Modeling Bicycle Crash Costs Using a Grid-Cell-Based Random Parameters Tobit Model

Kun Xie, Kaan Ozbay, Di Yang, Chuan Xu, Hong Yang

This is the author's version of a work that was presented at the Transportation Research Board 97th Annual Meeting. The final version can be found on the [Transportation Research Board website]().

MODELING BICYCLE CRASH COSTS USING A GRID-CELL-BASED RANDOM PARAMETERS TOBIT MODEL

Kun Xie, Ph.D.
Postdoctoral Associate, C^2SMART Center
Department of Civil and Urban Engineering
Center for Urban Science and Progress (CUSP)
Tandon School of Engineering
New York University
1 MetroTech Center, 19th Floor, Brooklyn, NY 11201, USA
Phone: +1-646-997-0547, E-mail: kun.xie@nyu.edu

Kaan Ozbay, Ph.D.
Professor & Director, C^2SMART Center
Department of Civil and Urban Engineering &
Center for Urban Science and Progress (CUSP)
Tandon School of Engineering
New York University (NYU)
6 MetroTech Center, 4th Floor, Brooklyn, NY 11201, USA
Phone: +1-646-997-0552, E-mail: kaan.ozbay@nyu.edu

Di Yang, Ph.D. Candidate
Graduate Research Assistant, C^2SMART Center
Department of Civil and Urban Engineering
Center for Urban Science and Progress (CUSP)
Tandon School of Engineering
New York University
1 MetroTech Center, Brooklyn, 19th Floor, NY 11201, USA
Phone: +1-646-997-0548, E-mail: dy855@nyu.edu

Chuan Xu, Ph.D. (Corresponding author)
Assistant Professor
School of Transportation and Logistics
National United Engineering Laboratory of Integrated and Intelligent Transportation
Southwest Jiaotong University
Visiting Postdoctoral Researcher at New York University
No. 111 of 1st North Section 2nd Ring Rd., Chengdu, Sichuan, China
Phone: +1-646-997-0559, Email: xuchuan@swjtu.edu.cn

Hong Yang, Ph.D.
Assistant Professor
Department of Modeling, Simulation & Visualization Engineering
Old Dominion University
4700 Elkhorn Ave, Norfolk, VA 23529, USA
Phone: +1-757-683-4529, E-mail: hyang@odu.edu

Word Count: 5500 (Text) + 3 Figures + 5 Tables=7500 words; Abstract: 282
Submission Date: August. 1, 2017

Prepared for Possible Presentation at the 97th TRB Annual Meeting

ABSTRACT
Bicyclists are among the most vulnerable road users in the urban transportation systems. It is critical to investigate the contributing factors to bicycle-related crashes and to identify the hotspots for efficient allocation of treatment resources. A grid-cell-based modeling framework was used to explore the overall safety patterns of bicyclists in Manhattan. A random parameters (RP) Tobit model was developed in the Bayesian framework to correlate transportation, land use, sociodemographic and social media data with bicycle crash costs. It is worth to mention that large-scale bicycle ridership data from 2014 to 2016 was obtained from Citi Bike, which is the largest bike sharing program in the United States, and used for model development. The proposed RP Tobit model could deal with left-censored crash cost data and was found to outperform the Tobit model by accounting for the unobserved heterogeneity across neighborhoods. Results indicated that bicycle ridership, bicycle rack density, subway ridership, taxi trip, bus stop density, population, and ratio of population under 14 were positively associated with bicycle crash cost, whereas residential ratio and median age had negative impact on bicycle crash cost. The RP Tobit model was used to estimate the cell-specific potential for safety improvement (PSI) for hotspot ranking. The advantages of using the RP Tobit crash cost model to capture PSI lie in: 1) injury severity is considered by being converted into unit costs, and 2) varying effects of certain explanatory variables are accounted for by incorporating random parameters. The cell-based hotspot identification method can provide a complete risk map for bicyclists with high resolutions. It was found that most locations with high PSIs either had unprotected bicycle lanes or were close to the access points to protected bicycle routes.

Keywords: Safety Analysis, Big Data, Bicycle Crashes, Hotspot Identification, Random Parameters Tobit Model

INTRODUCTION

Bicycling is a green travel mode without fuel consumption and emission, and can yield enormous public health benefits. The above features make bicycling a highly-advocated way of traveling in urban areas. Bicycling ridership has grown dramatically in the cities of United States recently. New York City (NYC) has witnessed 80% growth in daily bicycling ridership from 2010 to 2015, with the bike network undergoing unprecedented enhancement *(1)*. However, bicyclists are among the most vulnerable road users in the urban transportation system, and are prone to higher risk of injuries and fatalities when involved in traffic crashes compared with drivers and passengers in the vehicles. According to Traffic Safety Facts 2015 *(2)*, 70% of bicyclist fatalities occurred in urban areas, as opposed to 30% in rural areas. In United States, bicycle trips account for 1% of all trips but more than 2% of the road deaths *(3)*. According to historical crash data of Manhattan, NYC from 2005 to 2012, approximately 9.4% of bicycle crashes were involved with serious injuries and fatalities, while the proportion of serious injuries and fatalities was merely 5.0% for all crashes events. To enhance bicycling safety and enlarge its related benefits, it is critical to investigate the contributing factors to bicycle crashes and to identify the hotspots for efficient allocation of treatment resources.

Our previous study *(4)* proposes a grid-cell-based Tobit model for the risk analysis of pedestrians. The cell-structured framework enables high-resolution large-scale hotspot identification and provides the convenience to incorporate a variety of datasets into safety modeling, including taxi trips, subway ridership and social media. Additional benefits of using grid cells as units of analysis over the traditional methods based on facilities (e.g., intersections and road segments) include: 1) there is no need to decide whether crashes are intersection-related or road segment-related, which can be a complicated process *(5)*, especially in cities with high-density networks like Manhattan; and 2) there is no need to conduct road segmentation (i.e., splitting roadways at points where geometric and traffic characteristics change), which could be subjective and costly. Pedestrian risks are represented by crash costs in this study *(4)*, so that both crash frequency and injury severity can be taken into account.

This study adopts the cell-based crash cost modeling framework proposed in *(4)* to explore the overall safety patterns of bicyclists in Manhattan. A massive amount of data regarding crashes, transportation, land use, sociodemographic and social media in Manhattan are collected for model development. Citi Bike, officially opened in May 2013 in NYC, is the largest bike sharing program in the United States *(6)*. As of March 31, 2016, Citi Bike has 163,865 annual subscribers, 10,000 bicycles from 603 stations, and an average of 38,491 rides per day in 2016 *(6)*. The Citi Bike ridership data provides a great indicator to the exposure of bicyclists in Manhattan, which is essential to develop a proper crash cost model. In addition, this study advances the previous Tobit crash cost model by including random parameters to tackle the unobserved heterogeneity across different neighborhood. More reliable inferences can be made by accounting for varying effects of certain explanatory variables.

LITERATURE REVIEW

Literatures on bicyclist safety are reviewed, with focuses on statistical modeling, contribution factors and units of analysis, each of which is discussed in one subsection below. Table 1 summarizes the response variables, units of analysis, sample sizes, methodology and key contributing factors used in previous studies.

