Safety contributing factors analysis of elderly vulnerable road users: global and local perspectives

Xueyu Zhang
Ph.D. Candidate
School of Transportation Engineering, Tongji University
The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Shanghai, China
4800 Cao'an Road, Jiading District, Shanghai, China, 201804
Email: zhangxueyu@tongji.edu.cn

Xuesong Wang, Ph.D., Corresponding author
Professor
School of Transportation Engineering, Tongji University
The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Shanghai, China
4800 Cao'an Road, Jiading District, Shanghai, China, 201804
Email: wangxs@tongji.edu.cn

Mohamed Abdel-Aty, Ph.D.
Professor
Department of Civil, Environmental and Construction Engineering, University of Central Florida
Orlando, Florida, FL 32816, United States
Email: M.Aty@ucf.edu

George Yannis, Ph.D.
Professor
Department of Transportation Planning and Engineering, National Technical University of Athens
5 Heroon Polytechniou Str, GR-15773, Athens, Greece
Email: geyannis@central.ntua.gr

Guangzhu Luo
Ph.D. Student
School of Transportation Engineering, Tongji University
The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Shanghai, China
4800 Cao'an Road, Jiading District, Shanghai, China, 201804
Email: guangzhuluo@tongji.edu.cn

Word count: 1,716 words
Submission Date: November 20th, 2024

Acknowledgments

This study was sponsored by the Chinese National Science Foundation (52372335).

Statement of Significance

Increasing attention to elderly traffic safety is necessary to understand the relationship between elderly traffic safety and contributing factors on a spatial scale. However, zero crashes exist at the analysis unit for some specific types of crashes, and few studies have considered the spatial heterogeneity between elderly crash frequency and influencing variables. To fill these gaps, this study developed an approach to explore the effects of contributing factors for elderly vulnerable road users' (VRUs) crashes from global and local perspectives.

Author contribution statement

The authors confirm contribution to the paper as follows: study conception and design, data collection, analysis and interpretation of results, and draft manuscript preparation were provided by Xueyu Zhang, Xuesong Wang, Mohamed Abdel-Aty, George Yannis, and Guangzhu Luo. Authors reviewed the results and approved the final version of the manuscript.

INTRODUCTION

During the rapid increase in the elderly population, the demand for the elderly in terms of life, health and spirituality continues to grow. By the end of 2023, there are 297 million elderly people aged 60 and above in China, accounting for 21.1% of the total population. China has the largest elderly population in the world and it becomes particularly important to tackle the population aging while improving the road traffic safety of the elderly. However, due to their declining physical, cognitive and self-protection abilities, the elderly are the most vulnerable road users (VRUs) when it comes to traffic safety. The elderly (61+) account for 37.18% of total fatalities in 2021, up from 25.77% in 2015. The percentage of injured elderly in the total injuries increases from 15.50% in 2015 to 23.44% in 2019 (*1*).

The abundance of zero crashes in many analysis units for some specific types of crashes, makes traditional models unsuitable for accurate analysis in spatial modeling. Few studies have revealed the spatial heterogeneity in the effects of contributing factors on older VRU crashes. Ignoring this spatial heterogeneity can lead to inaccurate predictions. The GWRF model has the dual capability to capture effectively the spatial heterogeneity and nonlinear relationships. Therefore, the GWRF model is employed in this study to understand how factors contribute to older VRU crashes in different spatial areas.

To improve elderly safety, this study proposes an analytic approach for contributing factors analysis of elderly VRU safety, which uses global and local models to analyze elderly VRU crashes. The study framework mainly includes three parts: (1) data collection and preprocessing, (2) global and local safety modeling, and (3) analyzing global and local results, including global relative influences, marginal effects, variables influences that vary across spatial analysis units of elderly-involved pedestrian crashes and non-motorized vehicle (NMV) crashes.

METHODOLOGY

Global Safety Modeling

To deal with zero-inflated crash data, the study employed gradient tree-boosted Tweedie compound Poisson models (TDboost), proposed by Yang et al. (*30*). The Tweedie distribution offers an integrated framework to model over-dispersed (variance greater than the mean), under-dispersed (variance lesser than the mean), and zero-inflated (more numbers of zero).

