Accident Analysis and Prevention 165 (2022) 106518

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Uncertainty matters: Bayesian modeling of bicycle crashes with incomplete exposure data☆

Pengpeng Xu [a,b], Lu Bai [b], Xin Pei [c], S.C. Wong [b,d], Hanchu Zhou [e,f,*]

[a] School of Civil Engineering and Transportation, South China University of Technology, Guangzhou, China
[b] Department of Civil Engineering, The University of Hong Kong, Hong Kong, China
[c] Department of Automation, Tsinghua University, Beijing, China
[d] Guangdong – Hong Kong – Macau Joint Laboratory for Smart Cities, Hong Kong, China
[e] School of Traffic and Transportation Engineering, Central South University, Changsha, Hunan, China
[f] School of Data Science, City University of Hong Kong, Hong Kong, China

ARTICLE INFO

Keywords:
Bicycle crashes
Incomplete exposure
Simultaneous equations
Bayesian imputation
Spatial correlation
Cross validation

ABSTRACT

Background: One major challenge faced by neighborhood-level bicycle safety analysis is the lack of complete and reliable exposure data for the entire area under investigation. Although the conventional travel-diary surveys, together with the emerging smartphone fitness applications and bike-sharing systems, provide straightforward and valuable opportunities to estimate territory-wide bicycle activities, the obtained ridership suffers inherently from underreporting.

Methods: We introduced the Bayesian simultaneous-equation model as a sound methodological alternative here to address the uncertainty arising from incomplete exposure data when modeling bicycle crashes. The proposed method was successfully fitted to a crowdsourced dataset of 792 bicycle–motor vehicle (BMV) crashes aggregated from 209 neighborhoods over a 3-year period in Hong Kong.

Results: Our analysis empirically demonstrated the bias due to omission of activity-based exposure measures or to the direct use of cycling distance extracted from the travel-diary survey without correcting for incompleteness. By modeling bicycle activities and the frequency of BMV crashes simultaneously, we also provided new evidence that an expansion of bicycle infrastructure was likely associated with a significant increase in cycling levels and a substantial reduction in the risk of BMV crashes, despite a slight increase in the absolute number of BMV crashes.

Conclusions: Our approach is promising in adjusting for the uncertainty in raw exposure data, extrapolating the missing exposure values, and untangling the linkage among built environment, bicycle activities, and the frequency of BMV crashes within a unified framework. To promote safer cycling, designated facilities should be provided to consecutively separate cyclists from motor vehicles.

1. Introduction

Although the benefits of cycling have been well documented, cyclists represent one of the most vulnerable types of road users and broadly sustain higher risks of fatality and injury than motorists (Blaizot et al., 2013; Feleke et al., 2018). This problem is a particular concern in highly motorized areas without a mature cycling culture, where the fear of being involved in a collision becomes a substantial barrier to the popularity of cycling (Handy et al., 2014). A deeper understanding of the factors that contribute to bicycle crashes is therefore imperative if cycling is promoted as a safe and attractive mode of transport. Improvement in safety levels will also encourage more people to cycle regularly for daily travel, accompanied by health benefits, mobility options, independence, and fun.

Throughout the last two decades, modeling bicycle crashes involving contiguous spatial units, such as traffic analysis zones and census tracts, has attracted ongoing research interest (See Table A1 in Appendix A). This allows local authorities to identify the spatial pattern of bicycle crashes, to better determine the factors associated with the incidence of bicycle crashes, to forecast the safety burdens sustained by cyclists, and to recommend targeted countermeasures.

Previous studies have suggested that road-network characteristics of a neighborhood, such as its intersection types (Wei and Lovegrove, 2013; Chen, 2015), traffic-signal density (Wei and Lovegrove, 2013;

☆ This manuscript is handled by Associate Editor John Ivan

* Corresponding author at: School of Traffic and Transportation Engineering, Central South University, Changsha, Hunan 410075, China.
E-mail address: hanchuzhou@csu.edu.cn (H. Zhou).

https://doi.org/10.1016/j.aap.2021.106518
Received 14 January 2021; Received in revised form 8 October 2021; Accepted 29 November 2021
Available online 8 December 2021

Chen, 2015; Cai et al., 2016; Nashad et al., 2016; Lee et al., 2015; Osama and Sayed, 2017; Guo et al., 2018a; Kamel et al., 2019; Kamel and Sayed, 2020; Yasmin et al., 2021), road lengths (Wei and Lovegrove, 2013; Kamel et al., 2019), road function (Noland and Quddus, 2004; Gladhill and Monsere, 2012; Cai et al., 2016; Nashad et al., 2016; Saha et al., 2018; Sener et al., 2021; Yasmin et al., 2021), speed limits (Siddiqui et al., 2012; Chen, 2015; Lee et al., 2015), bicycle-lane lengths (Wei and Lovegrove, 2013; Chen, 2015; Yao and Loo, 2016; Osama and Sayed, 2017; Tasic et al., 2017; Yasmin et al., 2021), sidewalk lengths (Cai et al., 2016; Nashad et al., 2016), bus-only lane lengths (Kim et al., 2010; Yasmin et al., 2021), and network topology (Gladhill and Monsere, 2012; Osama and Sayed, 2016; Guo et al., 2018a; Kamel and Sayed, 2020) were closely associated with the frequency of bicycle crashes.

Land-use also has a significant influence on the frequency of bicycle crashes. Specifically, higher proportions of commercial (Kim et al., 2010; Gladhill and Monsere, 2012; Kamel et al., 2019; Ding et al., 2020; Ding et al., 2021) and green space (Ding et al., 2020; Ding et al., 2021) land-use have been found to be associated with an increased likelihood of bicycle crashes. Similar conclusion holds true for the effect of mixed land-use (Chen, 2015; Amoh-Gyimah et al., 2016; Chen et al., 2018a; Kamel and Sayed, 2020; Yasmin et al., 2021). Inconsistent results, however, were found for the effect of residential land-use, as Kim et al. (2010), Amoh-Gyimah et al. (2016), Yao and Loo (2016), and Yasmin et al. (2021) reported a significantly positive relationship between the proportion of residential land-use and the frequency of bicycle crashes, whereas Osama and Sayed (2017) drew the opposite conclusion.

In addition, the accessibility of facilities such as bus stops (Kim et al., 2010; Wei and Lovegrove, 2013; Chen et al., 2018a; Kamel et al., 2019; Sener et al., 2021), parking lots (Chen, 2015), hotels (Lee et al., 2015; Nashad et al., 2016), restaurants (Yasmin et al., 2021), hospitals (Yasmin et al., 2021), financial centers (Yasmin et al., 2021), and universities (Yasmin and Eluru, 2016) was found to significantly increase the likelihood of bicycle crashes.

In terms of demographic and socio-economic characteristics, neighborhoods with a denser population (Siddiqui et al., 2012; Cai et al., 2016; Chen et al., 2018a; Kamel et al., 2019; Lee et al., 2020; Sener et al., 2021; Yasmin et al., 2021), higher proportions of the male (Ding et al., 2021), elderly (Noland and Quddus, 2004; Amoh-Gyimah et al., 2016; Ding et al., 2020), minority-ethnic (Yasmin and Eluru, 2016; Saha et al., 2018), and less-educated (Saha et al., 2018; Lee et al., 2020) residents, lower levels of vehicle ownership (Noland and Quddus, 2004; Siddiqui et al., 2012; Amoh-Gyimah et al., 2016; Nashad et al., 2016; Saha et al., 2018; Sener et al., 2021), higher levels of alcohol consumption (Noland and Quddus, 2004), higher employment rates (Siddiqui et al., 2012; Lee et al., 2015; Cai et al., 2016; Nashad et al., 2016; Prato et al., 2016), higher poverty levels (Noland and Quddus, 2004; Kim et al., 2010; Prato et al., 2016; Yasmin and Eluru, 2016), and lower household incomes (Prato et al., 2016; Yao and Loo, 2016; Ding et al., 2020; Lee et al., 2020) were all associated with more bicycle crashes.

Exposure is indispensable in determinig the incidence of bicycle crashes (Vanparijs et al., 2015). Unlike micro-level analysis that the number of passing bicycles can be readily obtained from counting stations or field observations (Schepers et al., 2011; Strauss et al., 2013; Nordback et al., 2014; Strauss et al., 2015; Saad et al., 2019; Shiranibidabadi et al., 2020; Cai et al., 2021a), reliable bicycle-activity data across a broader region required for neighborhood-level analysis are largely unavailable. Earlier studies (Noland and Quddus, 2004; Kim et al., 2010; Siddiqui et al., 2012; Chen et al., 2018a) therefore used population or population density as a surrogate, as this information is routinely reported by local authorities. The use of such an aggregate measure, however, completely neglects the variation in bicycle activities within an area of interest, thereby suffering necessarily from the omitted-variable bias. Although the emerging global positioning systems have recently enabled researchers to estimate bicycle trajectories from smartphone apps (Saha et al., 2018; Sener et al., 2021), public bicycle rental systems (Ding et al., 2020; Ding et al., 2021), or bicycle sharing programs (Xie et al., 2021), the accessible ridership fails fairly to represent the cycling population on the roads, due to biased sampling toward active users (Lee and Senser, 2021). Comparatively, household travel-diary surveys provide a more complete and representative means of estimating territory-wide bicycle activities. Despite its widespread use as an acceptable source to measure bicycle exposure, the potential that these self-reported bicycle trips are subject to the undersampling given budget constraints is typically overlooked. Without an appropriate adjustment of the uncertainty arising from these incomplete exposure data, the aforementioned relationship between the frequency of bicycle crashes and various risk factors is likely biased (Kim and Mooney, 2016; Fridman et al., 2021).

By virtue of the Bayesian inference, we introduce the simultaneous-equation model to tackle the challenges associated with incomplete exposure data commonly encountered in modeling the frequency of bicycle crashes. The merits of our proposed method are illustrated by a crowdsourced dataset of 792 bicycle–motor vehicle (BMV) crashes aggregated from 209 neighborhoods over a 3-year period in Hong Kong. We focus particularly on the BMV crashes here because cyclists are most likely to be fatally and severely injured if being collied by a motor vehicle (Klassen et al., 2014). Such crash type also suffers less likely from underreporting (Imprialou and Quddus, 2019).

The rest of this paper is organized as follows. After a detailed description of the model formulation, the data collected for model calibration are introduced. We then present and interpret model results, briefly summarize and validate the potential of our proposed method, and conclude with a discussion on promising directions for future studies.

2. Methods

2.1. Model formulation

We modeled the frequency of BMV crashes as in previous studies (Osama and Sayed, 2016; 2017; Prato et al., 2016; Guo et al., 2018a; Kamel et al., 2019; Kamel and Sayed, 2020). Let Y_i denote the number of BMV crashes in the ith neighborhood during a period of 3 or 5 years. Aggregating crash data over such a time slot helps to avoid random fluctuations and the regression-to-the-mean phenomenon. V_i and B_i refer to the vehicle and bicycle volumes, respectively, and X_{ik} is the kth explanatory variable related to the contextual characteristics of neighborhood i. Given the potential nonlinear relationship between BMV crash counts and traffic volume (Elvik and Goel, 2019), we have:

$$\begin{aligned} Y_i &\sim Poisson(\lambda_i) \\ ln(\lambda_i) &= \beta_1 + \beta_2 ln(V_i) + \beta_3 ln(B_i) + \sum_{k=4}^{p} \beta_k X_{ik} + u_i + s_i \end{aligned} \tag{1}$$

where λ_i is the parameter of the Poisson model, i.e., the expected number of BMV crashes in the ith neighborhood, β_1 is the intercept, $\beta_k (k = 2, \ldots, p)$ refers to the kth regression coefficient to be estimated, and u_i denotes the unstructured effect, which has an exchangeable normal prior with a mean of 0 and a variance of σ_u^2. Consistent with Siddiqui et al. (2012), Chen (2015), Lee et al. (2015), Amoh-Gyimah et al. (2016), Osama and Sayed (2016), Prato et al. (2016), Osama and Sayed (2017), Chen et al., (2018a), Guo et al. (2018a), Saha et al. (2018), Cai et al. (2019), Kamel et al. (2019), Wen et al. (2019), and Zeng et al. (2021), s_i here is assigned as the intrinsic conditional autoregressive prior proposed by Besag et al. (1991) to account for the spatial correlation in the error terms:

$$s_i \big| s_{j \neq i} \sim Normal\left(\frac{\sum_j c_{ij} s_j}{\sum_j c_{ij}}, \frac{\sigma_s^2}{\sum_j c_{ij}} \right) \tag{2}$$

where c_{ij} represents the non-normalized weight, e.g., $c_{ij} = 1$ if neighborhood i is adjacent to j, otherwise $c_{ij} = 0$. σ_s^2 is the variance parameter. The proportion of unobserved variations due to spatially correlated ef-

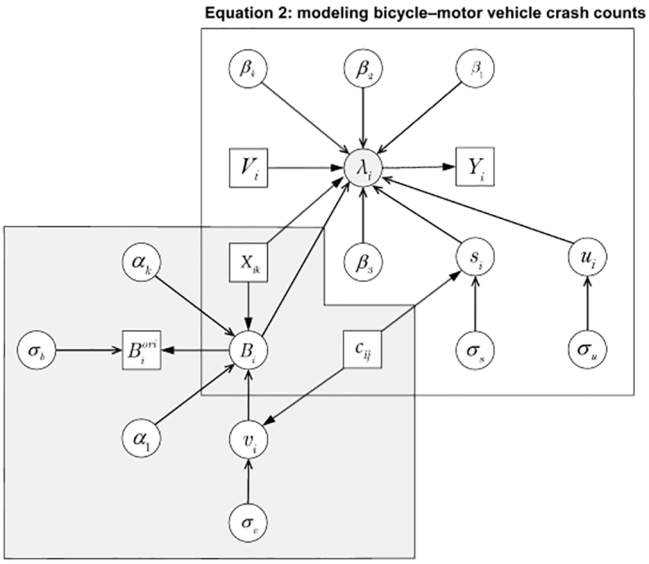

Fig. 1. Directed acyclic graph for the proposed simultaneous-equation model (the symbols are defined in Eqs. (1) to (5)).

fects can then be calculated as (Saha et al., 2018):

$$P_{\mathrm{BMV}} = \frac{\mathrm{sd}(\mathbf{s})}{\mathrm{sd}(\mathbf{s}) + \sigma_u} \tag{3}$$

where sd denotes the marginal standard deviation function, and **s** is the vector of s_i.