Table 1 Previous Studies on Cyclist Crash Models

Study	Response Variable	Dataset	Location	Methodology	Key Contributing Factors
Kim et al. (2010)[7]	Crash frequency	Gird cells (N=2,860)	United States (Honolulu)	Binary logistic regression	Road Characteristic: Bus stops, bus route length, number of intersections Sociodemographic: Population, job count, population below poverty line Land Use: Business, commercial mixed, high-density residential, low-density residential, military
Siddiqui et al.(2012)[8]	Crash frequency	Traffic analysis zones (N=1479)	United States (Florida)	Poisson-lognormal model with conditional autoregressive effects	Road Characteristic: Length of road with speed limit of 15 mph, length of road with speed limit of 35 mph, number of Intersection Sociodemographic: Median household income, dwelling unit, hotel unit, population density, percentage of households that has non-retired workers and zero automobile, percentage of households that has non-retired workers and one automobile, kindergarten through 12th grade school enrollment, long term parking cost, total employment
Wei and Lovegrove (2013)[9]	Crash frequency	Traffic analysis zones (N=500)	Canada (British Columbia)	Negative binomial model	Traffic Characteristic: Number of drive commuters, percentage of drive commuters. Road Characteristic: Total lane kilometers, bicycle lane kilometers, bus stops, traffic signals, intersection density, arterial–local intersection percentage
Strauss et al., (2013)[10]	Injury frequency	Intersections (N=647)	Canada (Montreal, Quebec)	Bivariate mixed Poisson mode with correlated lognormal error terms	Traffic Characteristic: Bicycle flows, motor vehicle flow, right turn flows, left turn flows Road Characteristic: Presence of bus stops, total crosswalk length, presence of raised median
Nordback et al., (2014)[3]	Crash frequency	Intersections (N=211)	United States (Boulder, Colorado)	Negative binomial	Traffic Characteristics: Annual average daily traffic (AADT), annual average daily bicyclist (AADB)
Zhang et al., (2015)[11]	Crash frequency	Census tracts (N=321)	United States (California, Alameda County)	Geographically weighted regression	Road Characteristic: Network betweenness centrality, overall clustering coefficient
Kaplan and Giacomo Prato, (2015)[12]	Crash Frequency and Severity	Road Links (N=272586)	Denmark (Copenhagen)	Multivariate Poisson-lognormal model with correlated	Traffic Characteristics: Bicycle kilometer travelled, car kilometer travelled, van kilometer travelled, heavy vehicle kilometer travelled

				autoregressive priors	Road Characteristic: Bicycle lane, bicycle path, one-way, give-way/stop, roundabout, traffic light Land Use: City center, high residential area, low residential area, industrial area
Chen, (2015)[13]	Crash frequency	Traffic analysis zones (N=707)	United States (Seattle)	Area-based Poisson lognormal random effects model	Traffic Characteristic: Total number of trips Road Characteristic: Number of 3-way intersections per hectare, length of on-arterial bike lanes per hectare, length of off-arterial bike lanes per hectare, number of parking signs per hectare, zonal mean of driving speed limits, number of traffic signals per hectare Land use: Entropy of mixing land use
Park et al, (2015)[14]	Crash frequency	Roadway segments (N=227)	United States (Florida)	Negative binomial model	Traffic Characteristics: AADT per lane, AADT Road Characteristics: Median width, bike lane width Demographic: Population density
Osama and Sayed, (2016)[15]	Crash frequency	Traffic analysis zones (N=134)	Canada (Vancouver)	Negative binomial model	Traffic Characteristics: Vehicle kilometer travelled (VKT), bike kilometer travelled (BKT) Road Characteristics: Intersections density, degree of bike network connectivity, average edge length, linearity, total length of bike network links, Average weighted slope of the bike network
Li et al, (2017)[16]	Crash frequency	Road segments (N=375)	United Kingdom (London)	Negative binomial model	Traffic Characteristics: AADT, AADB, speed Sociodemographic: Percentage of non-domestic buildings, index of multiple deprivation Road Characteristic: Road class, road type, density of Intersections

Statistical Modeling

According to the response variables, the previous studies on bicyclist safety mainly fall into two categories, crash severity studies and crash frequency studies. Statistical models and units of analysis differ greatly between those two categories.

Crash severity is usually treated as a categorical variable, and thus logistic regression model and its extensions are applied. Some researcher developed ordered logit models to address the ordered nature of crash severity and *(17-20)*. Generalized additive model *(19)* and spatial mixed logit model *(21)* were used to account for the spatial dependence among crashes sites. Some research recognized that the assumption that a monotonous effect of the independent variables on the dependent variables implied in ordered models is questionable and multinomial logit models using unordered response variable were built *(22)*. In addition to statistical models, machine learning methods including decision tree and Bayesian network analysis *(23)* were introduced to bicyclist-related crashes analysis, to overcome the shortages of regression models such as strong statistical assumptions, interactions among variables, and not satisfactorily handling variables with too many categories.

To capture the relationship between bicycle crash frequencies and contribution factors, power function and least square analysis *(24)*, linear regression *(25)* and general linear regression *(11)* methods were used. With the ability to accommodate the random, discrete, non-negative and sporadic nature of crash frequencies, Poisson-gamma *(3, 9, 14)* and Poisson-lognormal models *(8, 12, 13)* have been widely used to model bicyclist related crashes frequencies. To address the spatial correlation of the bicycle crash data, the model with correlated autoregressive priors was established *(12)*. In addition, Bayesian method *(8, 13)* was introduced to improve the model performance, and hierarchy structure *(13)* was proposed to account for the unobserved heterogeneity.

Only a few studies on bicyclist safety incorporated both crash severity and crash frequency into modeling. Kaplan and Prato *(12)* developed the multivariate Poisson-lognormal model to jointly model bicycle crash frequency and severity. But one disadvantage of the multivariate model is that crashes of certain severity levels (e.g., fatal crashes) could be rare and thus it would lead to unreliable estimation. A possible solution is to merge different severity levels, for instance, fatal and seriously injured crashes. But treating crashes of different severity levels equally would lead to information loss. Xie et al. (2017) *(4)* used crash costs, which were obtained from severity-weighted crash frequencies, to quantify crash risks. Crash cost is a continuous non-negative variable and could not fit in Poisson-based models. Tobit models *(26)* , which could accommodate censored, continuous dependent variables, is an applicable method, and it has been applied in modeling crash rate *(27, 28)* and crash cost *(4)*.

Contributing Factors

Contributing factors to bicycle crashes investigated in the previous studies fall into four categories: traffic characteristics, road characteristics, sociodemographic features, and land use.

The most essential contributing factors were recognized as bicyclist exposure indicators such as annual average daily bicycle volume (AADB) *(3, 16)*, bike kilometer travelled (BKT) *(15)* and their mathematical transformation *(12)*. Most bicyclist crashes were involved with motor vehicles, and thus annual average daily traffic (AADT) *(3, 10, 12, 14, 16)*, vehicle kilometer travelled *(15)* and their mathematical transformation were also found to be positively associated with bicycle crashes. From a different perspective, one review study *(29)* made a conclusion that with sufficient bicyclist users presented, bicyclist related crashes would decrease, which is called the “safety-in-numbers” effect . Other traffic characteristics like heavy vehicle kilometer travelled *(12)*, total number of trips *(13)* and speed *(16)* were also found to affect bicyclist related crashes.

Researchers paid much attention to road characteristics that affected bicyclist safety, since safety issues could be addressed by modifying the road design. Many studies focused on evaluating the impact of bicycle lanes on bicyclist safety *(16, 30)*. Contradicting findings on the impact of bicycle lanes *(9, 12)* suggested more work on this topic need to be done. Bicycle lane markings across the intersection *(30)* was found to improve bicyclist safety. Other road characteristics investigated included number/density of intersections *(7-9, 13, 15, 16)*, number of bus stops *(7, 9, 10)*, speed limit *(8)*, traffic signals *(9, 12)*. Additionally, some studies found road network structure indicators such as accessibility *(7)*, network betweenness, centrality and overall clustering coefficient *(11)* affected bicyclist safety and those findings could provide insights into road network planning.