The Tweedie compound Poisson distribution is a special class of distributions in the family of exponential dispersion distributions, and the probability density function of the family of exponential dispersion distributions can be expressed as

$$f_Y(y;\theta,\phi) = a(y,\phi)\exp\{\frac{y\theta-\kappa(\theta)}{\phi}\} \quad (1)$$

where $a(\cdot)$ and $\kappa(\cdot)$ are given functions, θ is a parameter in $\mathbb{R}$, and ϕ is the dispersion parameter in $\mathbb{R}^+$. For Tweedie models, the mean $\mathrm{E(Y)} = \mu = \kappa'(\theta)$ and the variance $\mathrm{Var(Y)} = \phi\kappa''(\theta)$, where $\kappa'(\theta)$ and $\kappa''(\theta)$ are the first and second derivatives of $\kappa(\theta)$, respectively. The power mean-variance relationship of Tweedie models is $\mathrm{Var(Y)} = \phi\mu^\rho$ for some index parameter $\rho \in (1,2)$, which gives $\kappa''(\theta) = \mu^\rho$, $\theta = \mu^{1-\rho}/(1-\rho)$, and $\kappa(\theta) = \mu^{2-\rho}/(2-\rho)$. The λ, α , γ can be reparametrized as

$$\lambda = \frac{\mu^{2-\rho}}{\phi(2-\rho)} \quad (2)$$
$$\alpha = \frac{2-\rho}{\rho-1} \quad (3)$$
$$\gamma = \phi(\rho-1)\mu^{\rho-1} \quad (4)$$

As a result, Equation (1) can be expressed as

$$f_Y(y|\theta,\phi,\rho) = a(y,\phi,\rho)\exp\{\frac{1}{\phi}(\frac{y\mu^{1-\rho}}{1-\rho} - \frac{\mu^{2-\rho}}{2-\rho})\} \quad (5)$$

A random variable *Y* is to obey a Tweedie compound Poisson distribution if its probability density function has the form of Equation (5) with $1 < \rho < 2$ and $\mu > 0$, denoted by $\mathrm{Tw}(\mu,\phi,\rho)$ where $1 < \rho < 2$ and $\mu > 0$.

The log-likelihood of the Tweedie compound Poisson model is

$$\log f_Y(y|\theta,\phi,\rho) = \frac{1}{\phi}\left(\frac{y\mu^{1-\rho}}{1-\rho} - \frac{\mu^{2-\rho}}{2-\rho}\right) + \log a(y,\phi,\rho) \quad (6)$$

The normalizing function $a(\cdot)$ can be expressed as

$$a(y,\phi,\rho) = \begin{cases} 1, for\ y = 0 \\ \frac{1}{y}\sum_{t=1}^{\infty} W_t(y,\phi,\rho) = \frac{1}{y}\sum_{t=1}^{\infty} \frac{y^{t\alpha}}{(\rho-1)^{t\alpha}\phi^{t(1+\alpha)}(2-\rho)^t t!\Gamma(t\alpha)}, for\ y > 0 \end{cases} \quad (7)$$

where $\alpha = (2-\rho)/(1-\rho)$ and $\sum_{t=1}^{\infty} W_t$ is an example of Wright's generalized Bessel function.

Following the settings of the above model, the crash Y_i is denoted by $\mathrm{Tw}(\mu_i,\phi,\rho)$, Assume that the expected crash μ_i is determined by a predictor function *F*:

$$\log\{\mu_i\} = \log\{E(Y_i|\mathbf{x}_i)\} = F(\mathbf{x}_i) \quad (8)$$

where $\mathbf{x}_i$ is a vector of independent variables.

The predictor function *F* is estimated by integrating the boosted Tweedie model into the tree-based gradient boosting algorithm. The TDboost model can provide the variable importance and partial dependence plots (PDPs) to interpret the global impact. In this study, 5-fold cross-validation is chosen to seek the estimated best number of trees.

Local Safety Modeling

The variable importance of the global safety analysis model (e.g. RF model) does not change with the spatial location of the samples. Additionally, Global models may not uncover the spatial dependence or heterogeneity in the associations among spatial data in geographical analysis of crash frequency. To account for spatial heterogeneity of factors' effects, the GWR model is the appropriate spatial statistical technique to capture the spatial non-stationarity by establishing a local equation at each analysis unit. The general expression for GWR model is written as:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{n} \beta_k(u_i, v_i)\, x_{ki} + \varepsilon_i \tag{9}$$

where y_i is the dependent variable for grid i, u_i and v_i is the coordinates of the center of grid i, ε_i is the residual, and $\beta_k(u_i, v_i)$ is the local regression coefficient estimate for the independent variable x_k at grid i.

