narayanamoorthy et al., 2013 and Vandenbulcke et al., 2014). This is likely attributed to the side street commercial activities raise the potential risk of a cyclist going into conflicts with motorized traffic.

The total road length has a positive direct and indirect association with cyclist-vehicle crashes. Which is attributed to the fact that road network length can be considered surrogate measure for traffic exposure. The signal density is positively associated with cyclist-vehicle crashes, as shown in Model A. More traffic signal implies higher likelihood of conflicts between different road users specifically between cyclists and motorists. Similarly, the presence of bus stops indicates more interaction between ground transit and cyclists. Model C reveal positive associations between cyclist-vehicle crashes and the population density as well as the number of households. The results are reasonable since these variables can be considered surrogate measures for traffic exposure, thereby explaining their positive associations with cyclist crashes, which is consistent with previous studies by Siddiqui et al. (2012) and Prato et al. (2016). The bike network linearity is positively associated with cyclist-vehicle crashes as cyclists and motorists tend to accelerate on the straight links, which would increase crash risk. Model A shows that the average edge length is negatively associated with cyclist-vehicle crashes, which agrees with a safety study done by Quintero et al. (2013) on Metro Vancouver transit network.

5.3. Summary and conclusions

Bayesian mediation analysis is employed to assess the mediated effects that some variables (network configuration, land use, and road facility) have on crash through their effects on bike exposure (by setting BKT as a mediator). Mediation occurs when: 1) a predictor of interest (e.g. network configuration) and a mediator (exposure level) are associated with some outcome (crashes); 2) the predictor of interest and the mediator are associated; 3) the mediator is assumed a causal consequence of the predictor of interest (the mediator is in the middle of the causal chain). The mediation effects could be direct, indirect (through a certain mediator), or both direct and indirect.

Bike kilometers travelled full Bayesian models showed that bike network slope and arterial-collector roads proportion were negatively associated with BKT, while the bike network coverage, road network length, recreational density, and commercial density were positively associated with the BKT level. Crash models were developed using full Bayesian techniques incorporating spatial effects. The crash models showed that all the variables in BKT models have significant direct effect on cyclist –vehicle crashes. The results showed that cyclist-vehicle crashes is negatively associated with average edge length, bike network coverage, the bike network slope, and recreational density, while signal density, bus stops density, population density, the number of households, bike network linearity, arterial-collector roads proportion, road network length, and commercial density have positive association with cyclist-vehicle crashes.

The importance of this study is that it differentiates between the direct and indirect effects (through exposure) of zonal characteristics on cyclist-vehicle crashes. The mediation analysis showed that some variables have different direct and indirect effects on cyclist-vehicle crashes. This indicates that while some variables may have negative direct association with crashes, their total effect on crashes can be positive after accounting for their effect through exposure. For example, the bike network coverage and recreational density have negative direct association with cyclist-vehicle crashes, and positive indirect association leading to positive total effect on cyclist-vehicle crashes. Furthermore, the Bayesian mediation results showed that the bike network slope has direct and indirect negative associations with cyclist-vehicle crashes. The commercial density and total road length showed positive direct and indirect associations with cyclist-vehicle crashes. Arterial-collector roads proportion was found to be directly positively associated with cyclist-vehicle crashes. Although, it has a negative indirect effect on cyclist-vehicle crashes through BKT, its total effect is positively associated with the cyclist-vehicle crashes.

This study has several limitations. Out-of-sample data is needed to validate the developed models. In addition, data from other North American or European cities could be used to test the developed model transferability and to validate the conclusions of this paper. Additional variables can be integrated in the bike kilometers travelled model to investigate further associations with BKT, including built environment variables (e.g. light poles), socio-demographic variables (e.g. employment density), additional facility characteristics (e.g. on-street parking, surface type, etc.), and additional bike network indicators (e.g. centrality, assortativity, etc.). As, this study did not consider potential mediating effects through vehicle exposure, future research should account for the mediating effects of vehicle exposure and bike exposure by simultaneously considering vehicle kilometers travelled and bike kilometers travelled as mediators. Finally, it should be noted that this study included only cyclist-vehicle interaction and ignored cyclist interaction with other road users, due to data limitation.

References

Aguero-Valverde, J., Jovanis, P.P., 2008. Analysis of road crash frequency with spatial models. Transp. Res. Rec. 2061 (1), 55–63.

Baron, R.M., Kenny, D.A., 1986. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J. Pers. Soc. Psychol. 51 (6), 1173.

Brooks, S., Gelman, A., 1998. Alternative methods for monitoring convergence of iterative simulations. J. Comput. Graph. Stat. 7, 434455.