Sociodemographic characteristics affecting bicyclist safety included population/population density *(7, 8, 14)*, income *(7, 16)*, and vehicle ownership *(8, 16)*. Regarding land use, Kim et al. *(7)* found the proportions of business, commercial mixed, high-density residential land were positively related to bicyclist related crashes, while the proportion of military land had the opposite effect.

Kaplan and Giacomo *(12)* found that the city center, industrial areas had more bicycle crashes and residential areas had fewer bicycle crashes. Chen *(13)* found the entropy of mixing land were positively associated with bicycle crash frequency.

Units of Analysis

Crash frequency models can be divided into two categories according to units of analysis, facility-based and zone-based. The facility-based models were usually developed based on transportation facilities such as intersections *(3, 10)* and roadway segments *(12, 14, 16)*, while zone-based models used units of analysis such as traffic analysis zones (TAZs) *(8, 9, 13, 15)*, census tracts *(11)*, and grid cells *(7)*.

Facility-based models can be used to investigate intersection/segment-related contributing factors directly, especially the traffic and geometrical design features. For instance, Strauss *(10)* found bicycle crashes at intersections were related to turning traffic flow volume, which is a feature hard to be included in a zone-based models. Park et al. *(14)* found median width and bike lane width of roadway segments had impact on bicycle crash frequencies. For zone-based models, aggregation needs to be carried out to capture the traffic, road, sociodemographic and land use features. Proportion/density *(13)*, mean *(16)* and summation *(8, 9, 15)* are widely used aggregation methods. Kim et al. (2010) *(7)* and Xie et al. (2017) *(4)* developed grid-cell-based framework, which provided a better way to include crashes on the boundaries of TAZs and census tracts and higher resolution for safety analysis.

DATA PREPARATION

Manhattan area below 116th Street is selected as our study area, since that is the region served by Citi Bike in Manhattan. The map of the study area was uniformly divided into a total of 4504 cells with the size of 300 feet by 300 feet.

We obtained three-year traffic crash data (2014-2016) from the NYC Open Data website[1]. A total of 4,667 crashes that got cyclists involved were identified. According to their injury severity, crashes were categorized into three types: no injury, injury and fatality.

Bicycle usage related data in Manhattan includes bicycle routes data, bicycle parking data, and Citi Bike trip data. The geographic information system (GIS) data of bicycle routes and bicycle parking were obtained from the New York City Department of Transportation (NYCDOT)[2]. There are 14,980 bicycle route segments in total and 2,649 segments of them installed between 2014 and 2016, which were not included in our previous study *(4)*. The length of bicycle routes within each cell was obtained by splitting bicycle routes at cell boundaries and then summing the split route segments by cells. There are three types of bicycle parking facilities, namely CityRacks, Bike Corrals, and Sheltered Bike Parking. CityRacks are individual bicycle racks located on sidewalks, while Bike Corrals and Sheltered Bike Parking facilities are assemblies of bicycle racks located in the curbside lane of the streets and sidewalks, respectively. The capacity of bicycle parking facilities can be represented by number of bicycle racks.

Citi Bike trip data from 2014 to 2016 was obtained from Citi Bike website[3]. The duration, origin, destination and attributes of the users are recorded and available to the public. It processes rich information regarding bicycle movements in NYC and is used in this study to estimate bicycle

---

[1] Source: https://opendata.cityofnewyork.us/
[2] Source: http://www.nyc.gov/html/dot/html/about/datafeeds.shtml#bikes
[3] Source: https://www.citibikenyc.com/system-data

ridership for each cell in the studied area in Manhattan. Detailed method in processing bicycle parking data and Citi Bike data are presented in the following subsection.

Traffic volume data in 2015 was obtained from NYSDOT[4] and vehicle miles traveled (VMT) for each grid cell was computed based on the method presented in *(4)* using this data. New York City yellow taxi data from 2014 to 2016 was obtained from New York City Taxi & Limousine Commission[5] (NYCTL). It contains the pick-up and drop-off coordinates of each trip and number of pick-ups and drop-offs of each cell could be obtained. The annual subway ridership for each station in Manhattan from 2014 to 2016 was obtained from the Metropolitan Transportation Authority[6] (MTA). The process of obtaining subway ridership per cell is presented in the following subsection.

Spatial Processing

The kernel density tool in ArcGIS 9.3 was employed to distribute the cost of each crash, the subway ridership of each station, and the effects of each bicycle rack and each bus stop, regarding the fact that attributes of one cell may have impact on its neighboring cells *(4)*. The value assigned to each cell is expressed in equation (1).

$$RC(s) = \sum_{i=1}^{n} \rho \left[1 - (\frac{d_{is}}{r})^2 \right]^2 C_i \tag{1}$$

$RC(s)$ is the value assigned to the raster cell s. C_i is the value possessed by point i. C_i is the unit crash cost for each bicycle crash, the ridership for each subway station, the number of racks for each bike parking station and one for each bus stop. The unit crash costs for no injury and fatality crashes were available from a report of National Safety Council *(31)*. The unit crash cost for injury crashes was computed based on the proportions of incapacitating injury, non-incapacitating injury and possible injury crashes from historical record. d_{is} is the distance from the location s to the point i, and r is the search radius (or bandwidth). For more details on the spatial processing method, please refer to our previous study *(4)*.

Citi Bike Trip Data

Data regarding the Citi Bike trips originated from the study area, as well as trips originated from Brooklyn and Queens and ended in the study area was collected. There are a total of 28,861,581 trips made among 660 origin stations and 685 destination stations. The average trips per day of the study area are approximately 26,357. Citi Bike has a policy of time limit per ride, i.e., 30 minutes for day pass and 45 minutes for annual membership, otherwise penalty will apply. The durations of about 99% of the Citi Bike trips are less than 60 minutes.

Previous studies *(32, 33)* assigned bicycle trips to road networks using GPS data, which could not be obtained from the Citi Bike data. We propose a new method to estimate the exposure of bicycles by assuming that after picking up from a Citi Bike station, bicyclists have the equal change of traveling in any directions. The procedure is described as follows:

Step 1. Calculate the trip distance (Euclidean distance) between the origin and destination stations for each trip.

[4] Source: https://gis.ny.gov/gisdata
[5] Source: http://www.nyc.gov/html/tlc
[6] Source: http://web.mta.info

Step 2. Let D be the maximum trip duration ($D = 60$ min in this study) and J be the total number of time intervals ($J = 6$ in this study). The j^{th} time interval can be represented by $\left[(j-1)\frac{D}{J}, j\frac{D}{J}\right]$. Label each trip with the time interval j by finding which time interval the trip duration belongs to.

Step 3. Obtain the number of trips n_{ij} and median trip distance d_{ij} for time interval j of station i ($i = 1, 2, 3, ..., I$, where I is the total number of origin stations). The median trip distance instead of mean trip distance is used to mitigate the impact of extreme distance values.

Step 4. Generate one circle buffer b_{ij} for station i's time interval j with the station i as the center and d_{ij} as the radius of the buffer.

Step 5. Count the number of cells N_{ij} covered by the buffer b_{ij} for station i and time interval j.