The GWR model can capture spatial heterogeneity or dependence and local variations of crash data, but they cannot reveal the non-linearity of independent variables. To address this limitation, the GWRF model is introduced, which integrates the concepts of GWR and traditional RF model and establishes a local model to characterize spatial heterogeneity effectively while revealing nonlinear relationships. The GWRF model can be expressed as the following equation:

$$y_i = f_{(u_i, v_i)}(x^*) + \varepsilon_i \tag{10}$$

where $f_{(u_i, v_i)}(x^*)$ is the prediction of GWRF model at grid i.

DATA PREPARATION

This study used a grid measuring 1 km × 1 km in Suzhou, China, for the case study area. We focused on pedestrian and NMV crashes involving elderly individuals aged 60 and over, which were selected as dependent variables, from 2019 to 2023. Based on the previous literature, the independent variables, that is, contributing factors, were selected mainly from three aspects including socio-economic, road network, public facility. Socio-economic data included population and housing prices. Road network features, including road network, nodes, and the total length of urban, rural, and national and provincial roads. Public facility data, including the points of interests (POIs) data, was used to reflect the mobility of elderly people. The number of metro exits, bus stops, healthcare, parks, markets were extracted directly from the POIs dataset.

RESULTS

Figure 1 illustrates the variable importance of independent variables in predicting elderly VRU crashes. Population stood out conspicuously in predicting elderly pedestrian crashes, with a contribution of 48.795%, accounting for a substantial proportion of the total importance. This finding is consistent with our intuition. Healthcare was the second most important independent variable with a contribution of 27.22%, respectively. Taken both together, population and healthcare have non-trivial effects on elderly pedestrian crashes.

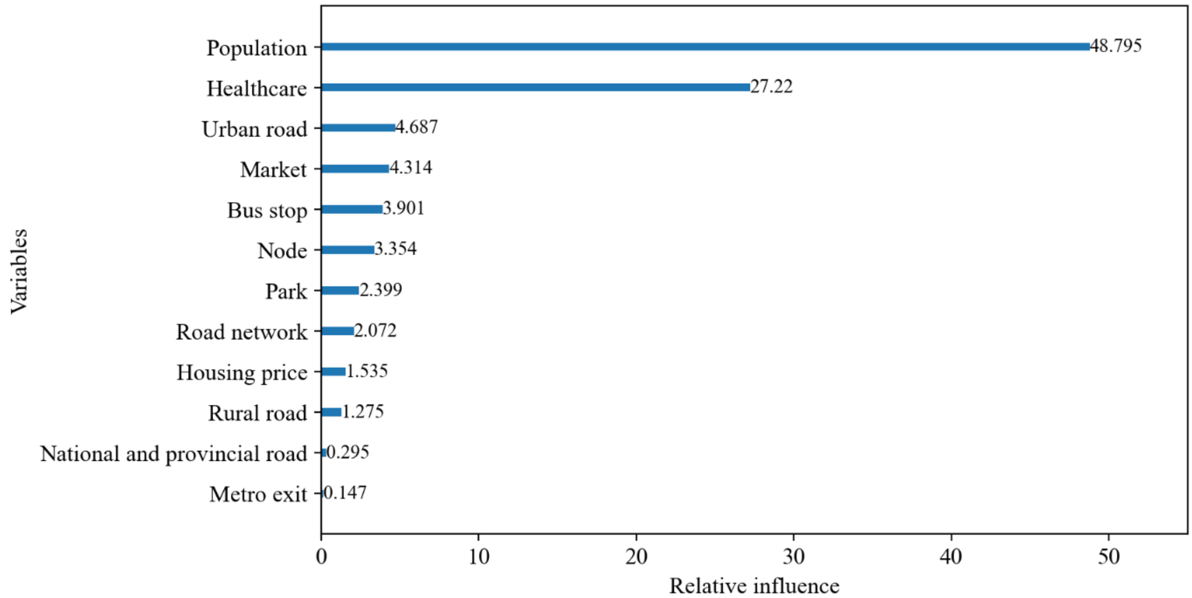


Figure 1 Relative contributions of independent variables to elderly pedestrian crashes

Figure 2 depicts the marginal effects of the top four contributing factors on predicting elderly pedestrian crashes based on the PDPs, after accounting for the average effects. All four contributing factors demonstrated nonlinear effects in predicting elderly pedestrian crashes. First, elderly pedestrian crashes increased substantially when population was under about 11,600 persons, that is, population was positively associated with elderly pedestrian crashes. When population exceeded about 11,600 persons, it had no additional effect on crashes. Similarly, there was an obvious increasing trend as the number of healthcare facilities increased from 0 to about 18.