Unlike vehicle volume accurately obtained from counting stations, bicycle activities extracted from existing data sources, such as the travel-diary survey, fitness applications, and public-bicycle rental systems, are more likely subject to uncertainty. Inspired by the classic measurement-error models developed by Xie et al. (2018) and Kamel and Sayed (2020), we assume B_i here as a latent variable[1] whose distribution can be well approximated by a function of X_{ik}, as bicycle activities in a neighborhood are closely associated with its land use, road configurations, facility accessibility, and socio-economic characteristics (Osama et al., 2017; Merlin et al., 2020). We therefore have:

$$\begin{aligned} ln\left(B_i^{ori}\right) &\sim Normal\left(ln(B_i), \sigma_b^2\right) \\ ln(B_i) &= \alpha_1 + \sum_{k=1}^{q} \alpha_k X_{ik} + v_i \end{aligned} \tag{4}$$

where B_i^{ori} is the bicycle volume (i.e., cycling distance or cycling time) originally extracted from the exposure data. Unlike Osama et al. (2017) who modeled B_i^{ori} as a normal distribution, we suggest a natural logarithmic transformation here given its inherently non-negative, continuous, and positively skewed nature. σ_b^2 refers to the variance parameter. Similar to s_i, v_i is added to adjust for the unobserved heterogeneity arising from the spatially correlated effects:

$$v_i \left| v_{j\neq i} \right. \sim Normal\left(\frac{\sum_j c_{ij} v_j}{\sum_j c_{ij}}, \frac{\sigma_v^2}{\sum_j c_{ij}}\right) \tag{5}$$

where σ_v^2 is the variance parameter. Likewise, to quantify the magnitude of spatial correlation in the error term, P_{BKT} is calculated as:

$$P_{\mathrm{BKT}} = \frac{\mathrm{sd}(\mathbf{v})}{\mathrm{sd}(\mathbf{v}) + \sigma_b} \tag{6}$$

where **v** refers to the vector of v_i.

Fig. 1 shows the directed graph for the proposed simultaneous-equation model, where the parameters to be estimated have been circled. Note that previous studies on joint modeling of active travel demand and crash frequencies typically assumed the exposure as a deterministic value (Miranda-Moreno et al., 2011; Strauss et al., 2013; Lee et al., 2019; Yasmin et al., 2021). Such a practice is statistically invalid and likely induces biases (van Smeden et al., 2021) when we are uncertain about the true values. Alternatively, by treating B_i as a latent variable whose distribution is readily inferred from the bicycle-activity equation, our approach is capable of appropriately adjusting for the uncertainty associated with B_i^{ori} during the estimation of BMV crash-frequency equation.

Since the frequentist method is not flexible to calibrate the aforementioned equations simultaneously, we used the Bayesian inference to estimate the parameters given its advantages in handling complex and high-dimensional data structures (Gelman et al., 2013; McElreath, 2020). Let $\boldsymbol{Y}$, $\boldsymbol{\lambda}$, $\boldsymbol{V}$, $\boldsymbol{B}$, $\mathbf{B}^{\mathbf{ori}}$, $\boldsymbol{X}$, $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, and $\boldsymbol{u}$ represent the vector of Y_i, λ_i, V_i, B_i, B_i^{ori}, X_{ik}, β_k, α_k, and u_i, respectively. $\boldsymbol{c}$ denotes the spatial weight matrix of c_{ij}. Based on Bayes' theorem, the posterior distribution of $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, σ_u, σ_s, σ_b, and σ_v given $\boldsymbol{Y}$, $\boldsymbol{V}$, $\mathbf{B}^{\mathbf{ori}}$, $\boldsymbol{X}$, and $\boldsymbol{c}$ can be estimated as:

$$\begin{aligned} &p\left(\boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma_u, \sigma_s, \sigma_b, \sigma_v \left| \boldsymbol{Y}, \boldsymbol{V}, \mathbf{B}^{\mathbf{ori}}, \boldsymbol{X}, \boldsymbol{c}.\right.\right) \propto \\ &p(\boldsymbol{Y}|\boldsymbol{\lambda}.) p(\boldsymbol{\lambda}|\boldsymbol{\beta}, \boldsymbol{V}, \boldsymbol{B}, \boldsymbol{X}, \boldsymbol{s}, \boldsymbol{u}) p(\boldsymbol{\beta}) p(\boldsymbol{s}) p(\boldsymbol{u}|\sigma_u.) p(\sigma_u) \times p\left(\boldsymbol{B} \left| \mathbf{B}^{\mathbf{ori}}, \sigma_b, \boldsymbol{\alpha}, \boldsymbol{X}, \boldsymbol{v}\right.\right) p(\sigma_b) p(\boldsymbol{\alpha}) p(\boldsymbol{v}) \end{aligned} \tag{7}$$

where $p(\boldsymbol{Y}|\boldsymbol{\lambda}.)$ refers to the likelihood of $\boldsymbol{Y}$, i.e., $L(\boldsymbol{\lambda}; \boldsymbol{Y})$. $p(\boldsymbol{\lambda}|\boldsymbol{\beta}, \boldsymbol{V}, \boldsymbol{B}, \boldsymbol{X}, \boldsymbol{s}, \boldsymbol{u})$ is the distribution of $\boldsymbol{\lambda}$ conditioned on $\boldsymbol{\beta}$, $\boldsymbol{V}$, $\boldsymbol{B}$, $\boldsymbol{X}$, $\boldsymbol{s}$, and $\boldsymbol{u}$. Analogous implications are applied to $p(\boldsymbol{u}|\sigma_u.)$ and $p\left(\boldsymbol{B}\left|\mathbf{B}^{\mathbf{ori}}, \sigma_b, \boldsymbol{\alpha}, \boldsymbol{X}, \boldsymbol{v}\right.\right)$. $p(\boldsymbol{\beta})$, $p(\boldsymbol{\alpha})$, $p(\sigma_u)$, $p(\sigma_b)$, $p(\boldsymbol{s})$, and $p(\boldsymbol{v})$ denote the prior distribution of $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, σ_u, σ_b, $\boldsymbol{s}$, and $\boldsymbol{v}$, respectively. Special caution is required, because although the univariate conditional prior distribution for $\boldsymbol{s}$ and $\boldsymbol{v}$ has been well defined by Eq. (2) and (5), respectively, the corresponding joint prior distributions become unidentifiable with an undefined mean and an infinite variance (Sun et al., 1999):

$$p(\boldsymbol{s}) \propto exp\left[-0.5\left(\sigma_s^2\right)^{-1} \sum_{\widetilde{ij}} c_{ij}\left(s_i - s_j\right)^2\right] \quad p(\boldsymbol{v}) \propto exp\left[-0.5\left(\sigma_v^2\right)^{-1} \sum_{\widetilde{ij}} c_{ij}\left(v_i - v_j\right)^2\right] \tag{8}$$

Another sound strength of the Bayesian updating is that missing values are treated as unknown quantities for which a joint posterior distribution can be inferred during model estimation (Lunn et al., 2012). Based on the Markov chain Monte Carlo methods wherein plausible values for the unknown quantities are synchronously simulated from their corresponding probability distributions, we can extrapolate the posterior-predictive distribution of $\mathbf{B}^{\mathbf{mis}}$ for observations with missing values of $\mathbf{B}^{\mathbf{ori}}$ after learning about $\boldsymbol{\alpha}$, σ_b, and σ_v from neighborhoods with

[1] Latent variables are those that are not exactly observed or measured but can be projected from a mathematical model using observed variables.

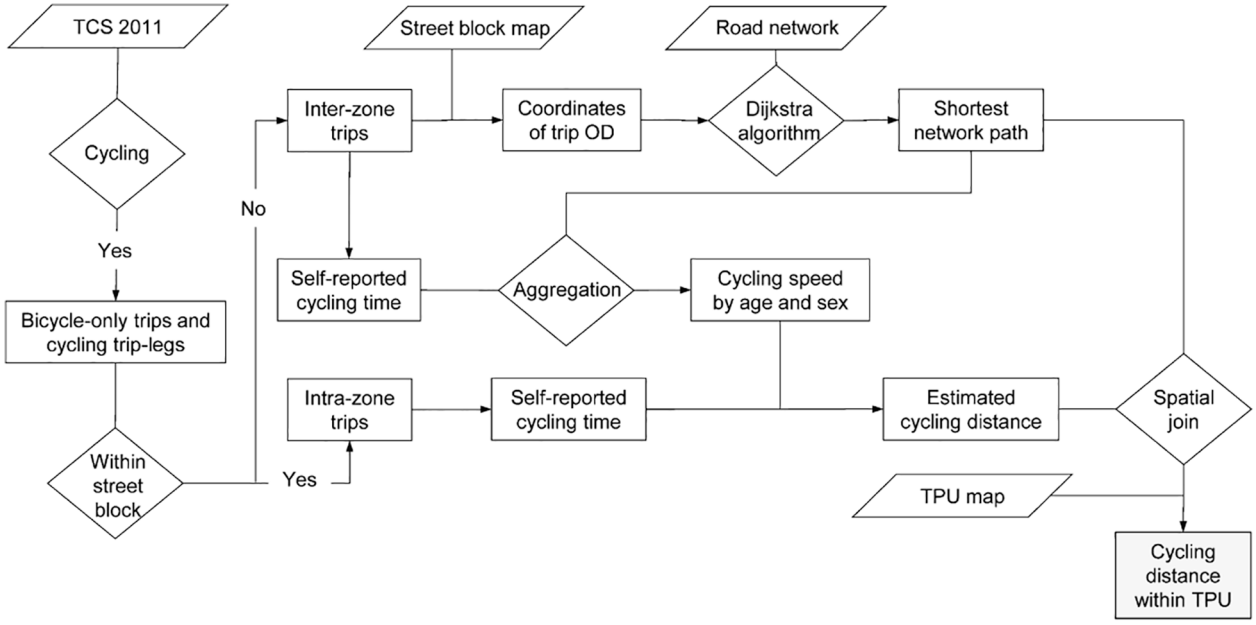

Fig. 2. Flowchart proposed to extract daily bicycle kilometers traveled at the neighborhood level based on the 2011 Hong Kong TCS data (TCS, OD, and TPU refer to the Travel Characteristics Survey, origin and destination, and tertiary planning unit, respectively).

complete records (i.e., $\mathbf{B}^{\text{com}}$), provided that $\mathbf{B}^{\text{mis}}$ and $\mathbf{B}^{\text{com}}$ are *conditionally* independent given $\boldsymbol{\alpha}$, σ_b, and σ_v. That is, the missing-at-random[2] (Rubin, 1976) is supposed to hold after including a wide range of variables that are the potential predictors of $\mathbf{B}^{\text{mis}}$ during the imputation process. On this basis, valid inferences are achievable even without explicitly modeling the missingness mechanism, which is fairly intractable to parameterize in empirical settings (Schafer and Graham, 2002; Sterne et al., 2009). We thus have:

$$p(\mathbf{B}^{\text{mis}}|\mathbf{B}^{\text{com}}.) \propto p(\mathbf{B}^{\text{mis}}|\boldsymbol{\alpha}, \sigma_b, \sigma_v, \boldsymbol{X}, \boldsymbol{c}.) p(\boldsymbol{\alpha}, \sigma_b, \sigma_v|\mathbf{B}^{\text{com}}, \boldsymbol{X}, \boldsymbol{c}) \tag{9}$$

where $p(\mathbf{B}^{\text{mis}}|\boldsymbol{\alpha}, \sigma_b, \sigma_v, \boldsymbol{X}, \boldsymbol{c}.)$ refers to the predictive distribution of $\mathbf{B}^{\text{mis}}$ conditioned on $\boldsymbol{\alpha}$, σ_b, σ_v, $\boldsymbol{X}$, and $\boldsymbol{c}$. $p(\boldsymbol{\alpha}, \sigma_b, \sigma_v|\mathbf{B}^{\text{com}}, \boldsymbol{X}, \boldsymbol{c})$ is the posterior distribution of $\boldsymbol{\alpha}$, σ_b, and σ_v given $\mathbf{B}^{\text{com}}$, $\boldsymbol{X}$, and $\boldsymbol{c}$, which is estimated via:

$$p(\boldsymbol{\alpha}, \sigma_b, \sigma_v|\mathbf{B}^{\text{com}}, \boldsymbol{X}, \boldsymbol{c}) \propto p(\mathbf{B}^{\text{com}}|\boldsymbol{\alpha}, \sigma_b, \sigma_v, \boldsymbol{X}, \boldsymbol{c}.) p(\boldsymbol{\alpha}) p(\sigma_b) p(\boldsymbol{\nu}) \tag{10}$$

where $p(\mathbf{B}^{\text{com}}|\boldsymbol{\alpha}, \sigma_b, \sigma_v, \boldsymbol{X}, \boldsymbol{c}.)$ is the likelihood of $\mathbf{B}^{\text{com}}$, i.e., $L(\boldsymbol{\alpha}, \sigma_b, \sigma_v, \boldsymbol{X}, \boldsymbol{c}; \mathbf{B}^{\text{com}})$.

To obtain the posterior estimates requires the specification of prior distributions. Compared with the vague or flat priors[3] (e.g., normal $(0,100^2)$), we conservatively suggest a weakly informative prior (i.e., normal (0,1)) to model α_k and β_k, acknowledging that the effects of contextual factors are typically small, particularly in terms of standardized predictors (Lemoine, 2019). Large effects arising from limited sample sizes and data noise will be treated skeptically unless there is sufficiently strong evidence supported by high-powered data (Gelman et al., 2017; Lemoine, 2019). Modest regularization provided by such a crafted prior also mitigates the *type I* error and improves the out-of-sample prediction for cases with low statistical power (McElreath, 2020). Because the commonly used inverse-gamma (ε, ε) prior is sensitive to the value of ε if the true variance is close to zero, a uniform (0,10) was assigned to $\sigma_{u,s,b,v}$ following Gelman (2006), Xu et al. (2017), Boggs et al. (2020), Dong et al. (2020), and Bai et al. (2021).

Once the uncertainty due to incomplete exposure data is adjusted, we can calculate the expected risk of collisions with motor vehicles per unit cycled to quantify the safety levels of cycling in an neighborhood (Jacobsen et al., 2015):

$$\text{Risk}_i = \frac{\widehat{\lambda}_i}{\widehat{B}_i} \tag{11}$$

Furthermore, similar in spirit to the path analysis (Miranda-Moreno et al., 2011; Strauss et al., 2013; Gargoum and El-Basyouny, 2016; Liu and Khattak, 2017; Kamel et al., 2019; Lee et al., 2019; Park et al., 2021), by modeling bicycle activities and BMV crashes simultaneously, the linkage among built environment, bicycle activities, and the frequency of BMV crashes can be explicitly untangled. Given $\widehat{\alpha}_k$, $\widehat{\beta}_3$, and $\widehat{\beta}_k$, one unit increase in X_k (X_k has been standardized with a mean of 0 and a standard deviation of 1) is expected to be associated with a $(exp(\widehat{\alpha}_k) - 1) \times 100\%$ increase in bicycle activities. The number of BMV crashes and the risk of collisions with motor vehicles would increase by $(exp(\widehat{\alpha}_k \times \widehat{\beta}_3 + \widehat{\beta}_k) - 1) \times 100\%$ and $(exp(\widehat{\alpha}_k \times \widehat{\beta}_3 + \widehat{\beta}_k - \widehat{\alpha}_k) - 1) \times 100\%$, respectively.

2.2. Data preparation

We fitted the model to a rich dataset collected from multiple sources in Hong Kong. As a highly urbanized area with a well-established public transit system, Hong Kong has a relatively lower level of bicycle use than other metropolitan regions worldwide (Hong Kong Transport Department, 2014). Although traffic injuries and the risk of collision when cycling have increased sharply in Hong Kong, bicycle safety has not been treated as a citywide concern and is under-researched (Xu et al., 2019a). It is thus urgent for local authorities to take action to improve the safety of those vulnerable road users. An accurate identification of the neighborhoods with a substantially higher risk for cycling, along with an in-depth understanding of the factors contributing to the incidence of bicycle crashes, becomes the first step toward the formulation of evidence-based countermeasures in targeted areas.