Step 6. Calculate the number of trips t_{ijm} for each cell m using equation (2).

$$t_{ijm} = \begin{cases} \frac{n_{ij}}{N_{ij}} & if\ cell\ m\ within\ buffer\ b_{ij} \\ 0 & otherwise \end{cases} \tag{2}$$

Step 7. Calculate the number of trips of each cell m by summing the number of trips generated by station i's time interval j using equation (3).

$$T_m = \sum_{i=1}^{I} \sum_{j=1}^{J} t_{ijm} \tag{3}$$

The pseudo code for the trip assigning process is presented in Algorithm 1.

Algorithm 1. Pseudo Code for Citi Bike Exposure Estimation

```
m = 1; t_ijm = 0; T_m = 0
While m ≤ M do
    i = 1;
    While i ≤ I do
        j = 1;
        While j ≤ J do
            Using equation (2) to calculate the number of trips t_ijm;
                T_m += t_ijm;
            j += 1;
        i += 1;
    m += 1;
```

This method overcomes the unpredictable nature of cyclists and distributes trips originated from one station as a whole instead of predicting the specific route each bicyclist chose, which may have so many alternatives especially in the grid network of Manhattan. The result of Citi Bike trip assignment is illustrated in Figure 1.

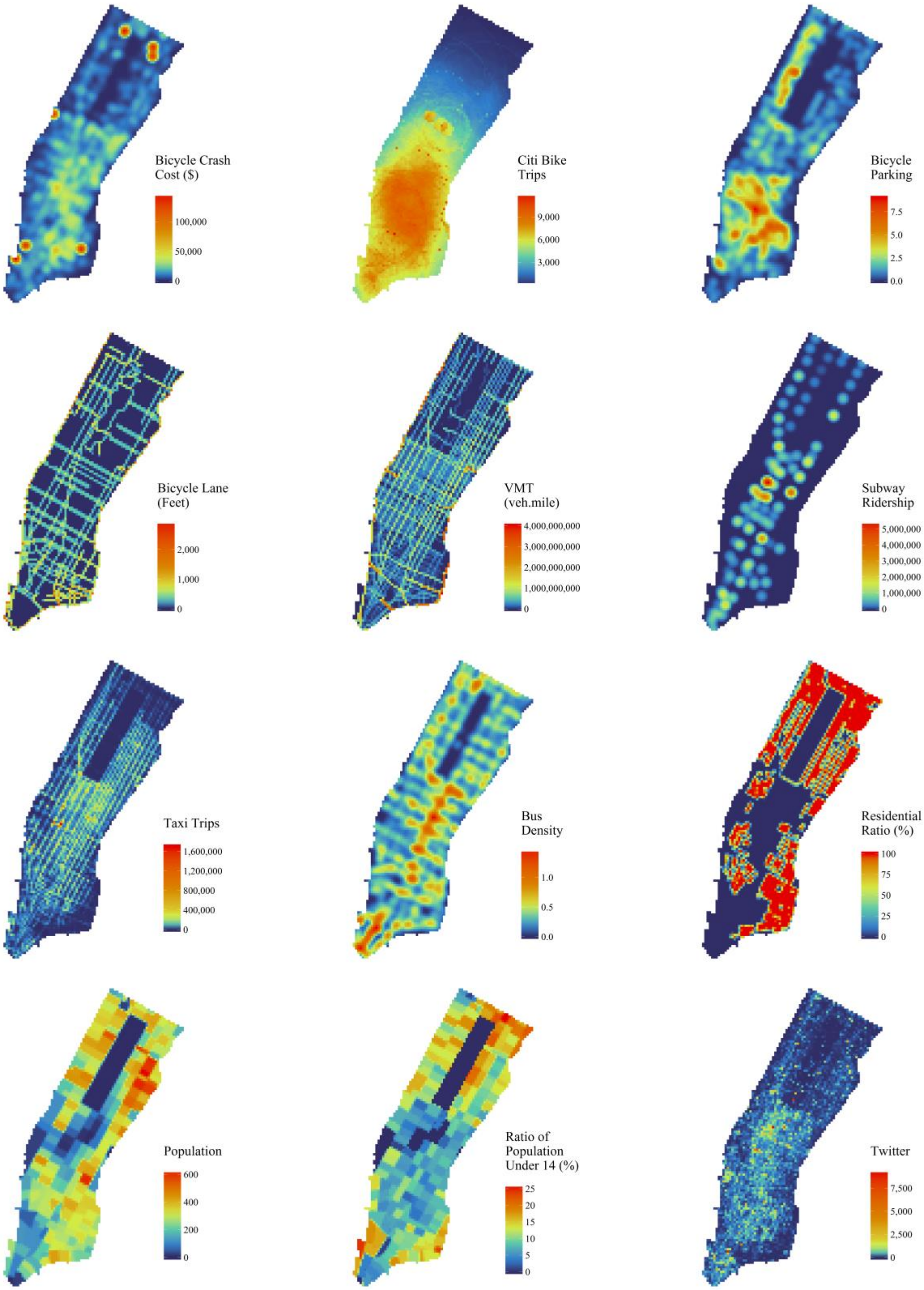


Figure 1 Spatial distributions of grid cell-specific features.

The descriptions and descriptive statistics of crashes, transportation, land use, demo-economic and social media data are listed in Table 2. Spatial distributions of some key variables are visualized in Figure 1.

Table 2 Descriptions and Descriptive Statistics of Key Variables (N=4,504 grid cells)

Variable	Description	Mean	Standard Deviation
Crash			
Crash cost	Annual average cost of bicyclists crashes after spatial processing ($) (35 zeros)	14,003.06	15,617.92
Transportation			
Bicycle ridership	Number of Citi Bike trips per cell	4,512.27	3,050.24
Bicycle rack density	Number of racks per cell	1.46	1.49
Bicycle lane	Total length of bicycle lanes (mile)	0.04	0.06
VMT	Annual vehicle miles traveled (10^3 veh.mile)	327.26	468.61
Truck ratio	The average ratio of truck flow to total flow	0.05	0.05
Subway ridership	Annual subway ridership after spatial processing (10^3)	187.46	430.63
Bus stop density	Number of bus stops after spatial processing	0.36	0.23
Taxi trip	Average of annual taxi pick-ups and drop-offs (10^3)	58.29	75.16
Land use			
Commercial ratio	The ratio of commercial zone area to the whole area (%)	36.74	42.74
Residential ratio	The ratio of residential zone area to the whole area (%)	42.05	43.73
Mixed ratio	The ratio of mixed zone area to the whole area (%)	8.63	25.64
Park ratio	The ratio of park area to the whole area (%)	12.56	30.44
Socio-demographic			
Population	Total population	239.86	149.85
Population under 14	The population under 14 years (%)	9.26	5.73
Population over 65	The population 65 years and over (%)	12.49	8.46
Male	The population of males	112.69	68.40
Female	The population of females	127.15	82.95
Median age	Median age of population	41.05	12.44
Median income	Median income per household (10^3 $)	81.53	49.19
Employed	Number of the employed	137.07	91.69
Unemployed	Number of the unemployed	9.88	7.45
Social media			
Tweet number	Average number of tweets per year	225.87	426.38