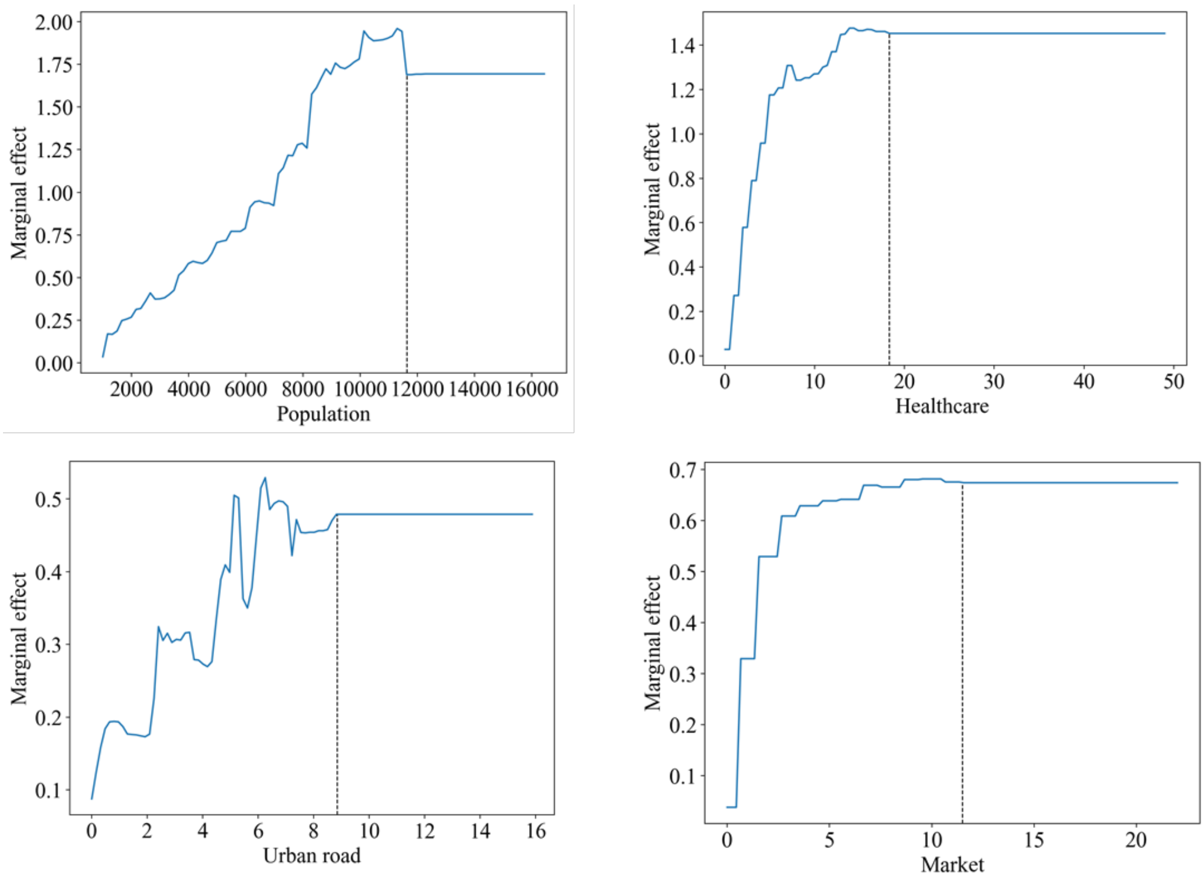
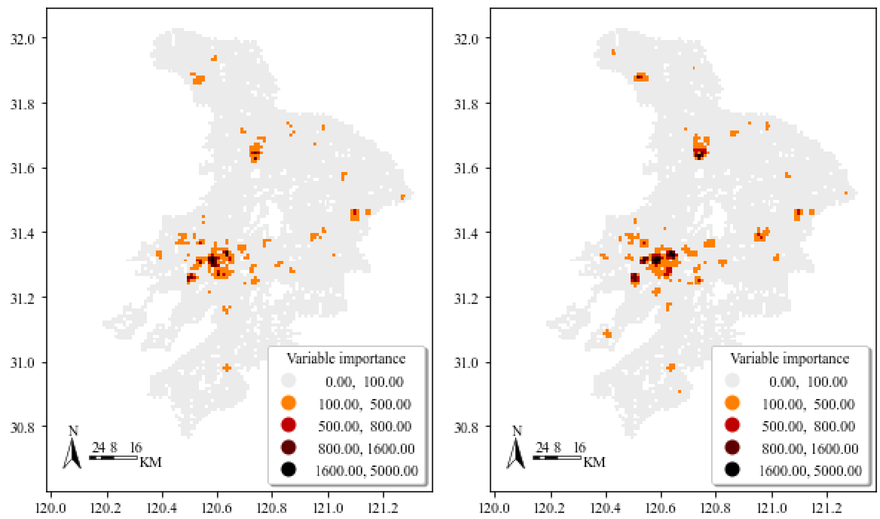