We obtained crash data from the Traffic Road Accident Database

[2] "Missing-at-random" is an assumption that justifies the analysis, not a property of the data. Such a hypothesis seems reasonable if a variable predictive of missing data is included, but not if that variable is omitted. In practice, as many as variables that are the potential predictors of missing data should thus be included to validate this imputation process.

[3] We avoid using the terminology of noninformative priors, as all priors actually contain some information. Occasionally, noninformative priors can strongly influence posterior distributions and are thus informative.

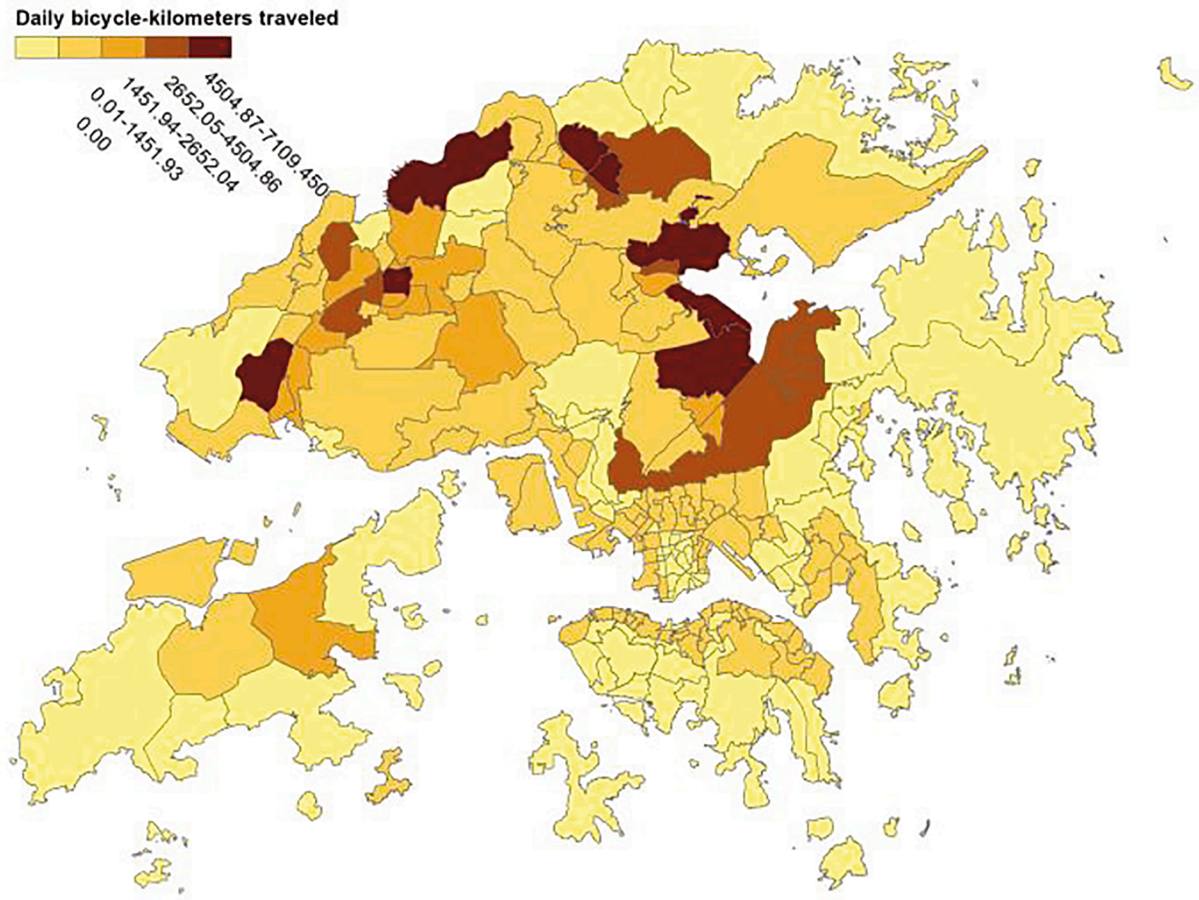

Fig. 3. Spatial distribution of daily kilometers cycled on working days from 2010 to 2012 in Hong Kong, extracted from the 2011 Travel Characteristics Survey.

System maintained by the Hong Kong Police Force (Dong et al., 2020; Zhou et al., 2020; Ye et al., 2021). These data are routinely collected by the police at crash scenes. Only crashes that result in injuries on public roads are recorded in the database (Xu et al., 2019b). Given available geographical coordinate information, we first mapped the crashes onto an ArcGIS map. A total of 792 BMV crashes were reported on normal weekdays (i.e., excluding weekends and public holidays) from 2010 to 2012. Of these, 99.24% were successfully geo-coded following the procedure of Loo (2006). These crashes were then allocated to 209 tertiary planning units (TPUs). We selected TPU as the basic unit of our analysis, because it is the smallest unit adopted by the Hong Kong Census Department when releasing the 2011 Population Census data and is capable of quantifying the built environmental factors at a relatively fine geographical scale (Yao and Loo, 2016; Dong et al., 2020; Su et al., 2021).

We retrieved vehicle volume from the Annual Traffic Census System. The Hong Kong Transport Department publicizes the annual vehicle-flow data recorded by counting stations installed throughout the territory. During 2010–2012, approximately 1600 counting stations including both core and coverage stations were surveyed, covering nearly all trafficable roads in Hong Kong (Hong Kong Transport Department, 2012). We then calculated the average daily vehicle kilometers traveled by multiplying the reported average annual daily traffic by the corresponding length of road segments.

Following Yao et al. (2015), Guo et al. (2017), Sze et al. (2019), Dong et al. (2020), and Su et al. (2021) who extracted pedestrian trajectories from the 2011 Travel Characteristics Survey data (Hong Kong Transport Department, 2014), we also used this latest household travel-diary survey released by the Hong Kong Transport Department to estimate the territory-wide cycling activities. Given the rich information on trip purpose, trip origin, trip destination, departure time, and arrival time collected from a random sample of 35,401 households (i.e., about 1.5% of domestic households) by a tailor-made questionnaire survey on working days between September 2011 and January 2012, an algorithm was elaborately proposed to estimate the daily bicycle kilometers traveled at the TPU level. As illustrated in Fig. 2, we first plotted all cycling trips, including bicycle-only trips and bicycle trip-legs on the ArcGIS map, with the centroid of street blocks as a trip's origin or destination. For inter-zone trips whose origin and destination were not within the same street block, the shortest network distance was generated using Dijkstra algorithm (Dijkstra, 1959) since cyclists typically choose the shortest path. We then estimated the cycling distance for intra-zone trips by multiplying the self-reported cycling time by average cycling speeds stratified by sex and age groups. These age- and sex-specific cycling speeds were calculated by dividing the total distance of the shortest cycling paths estimated for the inter-zone trips by the corresponding cycling time, as shown in Table A2. Finally, by spatially joining the TPU map with the street blocks, we obtained the cycling distance within each neighborhood. Compared with Yao and Loo (2016) who extracted bicycle trajectories based solely on inter-zone trips, our procedure was expected to produce more accurate estimates of bicycle activities by integration of inter-zone and intra-zone trips.

In total, the 2011 Travel Characteristics Survey estimated approximately 46,513 bicycle trips (44,716 inter-zone and 1,797 intra-zone trips), corresponding to 131,246 km cycled by Hong Kong residents per weekday between 2010 and 2012. Fig. 3 depicts the spatial distribution of these bicycle kilometers traveled across 209 TPUs. Of these, 85 neighborhoods had no observations, despite that 57 of which (i.e., about 67%) reported at least one bicycle crash during the period of interest.

We obtained the digital land-use data from the Hong Kong Planning Department. In addition to using the percentage to indicate the intensity of a particular land-use type within a neighborhood, we calculated the

Table 1
Characteristics of the 209 neighborhoods under investigation.

Variables	Mean	SD	Min	Max
Dependent variable				
Number of BMV crashes on working days during 2010–2012	3.76	5.81	0.00	50.00
Exposure variables				
Daily motor-vehicle kilometers traveled ($\times 10^3$)	127.70	153.79	0.36	1204.05
Daily bicycle kilometers traveled	627.97	1358.62	0.00	7109.45
Explanatory variables				
Land-use				
Commercial land-use (%)	4.36	8.91	0.00	53.36
Residential land-use (%)	21.88	16.28	0.00	80.70
Industrial land-use (%)	2.08	7.06	0.00	67.08
Institutional land-use (%)	9.67	10.44	0.00	66.67
Recreational land-use (%)	8.44	9.86	0.00	46.59
Special utilities (%)	16.55	13.77	0.00	55.64
Green space (%)	37.03	33.44	0.00	100.00
Land use mix	0.66	0.20	0.41	1.00
Road-network attributes				
Intersection density (/km)	4.43	1.81	1.00	14.31
Signalized intersections (%)	11.98	11.92	0.00	57.14
Roundabouts (%)	1.70	2.35	0.00	15.38
Three-leg intersections (%)	85.76	12.58	24.00	100.00
Four-leg intersections (%)	13.74	12.35	0.00	76.00
Intersections with five or more legs (%)	0.49	1.13	0.00	7.69
Road density (km/km^2)	10.52	6.98	0.62	37.49
Motorways (%)	6.06	8.45	0.00	57.32
Primary roads (%)	5.24	9.41	0.00	69.40
Secondary roads (%)	10.57	8.74	0.00	36.45
Tertiary roads (%)	10.26	11.07	0.00	77.94
Unclassified roads (%)	67.87	16.58	20.01	100.00
Bus-only lanes (km)	0.12	0.39	0.00	3.76
Bus-stop density (/km)	1.36	0.98	0.00	4.69
Bicycle paths (km)	1.02	2.65	0.00	20.35
Number of bicycle parking lots	2.09	5.48	0.00	47.00
Accessibility of public facilities				
Number of parking lots for motor vehicles	3.11	4.38	0.00	25.00
Number of tram stops	0.61	2.21	0.00	14.00
Number of metro entrances	2.45	4.08	0.00	19.00
Number of petrol stations	0.82	1.22	0.00	5.00
Number of supermarkets	3.23	3.44	0.00	15.00
Number of shopping malls	3.27	4.48	0.00	26.00
Number of convenience stores	6.45	7.83	0.00	49.00
Number of licensed hotels	5.93	24.60	0.00	294.00
Number of hospitals	0.21	0.64	0.00	4.00
Number of clinics	1.18	2.04	0.00	11.00
Number of schools	13.11	17.09	0.00	122.00
Number of police stations	0.22	0.46	0.00	3.00
Number of country parks	0.11	0.40	00.0	3.00
Number of libraries	0.89	1.25	0.00	9.00
Number of museums	0.09	0.45	0.00	5.00
Number of playgrounds	0.30	0.63	0.00	3.00
Number of sports centers	0.49	0.80	0.00	4.00
Number of sports grounds	0.13	0.34	0.00	1.00
Demographic characteristics				
Population density ($\times 10^3$/km^2)	30.60	35.30	0.02	155.98
Male population (%)	46.56	4.43	36.31	85.11
Population aged less than 15 (%)	11.92	2.88	0.00	20.76
Population aged between 15 and 24 (%)	11.00	3.00	2.20	27.25
Population aged between 25 and 44 (%)	33.33	5.16	21.04	57.30
Population aged between 45 and 64 (%)	30.29	3.65	15.36	43.77
Population aged 65 or above (%)	13.47	5.68	0.09	43.81
Population of Chinese ethnicity (%)	88.30	11.96	43.19	99.17
Educational characteristics (highest level attended)				
Population with primary education or below (%)	27.97	8.01	7.24	61.09
Population with secondary education (%)	44.70	7.36	17.35	84.27
Population with post-secondary education (%)	27.33	12.69	0.28	75.41
Economic characteristics				
Working population (%)	50.79	7.20	0.00	82.75
Working population who work at home (%)	12.25	9.72	0.00	50.56
Median monthly income (10^3)	13.86	5.68	0.00	40.00
Household characteristics				
Household density ($\times 10^3$/km^2)	10.66	12.74	0.00	58.09
Average household size	2.93	0.38	1.60	4.10
Median monthly household income ($\times 10^3$)	33.43	29.73	0.00	170.80
Median monthly household rent ($\times 10^3$)	8.06	11.71	0.00	76.00
Median rent to income ratio (%)	19.15	11.35	0.00	53.23
Population in public rental housing (%)	15.51	23.99	0.00	99.97
Population in subsidized ownership housing (%)	7.55	14.46	0.00	78.00
Population in permanent housing (%)	72.10	31.70	0.00	100.00
Population in non-domestic housing (%)	0.95	6.14	0.00	85.30
Population in temporary housing (%)	3.42	8.19	0.00	52.56

entropy index (Chen, 2015; Yasmin and Eluru, 2016; Tasic et al., 2017; Dong et al., 2020; Cai et al., 2021a; Yasmin et al., 2021) to quantify the mixture of land-use as:

$$\text{Entropy}_i = \frac{-\sum_{j=1}^{k_i} p_j^i ln\left(p_j^i\right)}{ln(k_i)} \tag{12}$$

where p_j^i denotes the percentage of land-use type $j(j = 1, 2, ..., 7)$ in TPU $i(i = 1, 2, ..., 209)$ and k_i represents the number of land-use types in the ith TPU. The entropy index ranges from 0 to 1, with a value towards one associated with more mixed land-use.

Our road-network data were derived from the node-link road centerline system provided by the Hong Kong Lands Department, from which various geometric characteristics, including intersection density, road density, percentages of different types of intersections, and percentages of road-segment length with different functional classifications, were obtained. We defined intersection density as the number of intersections divided by the length of road segments, while road density as the length of road segments per square kilometers. We crawled data about bicycle infrastructure, including the locations of bicycle parking lots and the length of bicycle paths, from the GeoInfo Map released by the Hong Kong SAR Government. These facilities were further validated by the shapefile derived from OpenStreetMap. Here we did not distinguish the on-street bicycle lanes (Pulugurtha and Thakur, 2015) from the off-road cycle tracks (Lusk et al., 2011), because in Hong Kong all the bicycle lanes are physically separated and dedicated to bicycle use (hereafter referred to as bicycle paths).

The points-of-interest data were also grabbed from the GeoInfo Map. Given the precise location information, the numbers of various public facilities were counted within each neighborhood. Finally, the demographic, educational, economic, and household characteristics were retrieved from the 2011 Population Census Report.

The variables available for model development, together with their descriptive statistics, are presented in Table 1.

3. Results and discussion

We calibrated the model using the freeware WinBUGS (Spiegelhalter et al., 2005). We tracked three parallel chains with diverse starting values, with the first 50,000 iterations discarded as burn-ins. Another 50,000 iterations were then performed, resulting in a sample of 150,000 for each parameter. To improve efficiency, we implemented *thinning*, by which only each 10th value from the draws was retained for inference (Lunn et al., 2012). We diagnosed model convergence by the Brooks-Gelman-Rubin statistic (Brooks and Gelman, 1998), visual examinations of the Markov chain Monte Carlo chains, and the ratios of Monte Carlo errors to the respective standard deviations of the estimates. As a

Table 2
Results of the bicycle-activity models, with the natural logarithm of daily bicycle kilometers traveled in 124 neighborhoods on working days during 2010–2012 in Hong Kong as the dependent variable.