METHODOLOGY

Tobit Model

Tobin (1958) *(34)* first proposed the Tobit model, which can accommodate left-censored dependent variables. Tobit models have been applied previously to model the crash costs in the study *(4)* and crash rates in the studies *(35)* and *(27)*. The Tobit model assumes that the dependent variable Y_i is equal to the latent variable Y_i^* when Y_i^* is positive and is equal to zero when Y_i^* is less than or equal to zero. Crash cost is non-negative and be regarded as a variable left-censored at zero. The Tobit model can be described as:

$$Y_{ij}^* = \beta_0 + \sum_{p=1}^{P} \beta_p X_{pij} + \varepsilon_{ij} \tag{4}$$

$$Y_{ij} = \begin{cases} Y_{ij}^* & if\ Y_{ij}^* > 0 \\ 0 & if\ Y_{ij}^* \le 0 \end{cases} \quad (5)$$

Y_{ij} is the average annual bicycle crash cost for i^{th} cell on j^{th} neighborhood ($i = 1,...,n_j$, n_j is the total number of cells on j^{th} neighborhood; $j = 1,...,J$, J is the total number of neighborhoods) in the study time period. Y_i^* is the latent variable for bicycle crash cost. X_{pij} are cell-specific explanatory variables ($p = 1,...,P$, P is the total number of explanatory variables). β_0 and β_p ($p = 1,...,P$) are the regression coefficients to be estimated. ε_{ij} is the error term which follows a normal distribution with mean zero and variance σ_ε^2.

Random Parameters (RP) Tobit Model

It is likely that the relation between contributing factors and bicycle crash cost can vary from one neighborhood to another. For instance, the increase of truck volume may have less influence on the crash risk in freeways but greater influence in the local roads. This variation in safety impacts is referred to as heterogeneity. Random parameters models, which allows some or all estimated parameters to vary spatially, can account for the unobserved heterogeneity *(36)*. The specification of the RP Tobit models *(37)* is given as:

$$Y_{ij}^* = \beta_{0j} + \sum_{p=1}^{P} \beta_{pj} X_{pij} + \varepsilon_{ij} \quad (6)$$

$$\beta_{pj} = \beta_p + \kappa_{pj} \quad (7)$$

β_{pj} ($p = 1,...,P$) are the random parameters to be estimated for the j^{th} neighborhood, with the mean β_p. κ_{pj} is a normally distributed term with mean 0 and variance $\sigma_p{}^2$. The Bayesian method provides advantages to flexibly accommodate complicated model structures *(38-40)* and is used to develop the RP Tobit model in this study.

Model Assessment

Deviance information criterion (DIC) is widely used as a comprehensive measure of fitting and complexity of Bayesian models *(41)*. Specifically, DIC is calculated as follows:

$$DIC = \overline{D(\theta)} + p_D \quad (8)$$

$D(\theta)$ is the Bayesian deviance of the estimated parameter θ; $\overline{D(\theta)}$ is the posterior mean of $D(\theta)$ and indicates how well the model fits the data; p_D defines the effective number of parameters and indicates the complexity of model structure. A difference in DIC that is larger than 5 suggests that the model with a smaller DIC should be favored *(36)*.

Besides DIC, widely used statistical measures including R-squared, Mean Squared Predictive Error (MSPE), and Mean Absolute Deviance (MAD) *(36, 42)* are used to indicate the goodness-of-fit of models.

MODELING RESULTS

This section presents modeling results and interpretation of variable estimates. Both the Tobit model and the RP Tobit model were developed in Bayesian framework to estimate the annual cost of bicycle-related crashes. Bayesian models were developed using the software WinBUGS *(41)*. The R2WinBUGS package *(43)* offered convenient functions to call WinBUGS from R, and thus an integrated process of data processing and Bayesian model development could be carried out in the R environment. Without credible prior information, uninformative priors were assumed *(44, 45)*. The Brooks-Gelman-Rubin (BGR) diagnostic *(46)* was used to assess the convergence of multiple chains. Considering convergence and time of updating, two MCMC chains of 20,000 iterations were run, and the first 10,000 samples were discarded as burn-in.

To conduct valid comparison, all the explanatory variables incorporated into the two models were kept the same. In the RP Tobit model, only if the estimated standard deviation (SD) of a random parameter was significantly positive and the incorporation of this random parameter would lead to lower DIC, the parameter was allowed to vary randomly across groups. Consequently, the coefficients of bicycle rack density and subway ridership were set to be random parameters. Variance Inflation factors (VIFs) were computed to diagnose multicollinearity of explanatory variables incorporated. A VIF greater than 5 indicates the existence of multicollinearity problem *(47)*. As shown in Table 3, no multicollinearity issue is detected, since all VIFs are far less than 5.

Table 3 Detection of Multicollinearity Using Variance Inflation Factors (VIF)

Variable	VIF
Citi Bike trip	1.86
Bicycle rack density	1.53
Subway ridership	1.28
Taxi trip	1.27
Bus stop density	1.24
Residential ratio	1.71
Population	1.90
Ratio of population under 14	1.56
Median age	1.53

The Bayesian posteriors of the Tobit and RP Tobit models are reported in Table 4. The 95% Bayesian Credible Interval (95% BCI) is used to examine the significance of estimations. Estimates can be regarded as significant at the 95% level if the BCIs do not cover 0 and vice versa *(48)*. Except residential ratio, ratio of population under 14 for the Tobit model and subway ridership, ratio of population under 14 for the RP Tobit model, all the other explanatory variables are found to be statistically significant. The SDs of two random parameters (bicycle rack density and subway ridership) are significantly positive, which provides evidences for the unobserved heterogeneity across neighborhoods.

Table 4 Posterior Summary of Bayesian Model Fitting

Parameter	Tobit			RP Tobit		
	Mean (SD)	95% BCI	Marginal Effect	Mean (SD)	95% BCI	Marginal Effect
Intercept	3899.47 (786.56)	(2424.00, 5458.00)	3851.85	2865.91 (828.59)	(1219.01, 4435.00)	2830.91
Transportation						
Bicycle ridership	0.96 (0.08)	(0.79, 1.12)	0.95	1.37 (0.08)	(1.22, 1.53)	1.35
Bicycle rack density	137.31 (17.24)	(102.80, 170.80)	135.63	28.28 (13.13)	(9.65, 56.81)	27.93
SD of bicycle rack density	-	-		1.95 (3.38)	(0.03, 12.35)	
Subway ridership (10^3)	2.44 (0.56)	(1.34, 3.53)	2.41	13.30 (7.56)	(-1.53, 28.31)	13.14
SD of subway ridership	-	-		34.87 (5.87)	(25.34, 48.34)	
Taxi trip	7.29 (3.20)	(1.09, 13.72)	7.20	11.27 (3.16)	(4.97, 17.55)	11.13
Bus stop density	4664.88 (716.86)	(3214.00, 6020.00)	4607.92	5323.48 (713.62)	(3891.97, 6714.00)	5258.47
Land use						
Residential ratio (%)	-10.08 (5.24)	(-20.37, 0.16)	-9.96	-15.04 (5.27)	(-25.40, -5.13)	-14.86
Sociodemographic						
Population	12.34 (1.74)	(9.02, 15.62)	12.19	13.74 (1.69)	(10.42, 16.95)	13.57
Ratio of population under 14 (%)	17.10 (9.79)	(-1.99, 36.02)	16.89	14.09 (9.60)	(-5.07, 33.30)	13.92
Median age	-42.90 (14.46)	(-71.11, -13.36)	-42.38	-38.79 (14.57)	(-67.42, -10.73)	-38.32

Statistical measures including DIC, R-squared, MSPE and MAD were used to assess the model performance, with results presented in Table 5. The DIC value of the RP Tobit (90,169,900) is 400 less than that of the Tobit model (90,170,300), and it indicates that the RP Tobit has better overall performance, although it is penalized by a higher p_D, which reflects the increasing complexity by including random parameters. According to R-square values, the RP Tobit model could explain 20% of the variance in the crash cost, which is much higher than 13% of the Tobit model. Additionally, the RP Tobit model has lower MSPE and MAD compared with the Tobit model, and shows substantial improvement in goodness-of-fit by allowing some parameters to vary across neighborhoods. The result further conforms the existence of neighborhood-specific unobserved heterogeneity.