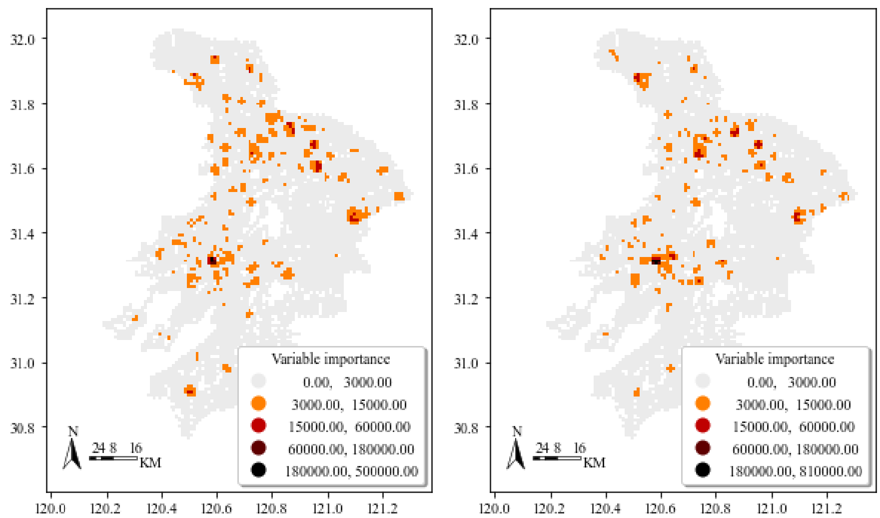
Figure 2 PDPs of key contributing factors on elderly pedestrian crashes

Figure 3 and **Figure 4** illustrate the spatial distribution of the variable importance for contributing factors in predicting elderly VRU crashes. Generally, the importance of population, healthcare, urban road, and market varied significantly across areas in the elderly pedestrian crashes model, showing a spatial clustering tendency with high importance in the downtown areas and central areas of county-level cities (**Figure 3**). In terms of population, it had a wider impact on elderly pedestrian crashes in the downtown areas compared with other factors. These results suggest that elderly pedestrian safety deserves more attention in downtown areas with high population. As shown in **Figure 4**, the importance of contributing factors for elderly NMV crashes was centrally dispersed in the central and northern areas, showing a block distribution tendency. This might be because NMVs travel further compared to walking, leading to a larger impact on elderly NMV crashes. Therefore, NMV safety needs to be taken care of in the northern county-level cities, focusing on areas with a high concentration of road facilities and public utilities.



(a) Population (b) Healthcare

Figure 3 Geographical distribution of variable importance of the major contributing factors in predicting elderly pedestrian crashes



(a) Population (b) Healthcare

Figure 4 Geographical distribution of variable importance of the major contributing factors in predicting elderly NMV crashes

CONCLUSIONS

Abundant zero crashes also exist at the macro-level for some specific types of crashes. In addition, few studies have considered the spatial heterogeneity between elderly crashes and influencing variables. The study tries to bridge these gaps, providing a comprehensive analytic framework from global and local perspectives to model the elderly VRU crashes, after which the contributing factors' safety effects were explored.