Buehler, R., Pucher, J., 2012. Cycling to work in 90 large American cities: new evidence on the role of bike paths and lanes. Transportation 39 (2), 409–432.

Chen, P., 2015. Built environment factors in explaining the automobile-involved bicycle crash frequencies: a spatial statistic approach. Saf. Sci. 79, 336–343.

Chen, L., Chen, C., Srinivasan, R., McKnight, C.E., Ewing, R., Roe, M., 2012. Evaluating the safety effects of bicycle lanes in New York City. Am. J. Public Health 102 (6), 1120–1127.

Daley, M., Rissel, C., 2011. Perspectives and images of cycling as a barrier or facilitator of cycling. Transp. Policy 18 (1), 211–216.

Detilleux, J., Reginster, J.Y., Chines, A., Bruyère, O., 2016. A Bayesian path analysis to estimate causal effects of bazedoxifene acetate on incidence of vertebral fractures, either directly or through non-linear changes in bone mass density. Stat. Methods Med. Res. 25 (1), 400–412.

Dill, J., Carr, T., 2003. Bicycle commuting and facilities in major US cities: if you build them, commuters will use them. Transp. Res. Rec. 1828, 116–123.

El-Basyouny, K., Sayed, T., 2009. Urban arterial accident prediction models with spatial effects. Transp. Res. Rec. 2102, 27–33.

El Esawey, M., Lim, C., Sayed, T., 2015. Development of a cycling data model: city of Vancouver case study. Can. J. Civ. Eng. 42 (12), 1000–1010.

Gargoum, S.A., El-Basyouny, K., 2016. Exploring the association between speed and safety: a path analysis approach. Accid. Anal. Prev. 93, 32–40.

Hamann, C., Peek-Asa, C., 2013. On-road bicycle facilities and bicycle crashes in Iowa, 2007–2010. Accid. Anal. Prev. 56, 103–109.

Harris, M.A., Reynolds, C.C., Winters, M., Chipman, M., Cripton, P.A., Cusimano, M.D., Teschke, K., 2011. The Bicyclists' Injuries and the Cycling Environment study: a protocol to tackle methodological issues facing studies of bicycling safety. Inj. Prev injuryprev-2011.

Hayes, A.F., 2013. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-based Approach. Guilford Press.

Huang, Y., Sun, D.J., Zhang, L.H., 2018. Effects of congestion on drivers' speed choice: assessing the mediating role of state aggressiveness based on taxi floating car data. Accid. Anal. Prev. 117, 318–327.

Jacobsen, P.L., 2003. Safety in numbers: more walkers and bicyclists, safer walking and bicycling. Inj. Prev. 9 (3), 205–209.

Judd, C.M., Kenny, D.A., 1981a. Estimating the Effects of Social Intervention. CUP Archive.

Judd, C.M., Kenny, D.A., 1981b. Process analysis: estimating mediation in treatment evaluations. Eval. Rev. 5 (5), 602–619.

Kansky, K.J., 1963. Structure of Transportation Networks: Relationships between Network Geometry and Regional Characteristics.

Kaplan, S., Giacomo Prato, C., 2015. A spatial analysis of land use and network effects on frequency and severity of cyclist–motorist crashes in the Copenhagen region. Traffic Inj. Prev. 16 (7), 724–731.

Karim, M., Wahba, M., Sayed, T., 2013. Spatial effects on zone-level collision prediction models. Transp. Res. Rec. 2398, 50–59.

Liu, J., Khattak, A.J., 2017. Gate-violation behavior at highway-rail grade crossings and the consequences: using geo-spatial modeling integrated with path analysis. Accid. Anal. Prev. 109, 99–112.

MacKinnon, D., 2012. Introduction to Statistical Mediation Analysis. Routledge.

Marshall, W.E., Garrick, N.W., 2011. Evidence on why bike-friendly cities are safer for all road users. Environ. Pract. 13 (1), 16–27.

Milfont, T.L., Sibley, C.G., 2016. Empathic and social dominance orientations help explain gender differences in environmentalism: a one-year Bayesian mediation analysis. Pers. Individ. Dif. 90, 85–88.

Miranda-Moreno, L.F., Strauss, J., Morency, P., 2011. Disaggregate exposure measures and injury frequency models of cyclist safety at signalized intersections. Transp. Res. Rec. 2236 (1), 74–82.

Mountain, L., Maher, M., Fawaz, B., 1998. The influence of trend on estimates of accident at junctions. Accid. Anal. Prev. 30 (5).

Narayanamoorthy, S., Paleti, R., Bhat, C.R., 2013. On accommodating spatial dependence in bicycle and pedestrian injury counts by severity level. Transp. Res. Part B Methodol. 55, 245–264.