Table 5 Assessment of Model Performance

Statistical Measures	Tobit	RP Tobit
DIC	90,170,300	90,169,900
$\overline{D(\theta)}$	90,170,291	90,169,872
p_D	9	28
R-squared	0.13	0.20
MSPE	2.11E+08	1.95E+08
MAD	7,476	7,141

The RP Tobit model, due to its comparatively better performance, is used to explore the impact of explanatory variables. The marginal effects of explanatory variables were computed (refer to *(4)* for the equation to obtain marginal effects for Tobit models) and reported in Table 4. Bicycle ridership was found to be positively associated with the crash cost. It is an anticipated result since more bicycle trips are associated with greater opportunities to be involved with crashes. The marginal effect of bicycle ridership can be interpreted as: one unit increase in bicycle ridership is expected to raise the bicycle crash cost by $1.35. Another bicycle exposure indicator namely, bicycle rack density was found to be statistically significant and one additional rack would lead to approximately \$27.93 more bicycle-related crash cost annually. This finding is quite intuitive since the number of bicycle racks are related to bicycle usage. The coefficients of subway ridership, bus stop density and taxi trip are all found to be significantly positive. A possible reason for that is regions with higher subway, bus and taxi usage generally have more activities, which also attract more bicyclists. Consistent with the findings by previous studies *(7, 12, 13)*, land use patterns could affect the risk of bicycling. The ratio of residential areas was found to have negative impact on crash cost, since compared to the commercial areas, the residential areas have relatively lower traffic and thus lower risk of involving bicycle crashes. Additionally, the RP Tobit model showed that ratio of population under 14 were positively associated with bicycle crash cost while median age negatively associated with bicycle crash cost. It is likely that young people have a higher chance to travel by bicycle and thus are associated with higher exposure to bicycle-related crash risks. Similar to the previous studies *(7, 8, 14)*, population was found to be a significant variable with positive effect on bicycle crashes.

HOTSPOT IDENTIFICATION

In this study, bicycle crash hotspots are defined as sites with excessive bicycle costs. Potential for Safety Improvement (PSI) is used as a criterion for hotspot ranking *(49, 50)*. PSI is the actual bicycle crash cost minus the expected cost of "similar" sites that can be estimated from the crash cost models. PSI can be regarded as the excessive crash cost that would be reduced after implementing proper countermeasures. PSI is given by:

$$PSI_i = Y_i - E(Y_i) \tag{9}$$

PSI_i is the potential for safety improvement for site i. $E(Y_i)$ is the expected average crash cost for sites that are similar to site i and is estimated using the RP Tobit model.

The advantages of using the proposed RP Tobit crash cost model to capture PSI lie in: 1) injury severity is considered by being converted into unit costs, and 2) varying effects of certain explanatory variables are accounted for by incorporating random parameters. PSI for each cell is presented in Figure 2, along with the bicycle lanes in Manhattan. The cell-based hotspot identification method provides a complete risk map for bicyclists with higher resolutions than methods based on census tracts or TAZs. Spatial clustering of high-risk cells can be observed from Figure 2, namely, cells with similar colors tend to gather together. Most locations with high PSIs either have unprotected bicycle lanes or are close to the access points to protected bicycle routes. Cells around protected bicycle paths are generally associated with low risks. The cell with the highest PSI is located in the neighborhood Morningside Heights. Its PSI value implies that the bicycle-related crash cost within this cell is approximately \$129,444 higher than "similar" sites. There are unobserved risk factors (e.g., improper bicycle lane design and traffic control) that contribute to the excessive amount of crash cost and have the potential to be addressed if proper countermeasures are implemented.

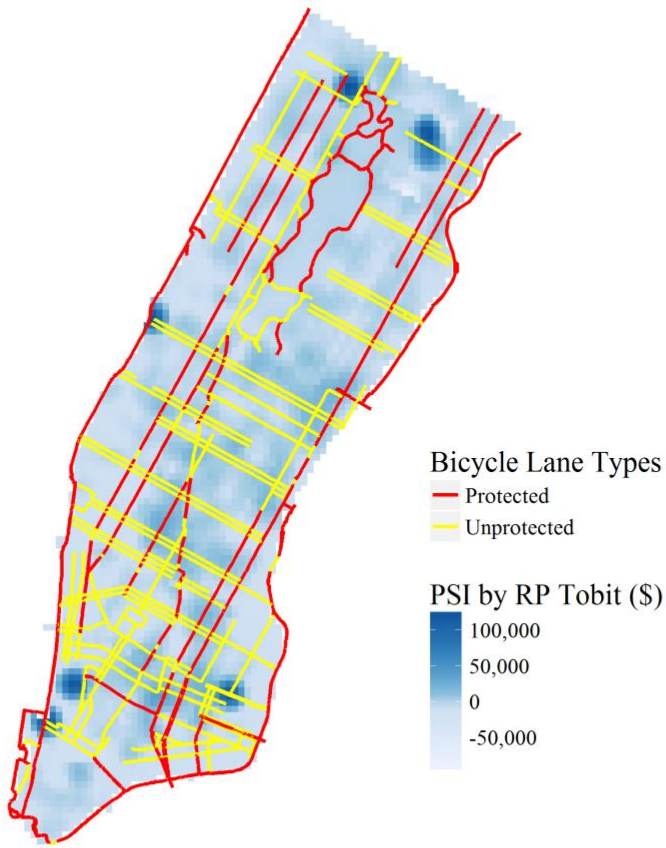
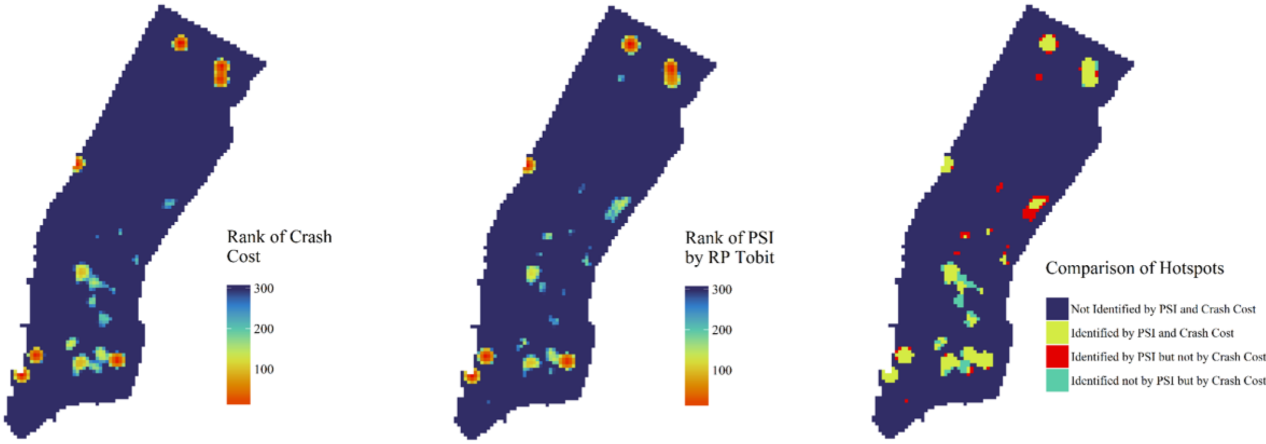


Figure 2 Cell-based potential for safety improvement (PSI) and bicycle lanes.