This study introduced a viable analytic framework to model the elderly VRU crashes globally and locally, exploring contributing factors' safety effects, and proposed some suggestions to improve the elderly safety. Furthermore, several questions merit further exploration. The study only used grids measuring 1km × 1km as analysis units to aggregate crashes, which might lead to the Modifiable Area Unit Problem (MAUP). The MAUP issue deserves in-depth study by comparing the model results using different analysis units, such as different size grids, traffic analysis zones, and buffer. Additionally, this paper considered travel modes for elderly crashes and the severity of elderly crashes could be further considered to explore the spatial relationship to provide suggestions for policy-makers, planners, and engineers.

REFERENCES

1. Bureau of Traffic Management of the Ministry of Public Security of PRC. China Road Traffic Accidents Statistics. Beijing, 2020.

2. Grisé, E., R. Buliung, L. Rothman, and A. Howard. A Geography of Child and Elderly Pedestrian Injury in The City of Toronto, Canada. *Journal of Transport Geography*, 2018: 66: 321-329.

3. Lee, S., J. Yoon, and A. Woo. Does Elderly Safety Matter? Associations Between Built Environments and Pedestrian Crashes in Seoul, Korea. *Accident Analysis and Prevention*, 2021. 144: 105621.

4. Jiang, M., Sato H, Diao X, G. I.M.A. Mothafer, and T. Yamamoto. Bicycle Accident Risk Factors for Different Age Groups in Nagoya, Japan. *Transportation research record: Journal of the Transportation Research Board*, 2023, 2677(5): 1402-1414.

5. Yu, C., W. Hua, C. Yang, S. Fang, Y. Li, and Q, Yuan. From Sky to Road: Incorporating the Satellite Imagery into Analysis of Freight Truck-Related Crash Factors. *Accident Analysis and Prevention*, 2024, 200: 107491.

6. Cai, Q., J. Lee, N. Eluru, and M. Abdel-Aty. Macro-Level Pedestrian and Bicycle Crash Analysis: Incorporating Spatial Spillover Effects in Dual State Count Models. *Accident Analysis and Prevention*, 2016, 93: 14–22.

7. Almasi, S. A., and H. R. Behnood. Exposure Based Geographic Analysis Mode for Estimating the Expected Pedestrian Crash Frequency in Urban Traffic Zones; Case Study of Tehran. *Accident Analysis and Prevention*, 2022, 168: 106576.

8. Wang, S., K. Gao, L. Zhang, B. Yu, and S. M. Easa. Geographically Weighted Machine Learning for Modeling Spatial Heterogeneity in Traffic Crash Frequency and Determinants In US. *Accident Analysis and Prevention*, 2024, 199: 107528.

9. Wu, D., Y. Zhang, and Q. Xiang. Geographically Weighted Random Forests for Macro-Level Crash Frequency Prediction. *Accident Analysis and Prevention*, 2024, 194: 107370.

10. Shankar, V. N., S. Sittikariya, and M. B. Shyu. Some Insights on Roadway Infrastructure Design for Safe Elderly Pedestrian Travel. IATSS Research, 2006, 30(1): 21–26.

11. Wang, Y., M. M. Haque, H. C. Chin. Elderly Pedestrian Injuries in Singapore. *Journal of Transportation Safety and Security*, 2017, 9(3): 273-300.

12. Silva, M. T. D., P. H. Lora, M. Massago, A. D. C. Dutra, J. L. Gabella, L. L. Silva, F. S. N. Carignano, E. M. D. Souza, and A. M. Obale, J. R. N. Vissoci, A. P. Joiner, C.A. Staton, O. K. Nihei, and L. D. Andrade. Built Environment Influence on The Incidence of Elderly Pedestrian Collisions in A Medium-Large City in Southern Brazil: A Spatial Analysis. *International Journal of Injury Control and Safety Promotion*, 2023, 30(3): 428-438.

13. Das, S., A. Bibeka, X. Sun, H. T. Zhou, and M. Jalayer. Elderly Pedestrian Fatal Crash-Related Contributing Factors: Applying Empirical Bayes Geometric Mean Method. *Transportation Research Record Journal of the Transportation Research Board*, 2019.

14. Gálvez-Pérez, D., B. Guirao, and A. Ortuño. Analysis of the Elderly Pedestrian Traffic Accidents in Urban Scenarios: The Case of The Spanish Municipalities. *International Journal of Injury Control and Safety Promotion*, 2024: 1-20.