	Non-spatial model		Spatial-error model†			
Variables	Mean	SD	95% BCI	Mean	SD	95% BCI
Intercept, α_1	5.21*	0.12	(5.01, 5.45)	5.15*	0.13	(4.90, 5.39)
Length of bicycle paths, α_2	0.50*	0.13	(0.23, 0.76)	0.45*	0.12	(0.21, 0.70)
Number of bicycle parking lots, α_3	0.42*	0.13	(0.16, 0.67)	0.36*	0.12	(0.12, 0.60)
Bus-stop density, α_4	–0.45*	0.12	(–0.69, –0.21)	–0.14	0.13	(–0.40, 0.12)
Population with secondary education (%), α_5	0.40*	0.16	(0.08, 0.72)	0.10	0.16	(–0.22, 0.42)
σ_b^2	1.75*	0.23	(1.35, 2.26)	1.08*	0.23	(0.65, 1.55)
σ_v^2				1.49*	0.70	(0.48, 3.15)
P_{BKT}				0.48*	0.06	(0.35, 0.60)
Goodness-of-fit measures						
Deviance information criterion			425.20			394.47
R^2			0.51			0.81
Mean absolute deviance			1.03			0.69
Mean squared prediction error			3.24			1.58
Moran's I test for residuals						
I			0.05*			–0.03
Z-score			2.33			–0.84
p-value			0.02			0.40

† The spatially correlated error term here was specified as the intrinsic conditional autoregressive prior distribution proposed by Besag et al. (1991).
SD: standard deviation.
BCI: Bayesian credible interval.
* denotes significance at 95% BCI.
Note: independent variables were standardized here with a mean of 0 and a standard deviation of 1.

rule of thumb, these ratios should be less than 0.05.

For model specification, we first conducted a multicollinearity test to ensure the non-inclusion of highly correlated variables. We then added all the uncorrelated variables to the initial model and selected covariates via backward elimination (Xie et al., 2018; Dong et al., 2020). The deviance information criterion (DIC; Spiegelhalter et al., 2002) was used to compare the alternative with different covariate subsets. The model with a lower value of DIC was statistically superior.

For the purpose of comparison, in addition to the spatial-error model, we also developed a non-spatial model to explain the geographical variations of bicycle activities. To highlight the benefits of adjusting for uncertainty in incomplete exposure data when modeling the frequency of BMV crashes, we first developed the model without bicycle volume as the benchmark. We then compared the results of our simultaneous-equation model with those derived from the candidate model suggested by Wei and Lovegrove (2013) and Yao and Loo (2016), wherein a deterministic value of one was imputed to the neighborhoods that initially reported zero bicycle kilometer traveled. Such a practice is motivated by that instead of entirely discarding a crucial variable due to incomplete information, it seems more desirable to fill in the missing data with plausible values. As such, five models were ultimately calibrated. Below we first present and interpret the results of bicycle-activity models, followed by those of BMV crash-frequency models. A good interpretation of parameter estimates also helps to justify the validity of the proposed method.

3.1. Bicycle-activity model results

Table 2 summarizes the results of the parameter estimates and goodness-of-fit measures for the spatial and non-spatial models, with the natural logarithm of daily bicycle kilometers traveled as the dependent variable. We used the 95% Bayesian credible interval (BCI) to determine whether the parameters differed credibly from zero. Variables that were insignificant in both models were excluded for parsimony and interpretability (Dong et al., 2020; Yasmin et al., 2021).

As shown in Table 2, the spatial-error model evidently outperformed with a significantly higher value of R^2 and lower values of DIC, mean absolute deviance, and mean squared prediction error, suggesting that model goodness-of-fit can be substantially improved by explicit consideration of spatial correlation in the error term. The superiority of the spatial-error model was also supported by that a moderate proportion of unobserved heterogeneity was explained by the spatially correlated effects (i.e., $\widehat{P}_{\mathrm{BKT}}$ resulted in a posterior estimate with a mean of 0.48 and a standard deviation of 0.06) and that the residuals of the non-spatial model were spatially correlated at 95% confidence level according to the Moran's I statistic (Moran, 1950), although they were assumed to be spatially independent in both models. This underlying hypothesis was violated and the non-spatial model likely produced biased parameters (Ziakopoulos and Yannis, 2020). It is not surprising to find that the variables of bus-stop density and the proportion of population with secondary education became less significant after an adjustment of spatially correlated effects. We therefore interpreted our parameter estimates based on the results of spatial-error model.

The configuration of bicycle facilities had a significant influence on bicycle activities. Specifically, the length of bicycle paths and the number of bicycle parking lots were positively associated with the number of bicycle kilometers traveled. This result was plausible, as the presence of bicycle facilities attracts more people to commute by bicycle (Pucher et al., 2010; Buehler and Pucher, 2012; Schoner and Levinson, 2014). Bicycle facilities are also more likely to be provided in neighborhoods with prevalence of cyclists.

Given the parameters estimated by the spatial-error model from 124 neighborhoods with records of bicycle activities, we inferred the posterior-predictive distribution of daily bicycle kilometers traveled for the remaining 85 neighborhoods whose bicycle activities were reported to be zero by the travel-diary survey. As illustrated in Fig. 4, the mean values of these posterior-predictive distributions ranged widely from 51.49 to 1658.00, indicating that the arbitrary imputation of a deterministic value for all neighborhoods with missing records following Wei and Lovegrove (2013) and Yao and Loo (2016) would necessarily ignore the variations in bicycle activities across the areas of interest. A closer look into the relationship between the frequency of BMV crashes and the extrapolated mean value of daily bicycle kilometers traveled suggests an absence of significant correlation, as the Spearman's correlation parameter was estimated at –0.12 with a p-value of 0.28. Given that neighborhoods without BMV crashes were not less likely to report bicycle activities, the missing-at-random hypothesis was validated partially as the missingness was not related to the response variable (Afghari et al., 2019). We therefore attribute these missing values to that cyclists were likely undersampled due to budget restrictions, particularly in our case that cycling was a minor mode primarily for recreational purposes, with a share of less than 1% of all trips made by local residents on working days (Xu et al., 2019a).

Overall, our spatial-error model estimated a total of 164,300 (95% BCI: 100,300 to 286,300) kilometers cycled by Hong Kong residents per weekday during 2010–2012. This figure was much higher than that originally extracted from the travel-diary survey (i.e., 131,246). Recall

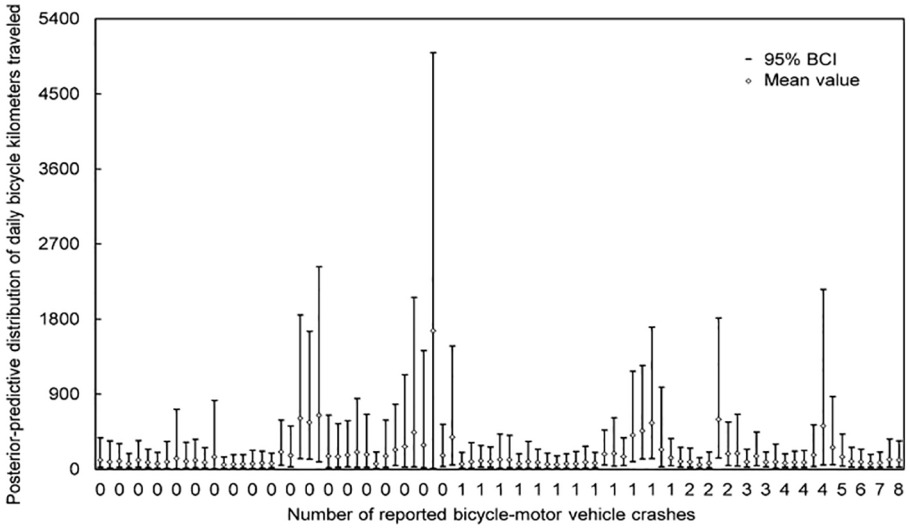

Fig. 4. Posterior-predictive distribution of bicycle kilometers traveled for 85 neighborhoods with missing exposure data, estimated by the spatial-error model in Table 1.

Table 3
Parameters estimated by the BMV crash-frequency models aggregated in 209 neighborhoods during 2010–2012 in Hong Kong.

	Model 1[†]			Model 2[‡]			Model 3[§]		
Variables	Mean	SD	95% BCI	Mean	SD	95% BCI	Mean	SD	95% BCI
Intercept, β_1	–2.54*	0.34	(–3.22, –1.88)	–2.53*	0.33	(–3.19, –1.89)	–3.40*	0.35	(–4.10, –2.71)
ln(vehicle kilometers traveled), β_2	0.72*	0.07	(0.58, 0.86)	0.65*	0.07	(0.51, 0.80)	0.68*	0.07	(0.54, 0.83)
ln(bicycle kilometers traveled), β_3	——	—	————	0.08*	0.03	(0.03, 0.14)	0.20*	0.07	(0.07, 0.34)
Length of bicycle paths, β_4	0.16*	0.06	(0.05, 0.27)	0.11	0.06	(–0.01, 0.22)	——	—	————
Land-use mix, β_5	0.23*	0.07	(0.09, 0.37)	0.20*	0.07	(0.06, 0.34)	0.22*	0.07	(0.10, 0.37)
Motorways (%), β_6	–0.35*	0.08	(–0.51, –0.19)	–0.36*	0.08	(–0.52, –0.20)	–0.35*	0.08	(–0.50, –0.19)
Thee-leg junctions (%), β_7	–0.29*	0.07	(–0.42, –0.16)	–0.29*	0.07	(–0.41, –0.16)	–0.30*	0.06	(–0.43, –0.18)
Median monthly income, β_8	–0.17*	0.08	(–0.33, –0.01)	–0.14	0.08	(–0.30, 0.02)	–0.16*	0.08	(–0.31, –0.01)
σ_u^2	0.05	0.04	(0.00, 0.16)	0.05	0.05	(0.00, 0.16)	0.05	0.04	(0.00, 0.15)
σ_s^2	0.88*	0.25	(0.46, 1.43)	0.74*	0.23	(0.34, 1.25)	0.64*	0.24	(0.24, 1.19)
var(s)	0.54*	0.10	(0.36, 0.75)	0.40*	0.09	(0.22, 0.61)	0.34*	0.12	(0.14, 0.61)
P_{BMV}	0.80*	0.10	(0.62, 0.98)	0.78*	0.11	(0.57, 0.98)	0.76*	0.12	(0.53, 0.98)
Goodness-of-fit measures									
Deviance information criterion	764.14	762.76	764.10						
Mean absolute deviance	1.76	1.76	1.78						
Mean squared prediction error	7.47	7.50	7.50						

[†] Without the inclusion of bicycle exposure measures.
[‡]Following Wei and Lovegrove (2013) and Yao and Loo (2016), a value of one was assigned to the neighborhoods if their bicycle kilometers traveled were originally estimated to be zero by the travel-diary survey data.
[§]Bayesian simultaneous-equation model. Here we modeled bicycle kilometers traveled as a latent variable whose distribution was inferred from the spatial-error model presented in Table 2. Since the length of bicycle path was not statistically significant, we removed it from the model to avoid multicollinearity.
SD: standard deviation.
BCI: Bayesian credible interval.
*denotes significance at 95% BCI.
Note: independent variables were standardized with a mean of 0 and a standard deviation of 1.

that the assignment of bicycle trips in our study depends on a strong assumption that cyclists choose the shortest path. In reality, the route choice of cyclists, however, is inherently complex and dynamic (Broach et al., 2012; Chen et al., 2018b). As a result, the measurement errors induced during the estimation of cycling distance based solely on the available trip-origin and destination information, together with the potential of underreporting of bicycle trips, plausibly lead to this underestimation of bicycle activities.

3.2. BMV crash-frequency model results

Table 3 presents the model results for the frequency of BMV crashes. Although the goodness-of-fit measures in terms of DIC, mean absolute deviance, and mean squared prediction error were not substantially different among the three candidate models, their significant variables were not completely identical. For example, the length of bicycle paths was statistically significant in the benchmark model but became totally insignificant after the inclusion of the exposure variable. Likewise, the variable of median monthly income was solely insignificant in model 2. A comparison between model 2 and 3 further indicates that the coefficient of bicycle volume increased sharply from 0.08 to 0.20. Its standard deviation increased accordingly from 0.03 to 0.07, resulting in a much wider confidence interval. Xie et al. (2018) and Kamel and Sayed (2020) also reported such a dilution bias and a similar underestimation of the uncertainty if measurement errors in the exposure were ignored. In addition, although σ_u^2 was relatively stable across the three models, *var*(s) reduced gradually from 0.54 to 0.34 among the three models, suggesting that an adjustment of the uncertainty in the raw exposure data

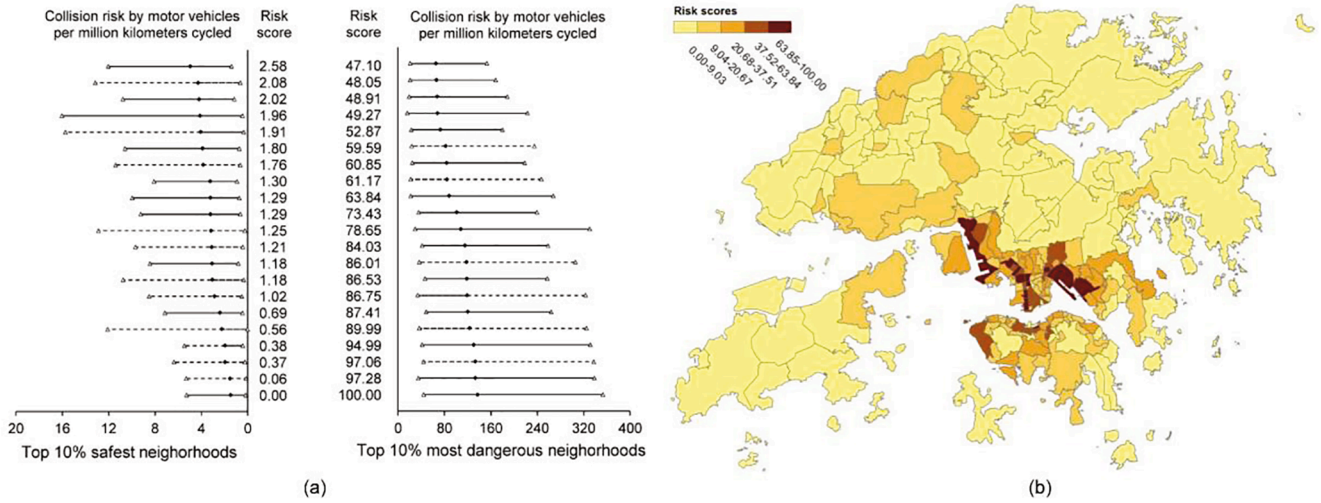

Fig. 5. Hot-zone identification results: (a) top 10% safest and most dangerous neighborhoods for cycling in terms of expected risk of collisions by motor vehicles per million kilometers cycled on working days during 2010–2012 in Hong Kong, estimated by the Bayesian simultaneous-equation model (solid dots: mean values; hollow dots: upper and lower bounds of the 95% BCI; dotted lines: neighborhoods whose raw exposure data were missing, with extrapolated values by the spatial-error model in Table 2; risk scores represent the normalized values of collision risk using the mix-max feature-scaling method); and (b) risk scores for cycling across the whole territory on working days during 2010–2012 in Hong Kong.

helps to reduce the unobserved heterogeneity arising particularly from the spatially correlated effects. These findings raise an alarm that either the omission of activity-based exposure measures or the neglect of the uncertainty in raw exposure data when modeling the frequency of bicycle crashes probably leads to biased estimates and invalid inferences.