Crash cost and PSI were used to rank the cells from the most dangerous to the safest. The cells ranked among the top 300 were labelled as hotspots. Figure 3 (a) and (b) demonstrate the rank of crash cost and the rank of PSI, respectively, where the red color indicates the high-ranked (or the most dangerous) hotspots and the blue color indicates the low-ranked (or the safest) hotspots and other cells not labelled as hotspots. Figure 3 (c) shows the difference in hotspot ranking using crash cost and PSI. In summary, there are 236 cells identified as hotspots by both crash cost and PSI, and 64 cells identified as hotspots by only one of the two ranking criteria. We manually examined the 64 cells labeled as hotspots by PSI but not by crash cost. It was found that 59 of those cells did not have protected bike lanes so that bicyclists in those cells are exposed directly to motor vehicles. Moreover, four of those cells were close to the access points to bridges/tunnels with complex road geometry and large disrupting traffic. PSI could identify hotspots with relative low crash costs, which would otherwise be neglected by crash cost ranking.



(a) (b) (c)

Figure 3 Comparisons of hotspots identified by potential for safety improvement (PSI) and crash cost.

SUMMARY AND CONCLUSIONS

This paper aims to investigate the overall safety patterns of bicyclists in Manhattan using a grid-cell-structured modeling framework. The study area was uniformly divided into 300×300 feet^2 grid cells, as the basic geographical units of analysis. The cost of each bicycle crash, weighted by injury severity, was assigned to the cells based on the relative distance to the crash site using a Kernel density function. Transportation, land use, socioeconomic and social media data were captured for each cell using spatial analysis tools. Large-scale datasets such as taxi pick-ups/drop-offs, geotagged tweets and Citi Bike ridership were collected. A new method to estimate the exposure of bicycles were proposed by assuming that after picking up from a Citi Bike station, bicyclists have the equal change of traveling in any directions. The density of bicycle racks was also obtained as another exposure indicator of bicyclists.

A random parameters (RP) Tobit model was developed in the Bayesian framework to model the cell-based bicycle-related crash cost. The proposed RP Tobit model could not only deal with left-censored crash cost data, but also account for the inter-neighborhood unobserved heterogeneity. The RP Tobit model was found to outperform the Tobit model by allowing its parameters to vary across neighborhoods. Thereby, the RP Tobit model was used to investigate the effects of contributing factors to bicycle-related crashes. Results indicated that bicycle ridership, bicycle rack density, subway ridership, taxi trip, bus stop density, population, and ratio of population under 14 were positively associated with bicycle crash cost, whereas residential ratio and median age had negative impact on bicycle crash cost.

The RP Tobit model was used to estimate the cell-specific potential for safety improvement (PSI), which was computed by using the actual bicycle-related crash cost minus the average cost of "similar" sites estimated. The advantages of using the proposed RP Tobit crash cost model to capture PSI lie in: 1) injury severity is considered by being converted into unit costs, and 2) varying effects of certain explanatory variables are accounted for by incorporating random parameters. The cell-based hotspot identification method has more flexibility for organizing the spatial data and can provide a high-resolution risk map for the whole study areas. It was found that most locations with high PSIs either had unprotected bicycle lanes or were close to the access points to protected bicycle routes. The cells with high PSI values were identified as hotspots of bicycle crashes, and were compared with hotspots identified by crash cost. It was found that PSI could identify hotspots with relative low crash costs, which would otherwise be neglected by crash cost ranking.

This study could help government agencies gain deeper insights into the overall safety patterns of bicyclists in Manhattan and assist them in making better decisions on the allocation of treatment resources. For future study, we consider including real-time data from sources such as connected vehicles, loop detectors and surveillance cameras to detect dynamic hotspots, and to support more proactive road safety management.

ACKNOWLEDGMENTS

The work is partially funded by Connected Cities for Smart Mobility towards Accessible and Resilient Transportation (C^2SMART) Center at the Tandon School of Engineering at New York University (NYU). The authors would like to thank the New York State Department of Transportation, the New York City Department of Transportation, the New York Metropolitan Transportation Council, the Metropolitan Transportation Authority, and the New York City Department of City Planning and Citi Bike for providing data for the study. The contents of this paper reflect views of the authors who are responsible for the facts and accuracy of the data presented herein. The contents of the paper do not necessarily reflect the official views or policies of the agencies.

REFERENCES

1. NYCDOT, Access 7/17/2017. Cycling in the city: cycling trends in NYC. *http://www.nyc.gov/html/dot/downloads/pdf/cycling-in-the-city.pdf*.
2. NHTSA, Access 7/17/2017. Traffic Safety Facts 2015 Data: Bicyclists and Other Cyclists. *https://crashstats.nhtsa.dot.gov/Api/Public/Publication/812382*.
3. Nordback, K., W.E. Marshall, and B.N. Janson, 2014. Bicyclist safety performance functions for a U.S. city. *Accident Analysis & Prevention* 65, pp. 114-122.
4. Xie, K., K. Ozbay, A. Kurkcu, and H. Yang, 2017. Analysis of Traffic Crashes Involving Pedestrians Using Big Data: Investigation of Contributing Factors and Identification of Hotspots. *Risk Analysis*, pp. 1-18.
5. Wang, X., M. Abdel-Aty, A. Nevarez, and J. Santos, 2008. Investigation of safety influence area for four-legged signalized intersections: nationwide survey and empirical inquiry. *Transportation Research Record: Journal of the Transportation Research Board*(2083), pp. 86-95.
6. Wikipedia, T.F.E., 2017. Citi Bike, https://en.wikipedia.org/wiki/Citi_Bike, Access July 20.
7. Kim, K., P. Pant, and E. Yamashita, 2010. Accidents and accessibility: Measuring influences of demographic and land use variables in Honolulu, Hawaii. *Transportation Research Record: Journal of the Transportation Research Board*(2147), pp. 9-17.
8. Siddiqui, C., M. Abdel-Aty, and K. Choi, 2012. Macroscopic spatial analysis of pedestrian and bicycle crashes. *Accident Analysis & Prevention* 45, pp. 382-391.
9. Wei, F. and G. Lovegrove, 2013. An empirical tool to evaluate the safety of cyclists: Community based, macro-level collision prediction models using negative binomial regression. *Accident Analysis & Prevention* 61, pp. 129-137.
10. Strauss, J., L.F. Miranda-Moreno, and P. Morency, 2013. Cyclist activity and injury risk analysis at signalized intersections: A Bayesian modelling approach. *Accident Analysis & Prevention* 59, pp. 9-17.