15. Soltani, A., M. Azmoodeh, and M. R. Qadikolaei. Post COVID-19 Transformation in the Frequency and Location of Traffic Crashes Involving Older Adults. *Transportation Research Record: Journal of the Transportation Research*, 2023.

16. Lee, J. S., P. C. Zegras, and E. Ben-Joseph. Safely Active Mobility for Urban Baby Boomers: The Role of Neighborhood Design. *Accident Analysis and Prevention*, 2013, 61: 153–166.

17. Luo, Y., Y Liu, Z. Tong, N. Wang, and L. Rao. Capturing gender-age thresholds disparities in built environment factors affecting injurious traffic crashes. *Travel Behaviour and Society*, 2023, 30: 21–37.

18. Kim, D. The Transportation Safety of Elderly Pedestrians: Modeling Contributing Factors to Elderly Pedestrian Collisions. *Accident Analysis and Prevention*, 2019, 131: 268-274

19. Amoh-Gyimah, R., M. Saberi, and M. Sarvi. Macroscopic Modeling of Pedestrian and Bicycle Crashes: A Cross-Comparison of Estimation Methods. *Accident Analysis and Prevention*, 2016, 93: 147-159

20. Huang H, Song B, Xu P, J. Lee, and M. Abdel-Aty. Macro and Micro Models for Zonal Crash Prediction with Application in Hot Zones Identification. *Journal of Transport Geography*, 2016, 54: 248-256.

21. Huang, H., M. Abdel-Aty, A. L. Darwiche. County-Level Crash Risk Analysis in Florida: Bayesian Spatial Modeling. *Transportation Research Record: Journal of the Transportation Research Board*, 2010, 2148: 27-37.

22. Wang, X., J. Yang, C. Lee, Z. Ji, and S. You, 2016. Macro-Level Safety Analysis of Pedestrian Crashes in Shanghai, China. *Accident Analysis and Prevention*, 2016, 96: 12–21.

23. Hezaveh, A. M., and C. R. Cherry. Applying a Home-Based Approach to the Understanding Distribution of Economic Impacts of Traffic Crashes. *Transportation Research Record Journal of the Transportation Research Board*, 2020, 2674(12): 360-371.

24. Rhee, K, R., J. Kim, Y. Lee, and G. F. Ulfarsson. Spatial Regression Analysis of Traffic Crashes in Seoul. *Accident Analysis and Prevention*, 2016, 91: 190-199.

25. Liu, J., S. Das, and M. N. Khan. Decoding the Impacts of Contributory Factors and Addressing Social Disparities in Crash Frequency Analysis. *Accident Analysis and Prevention*, 2024, 194: 107375.

26. Hadayeghi, A., A. S. Shalaby, and B. N. Persaud. Development of Planning Level Transportation Safety Tools Using Geographically Weighted Poisson Regression. *Accident Analysis and Prevention*, 2010, 42: 676-688

27. Gomes, M. T. J. T., F. Cunto, and A. R. D. Silva. Geographically Weighted Negative Binomial Regression Applied to Zonal Level Safety Performance Models. *Accident Analysis and Prevention*, 2017, 106: 254-261.

28. Rahman, M.S., M. Abdel-Aty, S. Hasan, and Q. Cai. Applying Machine Learning Approaches to Analyze the Vulnerable Road-Users' Crashes at Statewide Traffic Analysis Zones. *Journal of Safety Research*, 2019, 70: 275-288.

29. Wang, X., X. Zhang, and Y. Pei. A Systematic Approach to Macro-Level Safety Assessment and Contributing Factors Analysis Considering Traffic Crashes and Violations. *Accident analysis and Prevention*, 2024, 194: 107323.

30. Yang, Yi, W. Qian, and H. Zou . Insurance Premium Prediction via Gradient Tree-Boosted Tweedie Compound Poisson Models. *Journal of Business & Economic Statistics*, 2018, 36:3, 456-470.

31. Yang, Y., H. Zou, W. Qian, and G. Ridgeway. 2019. TDboost, R Package. Available online:https://cran.r-project.org/web/packages/TDboost/TDboost.pdf

32. Kalogirou, S., and S. Georganos, 2019. SpatialML, R Package. Available online: https://cran.r-project.org/web/packages/SpatialML/SpatialML.pdf.