As shown in Fig. 5(a), once the bias due to incomplete exposure data was adjusted, we could calculate the expected risk of collisions with motor vehicles per million kilometers cycled to quantify the safety of cycling across the whole territory, even for neighborhoods originally with missing bicycle-activity records. A mapping of the risk scores further helps to identify neighborhoods where cyclists were more likely to be collided by motor vehicles (Kamel et al., 2019, 2020). As Fig. 5(b) illustrates, these hot zones were mostly clustered in the dense urban areas of Hong Kong Island and Kowloon, where the road system has long been dominated by motor vehicles without adequate bicycle infrastructure (Xu et al., 2019a). Cycling in these motorist-oriented neighborhoods is not only extremely unsafe but also inconvenient and unpleasant, because cyclists have to share the roads crowded with motor vehicles.

Also importantly, an integrated interpretation of the results presented in Tables 2 and 3 revealed that the effect of bicycle facilities on the frequency of BMV crashes was fully mediated by bicycle activities. Accordingly, one unit increase in bicycle-path lengths (i.e, 2.65 km) and the number of bicycle parking lots (i.e., 5.48) in a neighborhood was associated with a roughly 58.14% (95% BCI: 23.11% to 100.70%) and 44.07% (95% BCI: 12.78% to 81.50%) increase in daily bicycle kilometers traveled, respectively. Owing to the non-linearity between bicycle volume and the number of BMV crashes, the risk of collisions by motor vehicles per kilometers cycled was expected to decrease by 29.80% (95% BCI: 14.82% to 44.07%) and 24.41% (95% BCI: 9.07% to 38.78%), although the absolute number of BMV crashes would increase slightly by 9.60% (95% BCI: 2.32% to 19.05%) and 7.63% (95% CI: 1.36% to 16.38%), respectively. Such a non-linear relationship, also known as the safety-in-numbers effect (Jacobsen, 2003; Bhatia and Wier, 2011; Jacobsen et al., 2015; Elvik and Bjørnskau, 2017; Elvik and Goel, 2019; Xu et al., 2019b), has been previously reported (Strauss et al., 2013; Tasic et al., 2017; Kamel et al., 2019; Lee et al., 2020; Cai et al., 2021b). Since a greater number of cyclists are more visible (Jacobsen et al., 2015), one plausible explanation should be that motorists adjust their behaviors such as by lowering driving speeds (Klieger and Savage, 2020) and being more willing to yield (Thompson et al., 2019) when encountering a group of cyclists. For another, the accumulation of experience gained from more cycling also enables people to ride skillfully (Schepers, 2012), thereby improving their riding performance on the roads.

Unlike Chen (2015), Prato et al. (2016), and Osama and Sayed (2017) who reported a significantly negative association between the length of off-street bicycle paths and the frequency of BMV crashes after controlling for bicycle exposure, the absence of a directly protective effect of separated bicycle lanes in our study is probably due to the hazards of riding on fragmented cycle tracks. Since most of the bicycle paths in Hong Kong are fragmentarily located in newly developed towns in the New Territories (Xu et al., 2019a), very often, people who would like to ride on cycle tracks have to use public roads for linking separate segments of these paths, which potentially increases the likelihood of involvement in a BMV crash (Loo and Tsui, 2010). We therefore conjecture that the safety benefits reaped from riding on physically seperated bicycle tracks have been completely compensated by the risk associated with their inconsecutive layouts. Further efforts are warranted to validate this hypothesis.

In addition to encouraging more people to cycle, another effective countermeasure to improve bicycle safety would be to restrict the use of motor vehicles. Besides the benefits of reduced congestion, less emission of pollutants, and decreased traffic noise, strategies to reduce traffic volume will lower both the number and risk of BMV crashes. As evidenced by our results, all things being equal, a neighborhood halving its vehicle kilometers traveled would expect an roughly 37.58% (i.e., $(1-0.5^{0.68})\times 100\%$) decrease in BMV crashes. Therefore, a shift from motor vehicles to other travel modes, such as public transits, cycling, and walking, should be vigorously advocated, particularly in urban settings.

Interestingly, land-use mixture was associated with an increased risk of BMV crashes. A visual examination via Google Street View suggests that traffic conditions in neighborhoods with mixed land-use in Hong Kong are considerably complex and diverse, as various road users share the same activity spaces. The frequent interactions between motor vehicles and cyclists in these neighborhoods undesirably increase the risk of collisions (Chen, 2015; Amoh-Gyimah et al., 2016; Chen et al., 2018a; Kamel and Sayed, 2020).

Road functional classification also had a significant relationship with the frequency of BMV crashes. Specifically, neighborhoods with a higher percentage of motorways were associated with a lower risk of BMV crashes. This result is intuitively reasonable, because in Hong Kong all motorways are limited-access roads aimed at serving fast-moving motor vehicles, on which cycling is prohibited (Dong et al., 2020).

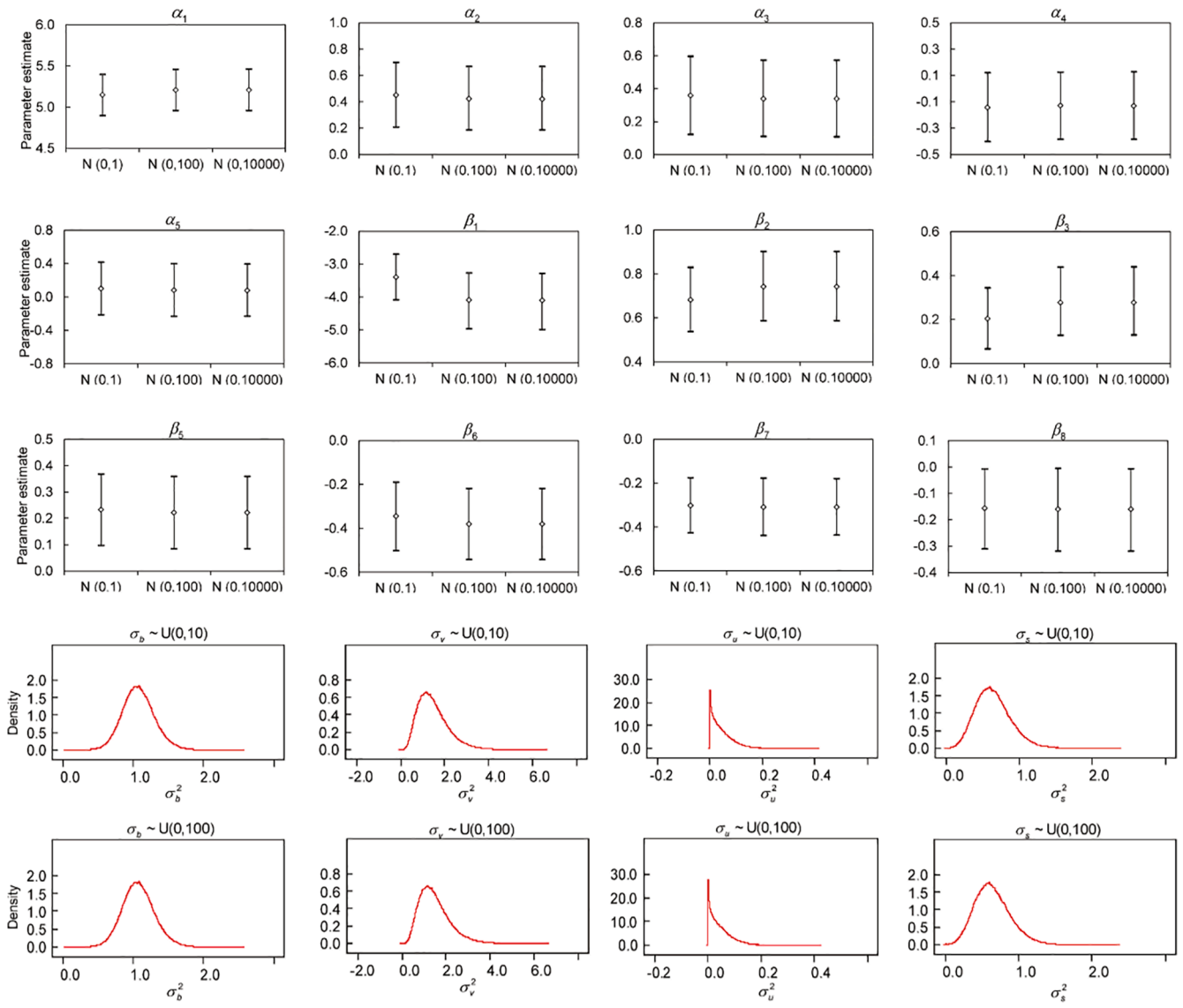

Fig. 6. Sensitivity to prior specifications, with parameters estimated by the Bayesian simultaneous-equation model presented in Table 3 (hollow dots: mean value; solid dots: upper and lower bounds of the 95% BCI).

Intersections are well-known as hazardous locations in road networks (Reynolds et al., 2009; Guo et al., 2018b; Bai and Sze, 2020). Since intersections with fewer legs generally possess fewer conflict points between cyclists and motor vehicles, it is not surprising that the proportion of intersections with three legs had a significantly negative relationship with the frequency of BMV crashes. Similar findings were also reported by Chen (2015).

Consistent with Kim et al. (2010), Prato et al. (2016), and Yao and Loo (2016), the negative relationship between median monthly income and the frequency of BMV crashes found here is plausibly attributed to two causes. First, people with higher incomes may have better awareness of safe cycling. Second, deprived neighborhoods probably have fewer facilities, such as bicycle boxes and raised bicycle crossings, designated for bicycle use.

Finally, P_{BMV} produced a posterior distribution with a mean of 0.76 and a standard deviation of 0.12, implying that an appreciable amount of unobserved heterogeneity was explained by the spatially correlated effects. Taken together with the result of P_{BKT}, spatial correlation should not be neglected when modeling bicycle activities and BMV crash frequencies aggregated at contiguous spatial units.

3.3. Validation

3.3.1. Sensitivity analysis of prior specifications

Given the nonexistence of ground-truth priors, as an integral part of the Bayesian analysis, sensitivity to alternative prior assumptions should be justified and reported formally (Gelman et al., 2017). In addition to the weakly informative prior (i.e., normal (0,1)) suggested by Lemoine (2019), we also considered the vague options, i.e., normal ($0,10^2$) and normal ($0,100^2$), for α_k and β_k, respectively. To determine the uniform distribution for $\sigma_{u,s,b,v}$, we first preset its upper value as 10 to stabilize the computations. We then gradually increased the value by a step of 10 until the model failed to converge (Boggs et al., 2020). The results are shown in Fig. 6. Generally, our parameters were insensitive and robust to prior specifications, as the posterior distributions almost always overlapped, with negligible variations in the range and shape. Only β_1 shrunk slightly toward zero. These results demonstrate that our data were sufficient to draw robust and credible inferences.

3.3.2. Cross-validation

To evaluate the extrapolation performance of the spatial-error model, we performed a leave-one-out cross-validation (James et al., 2013; Xu et al., 2021). Given the 124 neighborhoods with bicycle-

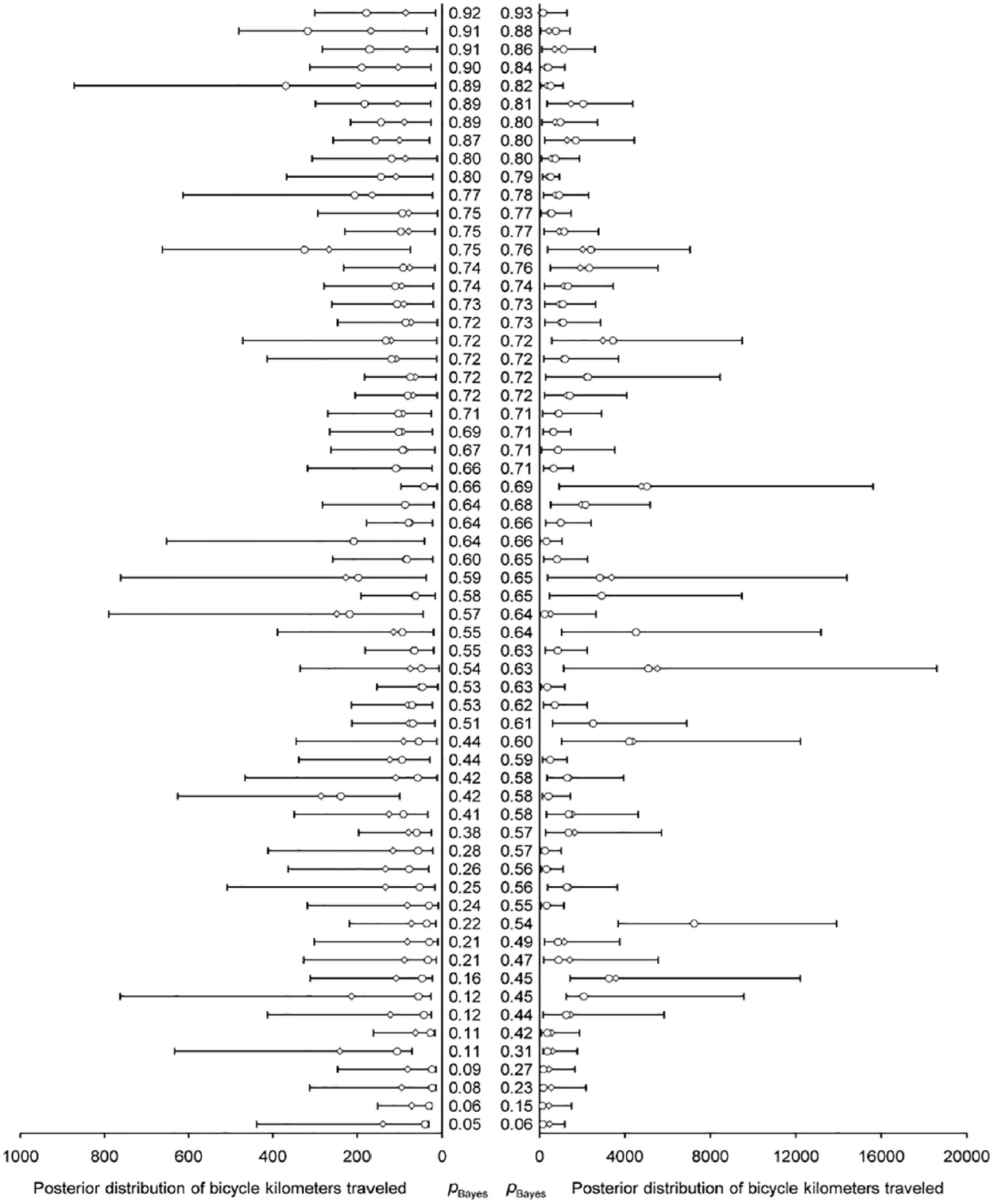

Fig. 7. Posterior-predictive distribution of daily bicycle kilometers traveled in the leave-one-out cross-validation, estimated by the spatial-error model (hollow circles: observed values; hollow rectangles: mean values; solid dots: upper and lower bounds of the 95% BCI).

activity records, one observation was left out as the validation whereas the remaining was combined for calibration. This process was repeated 124 times. We then estimated the posterior-predictive distribution of the validation samples using Eq. (9). We also calculated the Bayesian *p*-value[4] (Lunn et al., 2012) to assess whether the residual was significantly greater than zero. The results are presented in Fig. 7.