11. Zhang, Y., J. Bigham, D. Ragland, and X. Chen, 2015. Investigating the associations between road network structure and non-motorist accidents. *Journal of Transport Geography* 42, pp. 34-47.
12. Kaplan, S. and C. Giacomo Prato, 2015. A Spatial Analysis of Land Use and Network Effects on Frequency and Severity of Cyclist–Motorist Crashes in the Copenhagen Region. *Traffic Injury Prevention* 16(7), pp. 724-731.
13. Chen, P., 2015. Built environment factors in explaining the automobile-involved bicycle crash frequencies: a spatial statistic approach. *Safety science* 79, pp. 336-343.
14. Park, J., M. Abdel-Aty, J. Lee, and C. Lee, 2015. Developing crash modification functions to assess safety effects of adding bike lanes for urban arterials with different roadway and socio-economic characteristics. *Accident Analysis & Prevention* 74, pp. 179-191.
15. Osama, A. and T. Sayed, 2016. Evaluating the impact of bike network indicators on cyclist safety using macro-level collision prediction models. *Accident Analysis & Prevention* 97, pp. 28-37.
16. Li, H., D.J. Graham, and P. Liu, 2017. Safety effects of the London cycle superhighways on cycle collisions. *Accident Analysis & Prevention* 99, Part A, pp. 90-101.
17. Zahabi, S., J. Strauss, K. Manaugh, and L. Miranda-Moreno, 2011. Estimating potential effect of speed limits, built environment, and other factors on severity of pedestrian and cyclist injuries in crashes. *Transportation Research Record: Journal of the Transportation Research Board*(2247), pp. 81-90.
18. Kaplan, S., K. Vavatsoulas, and C.G. Prato, 2014. Aggravating and mitigating factors associated with cyclist injury severity in Denmark. *Journal of safety research* 50, pp. 75-82.
19. Chen, P. and Q. Shen, 2016. Built environment effects on cyclist injury severity in automobile-involved bicycle crashes. *Accident Analysis & Prevention* 86, pp. 239-246.
20. Heesch, K.C., J. Garrard, and S. Sahlqvist, 2011. Incidence, severity and correlates of bicycling injuries in a sample of cyclists in Queensland, Australia. *Accident Analysis & Prevention* 43(6), pp. 2085-2092.
21. Klassen, J., K. El-Basyouny, and M.T. Islam, 2014. Analyzing the severity of bicycle-motor vehicle collision using spatial mixed logit models: A City of Edmonton case study. *Safety science* 62, pp. 295-304.
22. Rifaat, S.M., R. Tay, and A. de Barros, 2011. Effect of street pattern on the severity of crashes involving vulnerable road users. *Accident Analysis & Prevention* 43(1), pp. 276-283.
23. Prati, G., L. Pietrantoni, and F. Fraboni, 2017. Using data mining techniques to predict the severity of bicycle crashes. *Accident Analysis & Prevention* 101, pp. 44-54.
24. Jacobsen, P.L., 2003. Safety in numbers: more walkers and bicyclists, safer walking and bicycling. *Injury Prevention* 9(3), pp. 205-209.
25. Love, D.C., A. Breaud, S. Burns, J. Margulies, M. Romano, and R. Lawrence, 2012. Is the three-foot bicycle passing law working in Baltimore, Maryland? *Accident Analysis & Prevention* 48, pp. 451-456.
26. Tobin, J., 1958. Estimation of Relationships for Limited Dependent Variables. *Econometrica* 26(1), pp. 24-36.
27. Chen, F., X. Ma, and S. Chen, 2014. Refined-scale panel data crash rate analysis using random-effects Tobit model. *Accident Analysis & Prevention* 73, pp. 323-332.
28. Zeng, Q., H. Wen, H. Huang, X. Pei, and S.C. Wong, 2017. A multivariate random-parameters Tobit model for analyzing highway crash rates by injury severity. *Accident Analysis & Prevention* 99, Part A, pp. 184-191.

29. Elvik, R. and T. Bjørnskau, 2017. Safety-in-numbers: A systematic review and meta-analysis of evidence. *Safety Science* 92, pp. 274-282.
30. Chen, L., C. Chen, R. Srinivasan, C.E. McKnight, R. Ewing, and M. Roe, 2012. Evaluating the safety effects of bicycle lanes in New York City. *American journal of public health* 102(6), pp. 1120-1127.
31. Council, N.S., 2012. Estimating the costs of unintentional injuries.
32. Hood, J., E. Sall, and B. Charlton, 2011. A GPS-based bicycle route choice model for San Francisco, California. *Transportation letters* 3(1), pp. 63-75.
33. Halldorsdottir, K., O.A. Nielsen, and C.G. Prato, 2013. Bicycle route choice model for the Greater Copenhagen Area, Denmark-Based on GPS observations. In: Proceedings of the 2nd Symposium of the European Association for Research in Transportation.
34. Tobin, J., 1958. Estimation of relationships for limited dependent variables. *Econometrica: Journal of the Econometric Society*, pp. 24-36.
35. Anastasopoulos, P.C., A.P. Tarko, and F.L. Mannering, 2008. Tobit analysis of vehicle accident rates on interstate highways. *Accident Analysis & Prevention* 40(2), pp. 768-775.
36. Xie, K., X. Wang, K. Ozbay, and H. Yang, 2014. Crash frequency modeling for signalized intersections in a high-density urban road network. *Analytic Methods in Accident Research* 2, pp. 39-51.
37. Anastasopoulos, P.C. and F.L. Mannering, 2009. A note on modeling vehicle accident frequencies with random-parameters count models. *Accident Analysis and Prevention* 41(1), pp. 153-159.
38. Mitra, S. and S. Washington, 2007. On the nature of over-dispersion in motor vehicle crash prediction models. *Accident Analysis and Prevention* 39(3), pp. 459-468.
39. Lan, B., B. Persaud, C. Lyon, and R. Bhim, 2009. Validation of a full Bayes methodology for observational before-after road safety studies and application to evaluation of rural signal conversions. *Accident Analysis and Prevention* 41(3), pp. 574-580.
40. Xie, K., X. Wang, H. Huang, and X. Chen, 2013. Corridor-level signalized intersection safety analysis in Shanghai, China using Bayesian hierarchical models. *Accident Analysis and Prevention* 50, pp. 25-33.
41. Spiegelhalter, D.J., N.G. Best, B.R. Carlin, and A. van der Linde, 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B: Methodological* 64, pp. 583-616.
42. Xie, K., K. Ozbay, Y. Zhu, S. Demiroluk, H. Yang, and H. Nassif, 2017. Modeling the Salt Usage during Snow Storms: An Application of Hierarchical Linear Models with Varying Dispersion.
43. Sturtz, S., U. Ligges, and A. Gelman, 2005. R2WinBUGS: a package for running WinBUGS from R. *Journal of Statistical software* 12(3), pp. 1-16.
44. Wang, X., K. Xie, M. Abdel-Aty, X. Chen, and P.J. Tremont, 2014. Systematic approach for hazardous-intersection identification and countermeasure development. *Journal of Transportation Engineering* 140(6), pp. 04014022.
45. Xie, K., K. Ozbay, H. Yang, J. Holgurin-Veras, and E.F. Morgul, 2015. Modeling the safety impacts of off-hour delivery programs in urban areas. *Transportation Research Record: Journal of the Transportation Research Board* 4784.
46. Brooks, S.P. and A. Gelman, 1998. General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics* 7(4), pp. 434-455.

47. O’brien, R.M., 2007. A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity* 41(5), pp. 673-690.
48. Gelman, A., 2004. Bayesian data analysis, 2nd ed., Chapman & Hall/CRC, Boca Raton, Florida.
49. Hauer, E., J. Kononov, B. Allery, and M. Griffith, 2002. Screening the road network for sites with promise. *Transportation Research Record: Journal of the Transportation Research Board*(1784), pp. 27-32.
50. Persaud, B., C. Lyon, and T. Nguyen, 1999. Empirical Bayes procedure for ranking sites for safety investigation by potential for safety improvement. *Transportation Research Record: Journal of the Transportation Research Board*(1665), pp. 7-12.