Nelson, A., Allen, D., 1997. If you build them, commuters will use them: association between bicycle facilities and bicycle commuting. Transp. Res. Rec. 1578, 79–83.

Osama, A., Sayed, T., 2016. Evaluating the impact of bike network indicators on cyclist safety using macro-level crash prediction models. Accid. Anal. Prev. 97, 28–37.

Osama, A., Sayed, T., 2017a. Evaluating the impact of socioeconomics, land use, built environment, and road facility on cyclist safety. Transp. Res. Rec. 2659, 33–42.

Osama, A., Sayed, T., 2017b. Investigating the effect of spatial and mode correlations on active transportation safety modeling. Anal. Methods Accid. Res. 16, 60–74.

Osama, A., Sayed, T., Bigazzi, A.Y., 2017. Models for estimating zone-level bike kilometers traveled using bike network, land use, and road facility variables. Transp. Res. Part A Policy Pract. 96, 14–28.

Parkin, J., Wardman, M., Page, M., 2008. Estimation of the determinants of bicycle mode share for the journey to work using census data. Transportation 35 (1), 93–109.

Prato, C.G., Kaplan, S., Rasmussen, T.K., Hels, T., 2016. Infrastructure and spatial effects on the frequency of cyclist-motorist collisions in the Copenhagen Region. J. Transp. Saf. Secur. 8 (4), 346–360.

Quintero, L., Sayed, T., Wahba, M.M., 2013. Safety models incorporating graph theory based transit indicators. Accid. Anal. Prev. 50, 635–644.

Sacchi, E., Sayed, T., 2015. Investigating the accuracy of Bayesian techniques for before–after safety studies: The case of a "no treatment" evaluation" Accid. Anal. Prev.

78, 138–145.
Schoner, J.E., Levinson, D.M., 2014. The missing link: bicycle infrastructure networks and ridership in 74 US cities. Transportation 41 (6), 1187–1204.
Shankar, V.N., Albin, R.B., Milton, J.C., Mannering, F.L., 1998. Evaluating median crossover likelihoods with clustered accident counts: an empirical inquiry using the random effects negative binomial model. Transp. Res. Rec. 1635, 44–48.
Siddiqui, C., Abdel-Aty, M., Choi, K., 2012. Macroscopic spatial analysis of pedestrian and bicycle crashes. Accid. Anal. Prev. 45, 382–391.
Spiegelhalter, David, et al., 2003. WinBUGS User Manual.
Strauss, J., Miranda-Moreno, L.F., Morency, P., 2013. Cyclist activity and injury risk analysis at signalized intersections: a Bayesian modelling approach. Accid. Anal. Prev. 59, 9–17.
Sümer, N., 2003. Personality and behavioral predictors of traffic accidents: testing a contextual mediated model. Accid. Anal. Prev. 35 (6), 949–964.
Teschke, K., Harris, M.A., Reynolds, C.C., Winters, M., Babul, S., Chipman, M., et al., 2012. Route infrastructure and the risk of injuries to bicyclists: a case-crossover study. Am. J. Public Health 102 (12), 2336–2343.
Vandenbulcke, G., Thomas, I., Int Panis, L., 2014. Predicting cycling accident risk in Brussels: a spatial case-control approach. Accid. Anal. Prev.
Wei, F., Lovegrove, G., 2013. An empirical tool to evaluate the safety of cyclists: community based, macro-level crash prediction models using negative binomial regression. Accid. Anal. Prev. 61, 129–137.
Winters, M., Teschke, K., Brauer, M., Fuller, D., 2016. Bike Score®: associations between urban bikeability and cycling behavior in 24 cities. Int. J. Behav. Nutr. Phys. Act. 13 (1), 18.
Xing, Y., Handy, S.L., Mokhtarian, P.L., 2010. Factors associated with proportions and miles of bicycling for transportation and recreation in six small US cities. Transp. Res. D Transp. Environ. 15 (2), 73–81.
Yuan, Y., MacKinnon, D.P., 2009. Bayesian mediation analysis. Psychol. Methods 14 (4), 301.
Yigitcanlar, T., Dur, F., 2010. Developing a sustainability assessment model: the sustainable infrastructure, land-use, environment and transport model. Sustainability 2 (1), 321–340.
Zhang, Yuanyuan, Bigham, John, Li, Zhibin, Ragland, David, Chen, Xiaohong, 2012. Associations between Road network connectivity and pedestrian-bicyclist accidents. 91st Annual Meeting of the Transportation Research Board.
Zhang, Q., Ge, Y., Qu, W., Zhang, K., Sun, X., 2018. The traffic climate in China: the mediating effect of traffic safety climate between personality and dangerous driving behavior. Accid. Anal. Prev. 113, 213–223.