As Fig. 7 illustrates, the observed data were statistically compatible with the predictions, as nearly all of the observed values (i.e., 123 out of 124) fell within the 95% BCI of the predictive-posterior distributions, resulting in a Bayesian *p*-value ranging from 0.05 to 0.93. We were therefore confident that our spatial-error model performed satisfactorily in imputing missing bicycle exposure data.

[4] Please refer to Appendix B for more details.

4. Conclusions

A major challenge faced by the neighborhood-level bicycle safety analysis is the unavailability of complete and reliable bicycle exposure data for the whole area under investigation. Although the conventional household travel-diary surveys combined with the emerging fitness-applications and public bicycle-rental systems provide straightforward and valuable opportunities to measure territory-wide bicycle activities, the obtained ridership data suffer inherently from under-sampling and unrepresentativeness. In this study, we introduced an ad hoc approach namely the Bayesian simultaneous-equation model to tackle the challenges of analyzing BMV crashes using incomplete exposure data. By simultaneously modeling bicycle activities and the frequency of BMV crashes, our method is promising in addressing the incompleteness in raw exposure data, extrapolating the bicycle activities for observations with missing values, and untangling the linkage of built environment, bicycle activities, and the frequency of BMV crashes within a unified framework.

After successful model calibrations on a rich dataset aggregated in 209 neighborhoods over a 3-year period in Hong Kong, we empirically revealed the existence of uncertainty in bicycle volume data. Our explanatory analysis further demonstrated the biases arising from omission of activity-based exposure measures or from the direct use of raw exposure data without correcting for the uncertainty. Specifically, an appropriate adjustment of the missing exposure values empowered us to quantify the risk of being collided by motor vehicles for cyclists throughout the whole territory. Such a thorough risk profile is of paramount importance for local authorities in identification of targeted neighborhoods to improve the safety and mobility of cyclists, because without controlling for exposure, it is incapable to determine whether the overrepresentation of BMV crashes occurs as a result of environmental hazards or simply due to more people cycling.

By jointly modeling bicycle activities and the frequency of BMV crashes, we added new insights that an expansion of bicycle infrastructure is very likely associated with a significant increase in cycling levels and a substantial reduction in the risk of collisions by motor vehicles, despite a slight increase in the absolute number of BMV crashes. This result should be robust enough, particularly after adjusting for the uncertainty in exposure data, the unobserved heterogeneity due to spatial correlation, and the confounding effects of more than 60 variables related to land use, road-network attributes, the accessibility of public facilities, and demographic and socio-economic characteristics. Based on our findings, the provision of dedicated facilities to consecutively separate cyclists from motor vehicles should be the top priority in the formulation of effective strategies to promote safer cycling (Reynolds et al., 2009; DiGioia et al., 2017; Prati et al., 2018). Special attention should also be paid to the neighborhoods characterized by massive motor vehicles, mixed land-use, numerous complex junctions, and higher deprived levels, because cycling in these areas sustained a significantly higher risk of collisions with motor vehicles.

Beyond this application to BMV crashes, our approach can be readily generalizable to other research domains, such as the epidemiological and environmental studies where complete exposure measurements are also rarely accessible (van Smeden et al., 2021). To encourage a more thorough validation of our method if additional data sources such as the bicycle route-choice models (Zimmermann et al., 2017), smartphone fitness-application data (Strauss et al., 2015; Chen et al., 2020; Cai et al., 2021b), and bike-sharing data (Guo et al., 2021) become available in future studies, following Lemoine (2019) and Dong et al. (2020), we have shared our programming codes which are simple to manipulate by the freeware software WinBUGS in the Appendix C. In addition, although the neighborhood-level analysis is insightful and effective in investigating the effects of planning-level variables on the frequency of BMV crashes, bicycle safety is also a microscopic concern, as BMV crashes are usually caused by conflicts between cyclists and motor vehicles. To achieve a deeper insight into the causes of BMV crashes, further efforts that integrate cross-sectional research designs with crash-reconstruction simulations and in-depth crash-causation investigations are highly recommended.

Funding

This study was supported by grants from the Natural Science Foundation of China (Project No. 71671100), the Guangzhou – Hong Kong – Macau Joint Laboratory Program of the 2020 Guangdong New Innovative Strategic Fund, Guangdong Science and Technology Department (Project No. 2020B1212030009), and the Fundamental Research Funds for the Central Universities of Central South University (Project No. 2019zzts868). S.C. Wong was also supported by the Francis S Y Bong Professorship in Engineering. The funders had no roles in study designs, data collection and analysis, decision to publish, or preparation of the manuscript.

CRediT authorship contribution statement

Pengpeng Xu: Conceptualization, Formal analysis, Data curation, Validation, Visualization, Writing – original draft. **Lu Bai:** Writing – review & editing. **Xin Pei:** Writing – review & editing. **S.C. Wong:** Resources, Funding acquisition, Writing – review & editing. **Hanchu Zhou:** Software, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We sincerely appreciated the three anonymous reviewers for their insightful, constructive, and thoughtful comments that have help us to revise our manuscript. We would also like to thank the Hong Kong Transport Department, Hong Kong Police Force, and Hong Kong Planning Department for providing access to the database used in this study. The views expressed are the authors' own and do not necessarily represent those of the Hong Kong Transport Department and the Hong Kong Police Force.

Data availability statement

The accident and travel-diary survey datasets are owned and were made accessible to the authors by the Hong Kong Transport Department. Access to the data can be obtained, subject to compliance with the terms of service, by signing an agreement with the Hong Kong Transport Department. We confirm that we did not have any special access to the data that others would not have.

Appendix A

Table A1
Studies of factors that influence the frequency of bicycle crashes at neighborhood levels over the past two decades.

Authors	Study region	Study period	Dependent variable	Spatial units	Research method	Exposure measures		Risk factors included			
						Vehicle volume	Bicycle volume	Landuse	Road facilities	POI	Population census
Noland and Quddus (2004)	GreatBritain	1979–1998	Fatally and seriously injured cyclists	11 standard statistical regions	NB model	×	Population	×	✓	×	✓
			Slightly injured cyclists								
Priyantha Wedagama et al. (2006)	Newcastle upon Tyne, UK	1998–2001	Bicyclist injuries	185 enumeration districts	NB model	Road length	Population density	✓	✓	×	✓
Kim et al. (2010)	Honolulu, Hawaii, US	2002–2004	666 bicycle crashes	2,860uniform grids	Logistic regression model	×	Population	✓	✓	×	✓
Gladhill and Monsere (2012)	Portland, Oregon, US	2005–2007	514 bicycle crashes	792 uniform grids	NB model	Vehicle miles traveled	×	×	✓	×	✓
Siddiqui et al. (2012)	Hillsborough and Pinellas counties, Florida, US	2005–2006	1640 bicycle crashes	1,479 traffic analysis zones	PL model with CAR prior	×	Population density	×	✓	×	✓
Wei and Lovegrove (2013)	Okanagan, British Columbia, Canada	2002–2006	185 bicycle–vehicle crashes	500 traffic analysis zones	NB model	Road length	Bicycle lane length	×	✓	×	✓
Chen (2015)	Seattle, US	2010–2013	1389 bicycle–motor vehicle crashes	707 traffic analysis zones	PL model with CAR prior	Number of trips	Number of bicycle trips	✓	✓	✓	✓
Lee et al. (2015)	Orange, Seminole, and Osceola Counties, Florida, US	2008–2009	854 bicycle–motor vehicle crashes	1116 traffic analysis zones	Multivariate PL model with CAR priors	Vehicle miles traveled	Number of bicycle commuters	×	✓	×	✓
			1189 pedestrian–motor vehicle crashes								
			87,477 motor vehicle-only crashes								
Amoh-Gyimah et al. (2016)	Melbourne, Australia	2010–2012	3887 bicycle crashes	289 statistical areas	NB model	Vehicle kilometers traveled	Number of bicycle commuters	✓	×	×	✓
			1116 bicycle crashes with fatal and severe injuries		NB model with CAR prior						
			2754 bicycle crashes with minor injuries		Random-parameters NB model						
Cai et al. (2016)	Florida, US	2010–2012	15307 bicycle crashes	8518 traffic analysis zones	Bayesian dual-state model with spatial spillover	Vehicle kilometers traveled	Number of bicycle commuters	×	✓	×	✓
Prato et al. (2016)	Copenhagen Region, Denmark	2009–2013	5349 bicycle–motor vehicle crashes	269 traffic analysis zones	PL model with CAR prior	Vehicle kilometers traveled	Bicycle kilometers traveled	×	✓	×	✓

(*continued on next page*)

Table A1 (*continued*)

Authors	Study region	Study period	Dependent variable	Spatial units	Research method	Exposure measures		Risk factors included			
Yao and Loo (2016)	Hong Kong, China	2001–2003	3198 bicycle crashes	282 tertiary planning units	NB model	Vehicle kilometers traveled	Bicycle kilometers traveled	✓	✓	×	✓
		2010–2012	4144 bicycle crashes	289 tertiary planning units							
Nashad et al. (2016)	Florida, US	2010–2012	15,307 bicycle crashes 16,240 pedestrian crashes	8,518 traffic analysis zones	Copula-based bivariate NB model	Vehicle miles traveled	Number of bicycle commuters	✓	✓	×	✓
Yasmin and Eluru (2016)	Montreal Island, Canada	2006–2010	4185 bicycle crashes	837 traffic analysis zones	Latent segmentation NB model	Number of vehicles	Number of bicycle commuters	✓	✓	✓	✓
	Toronto, Canada		5457 bicycle crashes	672 traffic analysis zones							
Osama and Sayed (2016)	Vancouver, Canada	2009–2013	1703 bicycle–motor vehicle crashes	134 traffic analysis zones	PL model with CAR prior	Vehicle kilometers traveled	Bicycle kilometers traveled	×	✓	×	×
Osama and Sayed (2017)	Vancouver, Canada	2009–2013	1703 bicycle–motor vehicle crashes	134 traffic analysis zones	PL model with CAR prior	Vehicle kilometers traveled	Bicycle kilometers traveled	✓	✓	×	✓
Tasic et al. (2017)	Chicago, US	2005–2012	7632 bicycle crashes	801 census tracts	Generalized additive model	Daily vehicle miles traveled	Number of bicycle trips generated	✓	✓	✓	✓
Chen et al., (2018a)	Beijing, China	2013	bicycle–motor vehicle crashes	905 traffic analysis zones	PL model with CAR prior	Road length	Population density	✓	✓	×	×
Guo et al., 2018a	Vancouver, Canada	2009–2013	1703 bicycle–motor vehicle crashes	134 traffic analysis zones	PL model	Vehicle kilometers traveled	Bicycle kilometers traveled	✓	✓	×	✓
					PL model with CAR prior Random-parameters PL model						
Saha et al. (2018)	Florida, US	2011–2014	11,355 bicycle crashes 4315 bicycle crashes with fatal and severe injuries	11,355 census block groups	PL model with CAR prior	Daily vehicle miles traveled	Number of commuters cycling to work Bicycle ridership recorded by Strava	×	✓	×	✓
Kamel et al. (2019)	Vancouver, Canada	2009–2013	1703 bicycle–vehicle crashes	134 traffic analysis zones	PL model with CAR prior	Vehicle kilometers traveled	Bicycle kilometers traveled	✓	✓	×	✓
Ding et al. (2020)	Greater London, UK	2012–2013	2795 bicycle crashes Bicycle crashes during warm season	88 middle layer super output areas	Random-parameter NB model	Average annual daily traffic	Bicycle trips recorded by the public bicycle rental system	✓	✓	×	✓

(*continued on next page*)

Table A1 (*continued*)

Authors	Study region	Study period	Dependent variable	Spatial units	Research method	Exposure measures		Risk factors included			
							Cycling time recorded by the public bicycle rental system				
Kamel and Sayed (2020)	Vancouver, Canada	2009–2013	Bicycle crashes during cold season 1703 bicycle–vehicle crashes	134 traffic analysis zones	Bayesian measurement-error model	Vehicle kilometers traveled	Bicycle kilometers traveled	✓	✓	×	×
Lee et al. (2020)	US	2014–2016	1004 fatal crashes involving bicycles	42 metropolitan statistical areas	Bayesian PL model	×	Bicycle trips	×	×	×	✓
Sener et al. (2021)	Austin District, US	2007–2014	2038 bicycle crashes	1053 census block groups	NB model	Average daily traffic	Bicycle miles Bicycle hours Bicycle miles traveled by Strava Percent of bicycle commuters	×	✓	×	✓
Ding et al. (2021)	Greater London, UK	2015–2016	1385 bicycle crashes	270 lower super output area	NB model	Vehicle kilometers traveled	Bicycle kilometers traveled estimated by the public rental system	✓	×	×	✓
Xie et al. (2021)	Manhattan, New York City, US	2014–2016	Severity-weighted bicycle crash costs	4504 grid cells	Random-parameter Tobit model	Vehicle miles traveled	Bicycle trips estimated by the bicycle sharing program	✓	✓	×	✓
Yasmin et al. (2021)	Central Florida, US	2010	997 bicycle crashes	4747 traffic analysis zones	A framework integrating bicycle demand and bicycle crash-count models	Annual average daily traffic	Bicycle trips predicted by the travel demand model	✓	✓	✓	✓

POI: points-of-interest.
NB: negative binomial.
PL: Poisson lognormal.
CAR: conditional autoregressive.

Table A2
Cycling speeds estimated for Hong Kong residents based on the 2011 Travel Characteristics Survey data, stratified by sex and age groups (unit: m/s).

Age groups	⩽ 14	15–24	25–34	35–44	45–54	55–64	⩾ 65
Male	2.22	2.30	2.75	2.85	2.50	2.11	2.09
Female	2.22	2.34	2.20	2.44	2.38	1.93	1.59

Appendix B

One might be interested in how to include the spatially correlated error terms during the cross-validation. Let's re-formulate Eq. (4) as follows:

$$ln\left(B_i^{ori}\right) = \alpha_1 + \sum_{k=1}^{q} \alpha_k X_{ik} + v_i + \varepsilon_i \tag{B1}$$

where ε_i is assumed to follow a normal distribution with a mean of 0 and a variance of σ_b^2. Taking an exponential transformation on both sides of Eq. (B1), we have:

$$B_i^{ori} = exp\left(\alpha_1 + \sum_{k=1}^{q} \alpha_k X_{ik} + v_i + \varepsilon_i\right) \tag{B2}$$

A challenge arises when applying the parameters estimated from the training samples to predict the number of bicycle kilometers traveled for validation sample $[i]$ (i.e., $B_{[i]}^{pred}$), because it is impossible to know the exact values of $v_{[i]}$ and $\varepsilon_{[i]}$. However, if we take a Taylor series expansion near the point ($v_{[i]} = 0$, $\varepsilon_{[i]} = 0$), $B_{[i]}^{pred}$ can be closely approximated using a quadratic polynomial as follows (Xu et al., 2021):

$$\begin{aligned} B_{[i]}^{pred} \cong\ & exp(\alpha_1 + \sum_{k=1}^{q} \alpha_k X_{[i]k}) + \frac{\partial B_{[i]}^{pred}}{\partial v_{[i]}}\left(v_{[i]} - 0\right) + \frac{\partial B_{[i]}^{pred}}{\partial \varepsilon_{[i]}}\left(\varepsilon_{[i]} - 0\right) \\ & + \frac{1}{2}\frac{\partial^2 B_{[i]}^{pred}}{\partial v_{[i]}^2}\left(v_{[i]} - 0\right)^2 + \frac{1}{2}\frac{\partial^2 B_{[i]}^{pred}}{\partial \varepsilon_{[i]}^2}\left(\varepsilon_{[i]} - 0\right)^2 \end{aligned} \tag{B3}$$

where $\frac{\partial B_{[i]}^{pred}}{\partial v_{[i]}}$ and $\frac{\partial B_{[i]}^{pred}}{\partial \varepsilon_{[i]}}$ refer to the first-order partial derivative of $B_{[i]}^{pred}$ to $v_{[i]}$ and $\varepsilon_{[i]}$, respectively. $\frac{\partial^2 B_{[i]}^{pred}}{\partial v_{[i]}^2}$ and $\frac{\partial^2 B_{[i]}^{pred}}{\partial \varepsilon_{[i]}^2}$ are the corresponding second-order partial derivatives.

Taking the expectation on both sides of Eq. (B3), we have:

$$\begin{aligned} E\left(B_{[i]}^{pred}\right) &= exp(\alpha_1 + \sum_{k=1}^{q} \alpha_k X_{[i]k}) + \frac{1}{2}\frac{\partial^2 B_{[i]}^{pred}}{\partial v_{[i]}^2} var(\boldsymbol{v}) + \frac{1}{2}\frac{\partial^2 B_{[i]}^{pred}}{\partial \varepsilon_{[i]}^2}\sigma_b^2 \\ &= \left\{1 + \frac{1}{2}var(\boldsymbol{v}) + \frac{1}{2}\sigma_b^2\right\} exp(\alpha_1 + \sum_{k=1}^{q} \alpha_k X_{[i]k}) \end{aligned} \tag{B4}$$

where $var(\boldsymbol{v})$ denotes the variance of $\boldsymbol{v}$. Note that there is no closed form for $\boldsymbol{v}$, although its conditional prior distribution has been well defined by Eq. (5) (Sun et al., 1999).

Accordingly, the variance of $B_{[i]}^{pred}$ can be estimated as:

$$var\left(B_{[i]}^{pred}\right) = \left(\frac{\partial B_{[i]}^{pred}}{\partial v_{[i]}}\right)^2 var(\boldsymbol{v}) + \left(\frac{\partial B_{[i]}^{pred}}{\partial \varepsilon_{[i]}}\right)^2 \sigma_b^2 = \left\{var(\boldsymbol{v}) + \sigma_b^2\right\}\left\{exp\left(\alpha_1 + \sum_{k=1}^{q} \alpha_k X_{[i]k}\right)\right\}^2 \tag{B5}$$

Although Eq. (B4) and Eq. (B5) provide a straightforward and analytic solution to approximate the mean and variance of the prediction for validation samples, respectively, they fail to accommodate all the distributional information such as skewness and kurtosis contained in v_i and ε_i. Fortunately, based on the Markov chain Monte Carlo methods by which plausible values for the unknown quantities are synchronously simulated from their corresponding probability distribution, we can readily estimate the posterior-predictive distribution of $B_{[i]}^{pred}$ conditional on $\boldsymbol{B}^{ori}$ as follows:

$$p\left(B_{[i]}^{pred} \middle| \boldsymbol{B}^{ori}.\right) \propto p\left(B_{[i]}^{pred} \middle| \boldsymbol{\alpha}, \sigma_b, \sigma_v, \boldsymbol{X}_{[i]}, \boldsymbol{c}\right) p\left(\boldsymbol{\alpha}, \sigma_b, \sigma_v \middle| \boldsymbol{B}^{ori}, \boldsymbol{X}_i, \boldsymbol{c}\right) \tag{B6}$$

where $\boldsymbol{B}^{ori}$, $\boldsymbol{X}_{[i]}$, and $\boldsymbol{X}_i$ represent the vector of B_i^{ori}, $X_{[i]k}$, and X_{ik}, respectively. $p(B_{[i]}^{pred}|\boldsymbol{\alpha}, \sigma_b, \sigma_v, \boldsymbol{X}_{[i]}, \boldsymbol{c})$ refers to the predictive distribution of $B_{[i]}^{pred}$ conditional on $\boldsymbol{\alpha}$, σ_b, σ_v, $\boldsymbol{X}_{[i]}$, and $\boldsymbol{c}$. $p(\boldsymbol{\alpha}, \sigma_b, \sigma_v | \boldsymbol{B}^{ori}, \boldsymbol{X}_i, \boldsymbol{c})$ is the posterior distribution of $\boldsymbol{\alpha}$, σ_b, and σ_v given $\boldsymbol{B}^{ori}$, $\boldsymbol{X}_i$, and $\boldsymbol{c}$, which is estimated via:

$$p(\boldsymbol{\alpha}, \sigma_b, \sigma_v | \boldsymbol{B^{ori}}, \boldsymbol{X}_i, \boldsymbol{c}) \propto p(\boldsymbol{B^{ori}} | \boldsymbol{\alpha}, \sigma_b, \sigma_v, \boldsymbol{X}_i, \boldsymbol{c}) p(\boldsymbol{\alpha}) p(\sigma_b) p(\boldsymbol{v}) \tag{B7}$$

where $p(\boldsymbol{B^{ori}} | \boldsymbol{\alpha}, \sigma_b, \sigma_v, \boldsymbol{X}_i, \boldsymbol{c})$ is the likelihood of $\boldsymbol{B^{ori}}$, i.e., $L(\boldsymbol{\alpha}, \sigma_b, \sigma_v, \boldsymbol{X}_i, \boldsymbol{c}; \boldsymbol{B^{ori}})$.

We can then calculate the Bayesian p-value to justify whether the discrepancy between $B_{[i]}^{pred}$ and $B_{[i]}^{ori}$ is significantly greater than zero:

$$p_{Bayes} = Pr\left(B_{[i]}^{pred} \leqslant B_{[i]}^{ori} | \boldsymbol{B^{ori}}.\right) \tag{B8}$$

If $B_{[i]}^{ori}$ falls well within the 95% BCI of $B_{[i]}^{pred}$, resulting in a p_{Bayes} value ranging from 0.05 to 0.95, we can conclude that the observed value is statistically compatible with the prediction.

Appendix C

The BUGS codes for the calibration of the Bayesian simultaneous-equation model presented in Table 2:

```
Model {
for (i in 1:209) { #Missing values of LnCYC were denoted as NA, by which the BUGS will automatically generate values
  from the posterior-predictive distribution. More technical details please refer to Lunn et al. (2012, page 186)
  LnCYC[i] ~ dnorm(mu.cyc[i],tau.b) # Equation to model bicycle activities
  mu.cyc[i] <- alpha[1] + alpha[2] * CYC_Track[i] + alpha[3] * CYC_Parking[i] + alpha[4] * Bus_S[i] + alpha[5] * Edu2
  [i] + v[i]
  CYC[i] <- exp(mu.cyc[i]) # Estimated daily bicycle kilometers traveled
  CMV[i] ~ dpois(mu.cmv[i]) # Equation to model BMV crash frequencies
  log(mu.cmv[i]) <- beta[1] + beta[2] * LnVKT[i] + beta[3] * mu.cyc[i] + beta[4] * Mix_LU[i] + beta[5] * P_Mt[i] +
  beta[6] * Jun_3L[i] + beta[7] * MINC[i] + u[i] + s[i]
  u[i] ~ dnorm(0,tau.u)
  Risk[i] <- 1,000,000 * mu.cmv[i]/(CYC[i] * 741) # Risk of collisions with motor vehicles per million kilometers cycled
  on working days between 2010 and 2012
  ypred[i] ~ dpois(mu.cmv[i]) # Predicted crash frequencies based on complete data
  PPL.cmv[i] <- abs(ypred[i]-CMV[i])
  PPL.cmv2[i] <- pow(ypred[i]-CMV[i],2)
}
Sum.CYC <- sum(CYC[]) # Total "true" number of daily bicycle kilometers traveled
MAD.cmv <- mean(PPL.cmv[]) # Mean absolute deviance
MSPE.cmv <- mean(PPL.cmv2[]) # Mean squared prediction error
# Elasticity analysis. Note that contextual factors have been normalized with a mean of 0 and a standard deviation of 1
Ela.cyc.track <- exp(alpha[2]) − 1
Ela.cyc.parking <- exp(alpha[3]) − 1
Ela.cmv.track <- exp(alpha[2] * beta[3]) − 1
Ela.risk.track <- exp(alpha[2] * beta[3] - alpha[2]) − 1
Ela.cmv.parking <- exp(alpha[3] * beta[3]) − 1
Ela.risk.parking <- exp(alpha[3] * beta[3] - alpha[3]) − 1
# Prior specifications
for (i in 1:5) {alpha[i] ~ dnorm(0,1)} # Weakly informative prior suggested by Lemoine (2019)
for (k in 1:7) {beta[k] ~ dnorm(0,1)}
v[1:209] ~ car.normal(adj[], weights[], num[], tau.v) # Conditional autoregressive prior proposed by Besag et al. (1991)
s[1:209] ~ car.normal(adj[], weights[], num[], tau.s)
for (t in 1:sumNumNeigh) {weights[t] ⟨−1}
tau.v <- pow(sigma.v,-2)
var.v <- pow(sigma.v,2)
sigma.v ~ dunif(0,10)
sd.v <- sd(v[])
tau.b <- pow(sigma.b,-2)
var.b <- pow(sigma.b,2)
sigma.b ~ dunif(0,10)
P.BKT <- sd.v/(sd.v + sigma.b) # Proportion of unobserved heterogeneity due to spatially correlated effects
tau.u <- pow(sigma.u,-2)
var.u <- pow(sigma.u,2)
sigma.u ~ dunif(0,10)
tau.s <- pow(sigma.s,-2)
var.s <- pow(sigma.s,2)
sigma.s ~ dunif(0,10)
sd.s <- sd(s[])
P.BKT <- sd.s/(sd.s + sigma.u) # Proportion of unobserved heterogeneity due to spatially correlated effects
}
```

References

Afghari, A.P., Washington, S., Prato, C., Haque, M.M., 2019. Contrasting case-wise deletion with multiple imputation and latent variable approaches to dealing with missing observations in count regression models. Anal. Methods Acid. Res. 24, 100104.

Amoh-Gyimah, R., Saberi, M., Sarvi, M., 2016. Macroscopic modeling of pedestrian and bicycle crashes: a cross-comparison of estimation methods. Accid. Anal. Prev. 93, 147–159.

Bai, L., Sze, N.N., 2020. Red light running behavior of bicyclists in urban area: effects of bicycle type and bicycle group size. Travel Behav. Soc. 21, 226–234.

Bai, L., Wong, S.C., Xu, P., Chow, A., Lam, W., 2021. Calibration of stochastic link-based fundamental diagram with explicit consideration of speed heterogeneity. Transp. Res. Pt. B-Methodol. 150, 524–539.

Besag, J., York, J., Molli, E.A., 1991. Bayesian image restoration with two applications in spatial statistics. Ann. Inst. Stat. Math. 43 (1), 1–59.

Bhatia, R., Wier, M., 2011. "Safety in numbers" re-examined: can we make valid or practical inferences from available inference? Accid. Anal. Prev. 43 (1), 235–240.

Blaizot, S., Papon, F., Haddak, M.M., Amoros, E., 2013. Injury incidence rates of cyclists compared to pedestrians, car occupants and powered two-wheeler riders, using a medical registry and mobility data, Rhône County, France. Accid. Anal. Prev. 58, 35–45.

Boggs, A.M., Wali, B., Khattak, A.J., 2020. Exploratory analysis of automated vehicle crashes in California: a text analytics & hierarchical Bayesian heterogeneity-based approach. Accid. Anal. Prev. 135, 105354.

Broach, J., Dill, J., Gliebe, J., 2012. Where do cyclists ride? A route choice model developed with revealed preference GPS data. Transp. Res. Pt. A-Policy Pract. 46 (10), 1730–1740.

Brooks, S.P., Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. J. Comput. Graph. Stat. 7 (4), 434–455.

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.

Cai, Q., Abdel-Aty, M., Castro, S., 2021a. Explore effects of bicycle facilities and exposure on bicycle safety at intersections. Int. J. Sustain. Transp. 15 (8), 592–603.

Cai, Q., Abdel-Aty, M., Mahmoud, N., Ugan, J., Al-Omari, M.M.A., 2021b. Developing a grouped random parameter beta model to analyze drivers' speeding behavior on urban and suburban arterials with probe speed data. Accid. Anal. Prev. 161, 106386.

Cai, Q., Abdel-Aty, M., Sun, Y., Lee, J., Yuan, J., 2019. Applying a deep learning approach for transportation safety planning by using high-resolution transportation and land use data. Transp. Res. Part A: Policy Pract. 96, 14–28.

Cai, Q., Lee, J., Eluru, N., Abdel-Aty, M., 2016. Macro-level pedestrian and bicycle crash analysis: incorporating spatial spillover effects in dual state count models. Accid. Anal. Prev. 93, 14–22.

Chen, C., Wang, H., Roll, J., Nordback, K., Wang, Y., 2020. Using bicycle app data to develop safety performance functions (SPFs) for bicyclists at intersections: a generic framework. Transp. Res. Pt. A-Policy Pract. 132, 1034–1052.

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

Chen, P., Shen, Q., Childress, S., 2018a. A GPS data-based analysis of built environment influences on bicyclist route preferences. Int. J. Sustain. Transp. 12 (3), 218–231.

Chen, P., Sun, F., Wang, Z., Gao, X., Jiao, J., Tao, Z., 2018b. Built environment effects on bike crash frequency and risk in Beijing. J. Safety Res. 64, 135–143.

DiGioia, J., Watkins, K.E., Xu, Y., Rodgers, M., Guensler, R., 2017. Safety impacts of bicycle infrastructure: a critical review. J. Safety Res. 61, 105–119.

Dijkstra, E., 1959. A note on two problems in connection with graphs. Numer. Math. 1 (1), 269–271.

Ding, H., Sze, N.N., Li, H., Guo, Y., 2020. Roles of infrastructure and land use in bicycle crash exposure and frequency: a case study using Greater London bike sharing data. Accid. Anal. Prev. 144, 105652.

Ding, H., Sze, N.N., Guo, Y., Li, H., 2021. Role of exposure in bicycle safety analysis: effect of cycle path choice. Accid. Anal. Prev. 153, 106014.

Dong, N., Meng, F., Zhang, J., Wong, S.C., Xu, P., 2020. Towards activity-based exposure measures in spatial analysis of pedestrian–motor vehicle crashes. Accid. Anal. Prev. 148, 105777.

Elvik, R., Bjørnskau, T., 2017. Safety-in-numbers: a systematic review and meta-analysis of evidence. Saf. Sci. 92, 274–282.

Elvik, R., Goel, R., 2019. Safety-in-numbers: an updated meta-analysis of estimates. Accid. Anal. Prev. 129, 136–147.

Feleke, R., Scholes, S., Wardlaw, M., Mindell, J.S., 2018. Comparative fatality risk for different travel modes by age, sex, and deprivation. J. Transp. Health 8, 307–320.

Fridman, L., Rothman, L., Howard, A.W., Hagel, B.E., Macarthur, C., 2021. Methodological considerations in MVC epidemiological research. Inj. Prev. 27, 155–160.

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

Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models. Bayesian Anal. 1, 515–533.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2013. Bayesian Data Analysis, 3rd ed.

Gelman, A., Simpson, D., Betancourt, M., 2017. The prior can often only be understood in the context of the likelihood. Entropy 19, 555.

Gladhill, K., Monsere, C.M., 2012. Exploring traffic safety and urban form in Portland, Oregon. Transp. Res. Rec. 2318, 63–74.

Guo, Q., Xu, P., Pei, X., Wong, S.C., Yao, D., 2017. The effect of road network patterns on pedestrian safety: a zone-based Bayesian spatial modeling approach. Accid. Anal. Prev. 99, 114–124.

Guo, Y., Osama, A., Sayed, T., 2018a. A cross-comparison of different techniques for modeling macro-level cyclist crashes. Accid. Anal. Prev. 113, 38–46.

Guo, Y., Li, Z., Wu, Y., Xu, C., 2018b. Exploring unobserved heterogeneity in bicyclists' red-light running behaviors at different crossing facilities. Accid. Anal. Prev. 115, 118–127.

Guo, Y., Yang, L., Lu, Y., Zhao, R., 2021. Dockless bike-sharing as a feeder model of metro commute? The role of the feeder-related built environment: analytical framework and empirical evidence. Sustain. Cities Soc. 65, 102594.

Handy, S., Wee, B.V., Kroesen, A.M., 2014. Promoting cycling for transport: research needs and challenges. Transp. Rev. 34 (1), 4–24.

Hong Kong Transport Department, 2012. The Annual Traffic Census 2011. https://www.td.gov.hk/en/publications_and_press_releases/publications/free_publications/the_annual_traffic_census_2011/index.html.

Hong Kong Transport Department, 2014. Travel Characteristics Survey 2011 Final Report. https://www.td.gov.hk/filemanager/en/content_4652/tcs2011_eng.pdf.

Imprialou, M., Quddus, M., 2019. Crash data quality for road safety research: current state and future directions. Accid. Anal. Prev. 130, 84–90.

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

Jacobsen, P.L., Ragland, D.R., Komanoff, C., 2015. Safety in numbers for walkers and bicyclists: exploring the mechanisms. Inj. Prev. 21 (4), 217–220.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning. Springer Press, NY.

Kamel, M.B., Sayed, T., 2020. Cyclist-vehicle crash modeling with measurement error in traffic exposure. Accid. Anal. Prev. 144, 105612.

Kamel, M.B., Sayed, T., Bigazzi, A., 2020. A composite zonal index for biking attractiveness and safety. Accid. Anal. Prev. 137, 105439.

Kamel, M.B., Sayed, T., Osama, A., 2019. Accounting for mediation in cyclist-vehicle crash models: a Bayesian mediation analysis approach. Accid. Anal. Prev. 131, 122–130.

Kim, J.H., Mooney, S.J., 2016. The epidemiologic principles underlying traffic safety study designs. Int. J. Epidemiol. 45 (5), 1668–1675.

Kim, K., Pant, P., Yamashita, E., 2010. Accidents and accessibility: measuring influences of demographic and land use variables in Honolulu, Hawaii. Transport. Res. Rec. 2673 (4), 898–906.

Klassen, J., El-Basyouny, K., Islam, M.T., 2014. Analyzing the severity of bicycle-motor vehicle collision using spatial mixed logit models: a city of Edmonton case study. Saf. Sci. 62, 295–304.

Klieger, J., Savage, I., 2020. Motor-vehicle drivers' behavioral response to increased bicycle traffic. J. Safety Res. 74, 97–102.

Lee, J., Abdel-Aty, M., Cai, Q., 2020. Investigation of safety-in-numbers for pedestrians and bicyclists at a macroscopic level with various exposure variables. Transp. Res. Rec. 2674 (12), 568–580.

Lee, J., Abdel-Aty, M., Huang, H., Cai, Q., 2019. Transportation safety planning approach for pedestrians: an integrated framework of modeling walking duration and pedestrian fatalities. Transp. Res. Rec. 2637 (4), 898–906.

Lee, J., Abdel-Aty, M., Jiang, X., 2015. Multivariate crash modeling for motor vehicle and non-motorized modes at the macroscopic level. Accid. Anal. Prev. 78, 146–154.

Lee, K., Senser, I.N., 2021. Strava metro data for bicycle monitoring: a literature review. Transp. Rev. 41 (1), 27–47.

Lemoine, N.P., 2019. Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. Oikos 128, 912–928.

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.

Loo, B.P.Y., 2006. Validating crash locations for quantitative spatial analysis: a GIS-based approach. Accid. Anal. Prev. 38 (5), 879–886.

Loo, B.P.Y., Tsui, K.L., 2010. Bicycle crash casualties in a highly motorized city. Accid. Anal. Prev. 42, 1902–1907.

Lusk, A.C., Furth, P.G., Morency, P., Miranda-Moreno, L.F., Willettt, W.C., Dennerlein, J. T., 2011. Risk of injury for bicycling on cycle tracks versus in the street. Inj. Prev. 17 (2), 131–135.

Lunn, D., Jackson, C., Best, N., Thomas, A., Spiegelhalter, D., 2012. The BUGS Book: A Practical Introduction to Bayesian Analysis.

Merlin, L.A., Guerra, E., Dumbaugh, E., 2020. Crash risk, crash exposure, and the built environment: a conceptual review. Accid. Anal. Prev. 134, 105244.

McElreath, R., 2020. Statistical Rethinking: A Bayesian Course with Examples in R and Stan (Second Edition). Chapman & Hall/CRC: Boca Raton, FL.

Miranda-Moreno, L.F., Morency, P., El-Geneidy, A.M., 2011. The link between built environment, pedestrian activity and pedestrian–vehicle collision occurrence at signalized intersections. Accid. Anal. Prev. 43, 1624–1634.

Moran, P.A.P., 1950. Notes on continuous stochastic phenomena. Biometrika 37 (1), 17–23.

Nashad, T., Yasmin, S., Eluru, N., Lee, J., Abdel-Aty, M., 2016. Joint modeling of pedestrian and bicycle crashes: copula-based approach. Transp. Res. Rec. 2601, 119–127.

Noland, R.B., Quddus, M.A., 2004. Analysis of pedestrian and bicycle casualties with regional panel data. Transp. Res. Rec. 1897, 28–33.

Nordback, K., Marshall, W.E., Janson, B.N., 2014. Bicyclist safety performance functions for a U.S. city. Accid. Anal. Prev. 65, 114–122.

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

Osama, A., Sayed, T., 2017. 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., Bigazzi, A.Y., 2017. Models for estimating zone-level bike kilometers traveled using bike network, land use, and road facilities variables. Transp. Res. Part A: Policy Pract. 96, 14–28.
Park, E.S., Fitzpatrick, K., Das, S., Avelar, R., 2021. Exploration of the relationship among roadway characteristics, operating speed, and crashes for city streets using path analysis. Accid. Anal. Prev. 150, 105896.
Prati, G., Puchades, V.M., Angelis, M.D., Fraboni, F., Pietrantoni, L., 2018. Factors contributing to bicycle–motor vehicle collisions: a systematic literature review. Transp. Rev. 38 (2), 184–208.
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.
Priyantha Wedagama, D.M., Bird, R.N., Metcalfe, A.V., 2006. The influence of urban land-use on non-motorised transport casualties. Accid. Anal. Prev. 38 (6), 1049–1057.
Pucher, J., Dill, J., Handy, S., 2010. Infrastructure, programs, and policies to increase bicycling: an international review. Prev. Med. 50, S106–S125.
Pulugurtha, S.S., Thakur, V., 2015. Evaluating the effectiveness of on-street bicycle lane and assessing risk to bicyclists in Charlotte, North Carolina. Accid. Anal. Prev. 76, 34–41.
Reynolds, C.C., Harris, M.A., Teschke, K., Cripton, P.A., Winters, M., 2009. The impact of transportation infrastructure on bicycling injuries and crashes: a review of the literature. Environ. Health 8, 47.
Rubin, D.B., 1976. Inference and missing data. Biometrika 63, 581–592.
Saad, M., Abdel-Aty, M., Lee, J., Cai, Q., 2019. Bicycle safety analysis at intersections from crowdsourced data. Transp. Res. Rec. 2673 (4), 1–14.
Saha, D., Alluri, P., Gan, A., Wu, W., 2018. Spatial analysis of macro-level bicycle crashes using the class of conditional autoregressive models. Accid. Anal. Prev. 118, 166–177.
Sener, I.N., Lee, K., Hudson, J.G., Martin, M., Dai, B., 2021. The challenge of safe and active transportation: macrolevel examination of pedestrian and bicycle crashes in the Austin District. J. Transp. Saf. Secur. 13 (5), 525–551.
Schafer, J.L., Graham, J.W., 2002. Missing data: our view of the state of the art. Psychol. Methods 7 (2), 147–177.
Schepers, J.P., Kroeze, P.A., Sweers, W., Wüst, J.C., 2011. Road factors and bicycle–motor vehicle crashes at unsignalized priority intersections. Accid. Anal. Prev. 43, 853–861.
Schepers, P., 2012. Does more cycling also reduce the risk of single-vehicle crashes? Inj. Prev. 18, 240–245.
Schoner, J.E., Levinson, D.M., 2014. The missing link: bicycle infrastructure networks and ridership in 74 US cities. Transportation 41 (6), 1187–1204.
Shirani-bidabadi, N., Mallipaddi, N., Haleem, K., Anderson, M., 2020. Developing bicycle-vehicle crash-specific safety performance functions in Alabama using different techniques. Accid. Anal. Prev. 146, 105735.
Siddiqui, C., Abdel-Aty, M., Choi, K., 2012. Macroscopic spatial analysis of pedestrian and bicycle crashes. Accid. Anal. Prev. 45, 382–391.
Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. J. R. Stat. Soc. Series B Stat. Methodol. 64 (4), 583–639.
Spiegelhalter, D.J., Thomas, A., Best, N., Lunn, D., 2005. WinBUGS User Manual. MRC Biostatistics Unit, Cambridge, UK.
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.
Strauss, J., Miranda-Moreno, L.F., Morency, P., 2015. Mapping cyclist activity and injury risk in a network combining smartphone GPS data and bicycle counts. Accid. Anal. Prev. 83, 132–142.
Sterne, J.A.C., White, I.R., Carlin, J.B., Spratt, M., Royston, P., Kenward, M.G., Wood, A. M., Carpenter, J.R., 2009. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 338, b2393.
Su, J., Sze, N.N., Bai, L., 2021. A joint probability model for pedestrian crashes at macroscopic level: roles of environment, traffic, and population characteristics. Accid. Anal. Prev. 150, 105898.
Sun, D., Tsutakawa, R., Speckman, P.L., 1999. Bayesian inference for CAR(1) models with noninformative priors. Biometrika 86 (2), 341–350.
Sze, N.N., Su, J., Bai, L., 2019. Exposure to pedestrian crash based on household survey data: effect of trip purpose. Accid. Anal. Prev. 128, 17–24.
Tasic, I., Elvik, R., Brewer, S., 2017. Exploring the safety in numbers effect for vulnerable road users on a macroscopic scale. Accid. Anal. Prev. 109, 36–46.
Thompson, J.H., Wijnands, J.S., Mavoa, S., Scully, K., Stevenson, M.R., 2019. Evidence for the 'safety in density' effect for cyclists: validation of agent-based modelling results. Inj. Prev. 25, 379–385.
van Smeden, M., de Vries, B.B.L.P., Nab, L., Groenwold, R.H.H., 2021. Approaches to addressing missing values, measurement error, and confounding in epidemiological studies. J. Clin. Epidemiol. 131, 89–100.
Vanparijs, J., Panis, L.I., Meeusen, R., de Geus, B., 2015. Exposure measurement in bicycle safety analysis: a review of the literature. Accid. Anal. Prev. 84, 9–19.
Wei, F., Lovegrove, G., 2013. An empirical tool to evaluate the safety of cyclists: community based, macro-level collision prediction models using negative binomial regression. Accid. Anal. Prev. 61, 129–137.
Wen, H., Zhang, X., Zeng, Q., Sze, N., 2019. Bayesian spatio-temporal model for the main and interaction effects of roadway and weather characteristics on freeway crash incidence. Accid. Anal. Prev. 132, 105249.
Xie, K., Ozbay, K., Yang, D., Xu, C., Yang, H., 2021. Modeling bicycle crash costs using big data: a grid-cell-based Tobit model with random parameters. J. Transp. Geogr. 91, 102953.
Xie, S.Q., Dong, N., Wong, S.C., Huang, H., Xu, P., 2018. Bayesian approach to model pedestrian crashes at signalized intersections with measurement errors in exposure. Accid. Anal. Prev. 121, 285–294.
Xu, P., Dong, N., Wong, S.C., Huang, H., 2019a. Cyclists injured in traffic crashes in Hong Kong: a call for action. PLoS One 14 (8), e0220785.
Xu, P., Huang, H., Dong, N., Wong, S.C., 2017. Revisiting crash spatial heterogeneity: a Bayesian spatially varying coefficients approach. Accid. Anal. Prev. 98, 330–337.
Xu, P., Xie, S., Dong, N., Wong, S.C., Huang, H., 2019b. Rethinking safety in numbers: are intersections with more crossing pedestrians really safer? Inj. Prev. 25 (1), 20–25.
Xu, P., Zhou, H., Wong, S.C., 2021. On random-parameter count models for out-of-sample crash prediction: accounting for the variances of random-parameter distributions. Accid. Anal. Prev. 159, 106237.
Yao, S., Loo, B.P.Y., Lam, W.W.Y., 2015. Measures of activity-based pedestrian exposure to the risk of vehicle-pedestrian collisions: space-time vs. potential path tree methods. Accid. Anal. Prev. 75, 320–332.
Yao, S., Loo, B.P.Y., 2016. Safety in numbers for cyclists beyond national-level and city level data: a study on the non-linearity of risk within the city of Hong Kong. Inj. Prev. 22 (6), 379–395.
Yasmin, S., Bhowmik, T., Rahman, M., Eluru, N., 2021. Enhancing non-motorist safety by simulating trip exposure using a transportation planning approach. Accid. Anal. Prev. 156, 106128.
Yasmin, S., Eluru, N., 2016. Latent segmentation based count models: analysis of bicycle safety in Montreal and Toronto. Accid. Anal. Prev. 95, 157–171.
Ye, Y., Wong, S.C., Meng, F., Xu, P., 2021. Right-looking habit and maladaptation of pedestrians in areas with unfamiliar driving rules. Accid. Anal. Prev. 150, 105921.
Zeng, Q., Xu, P., Wang, X., Wen, H., Hao, W., 2021. Applying a Bayesian multivariate spatio-temporal interaction model based approach to rank sites with promise using severity-weighted decision parameters. Accid. Anal. Prev. 95, 106190.
Zhou, H., Yuan, C., Dong, N., Wong, S.C., Xu, P., 2020. Severity of passenger injuries on public buses: a comparative analysis of collision injuries and non-collision injuries. J. Safety Res. 74, 55–69.
Ziakopoulos, A., Yannis, G., 2020. A review of spatial approaches in road safety. Accid. Anal. Prev. 135, 105323.
Zimmermann, M., Mai, T., Frejinger, E., 2017. Bike route choice modeling using GPS data without choice sets of paths. Transp. Res. Pt. C-Emerg. Technol. 75, 183–196.