Transport and Telecommunication, 2017, volume 18, no. 4, 282–288
Transport and Telecommunication Institute, Lomonosova 1, Riga, LV-1019, Latvia
DOI 10.1515/ttj-2017-0025

MACRO-LEVEL MODELING OF URBAN TRANSPORTATION SAFETY: CASE-STUDY OF MASHHAD (IRAN)

***Mehdi Mohammadi*[1], *Gholamali Shafabakhsh*[2], *Ali Naderan*[3]**

[1,2] *Semnan University, Faculty of Civil Engineering*
Semnan, Iran
[3] *Science and Research Branch, Islamic Azad University*
Tehran, Iran
[1] *+9823 31533798, mmahdii@semnan.ac.ir*
[2] *+9823 31533786, shafabakhsh@semnan.ac.ir*
[3] *+9821 77725413, naderan@srbiau.ac.ir*

Transportation safety can be aimed at the planning stage in order to adopt safety management and evaluate the long-time policies. The main objective of this research was to make use of crash prediction models in urban transportation planning process. As such, it was attempted to gather data on the results of transportation master plan as well as Mashhad urban crash database. Two modelling method, generalized linear model with negative binomial distribution and geographically weighted regression, were considered as the methods used in this research. Trip variables, including trip by car, trip by bus, trip by bus services and trip by school services, were significant at 95%. The results indicated that both finalized models were competent in predicting urban crashes in Mashhad. Regarding to results urban transportation safety will be improved by changing the modal share for example from private car to bus. The application of the process presented in this study can improve the urban transportation safety management processes and lead to more accurate prediction in terms of crashes across urban traffic areas.

Keywords: safety planning, urban transportation, macro-level modelling, negative binomial, geographically weighted regression

1. Introduction

Generally speaking, traffic safety may be approached in terms of three stages: planning, designing and operation of transportation systems. Given the first stage, it may be argued that safety aims at managing transportation safety and regarding the second stage, this notion is directed towards safety engineering. Regarding the operation of transportation systems, it aims at inspection and evaluation of safety embedded in facilities and equipment of existing transportation systems.

Those individuals in charge of transportation and traffic safety represent two types of approaches in relation to traffic safety: proactive and reactive approaches. Proactive approach refers to a situation in which no serious safety issue or crash has occurred. Accordingly, this approach mainly includes identifying causes of the crash and providing macro solutions to prevent major crashes. Conversely, reactive approach refers to a situation in which the actual safety issue and crash has occurred and, thus, it mainly attempts to prevent the occurrence of similar crashes. Besides, it includes physical modifications in the road networks at the local level (Naderan, 2010).

Having combined the above-mentioned descriptions, it can be concluded that the safety in the planning and designing stages are categorized into the proactive approach and, then, safety in the operation stage is categorized into the reactive approach. It is clear that each of these approaches as well as each level of safety studies, in turn, is important and necessary. However, it seems that the safety issue, especially in the operation stage, has been mainly addressed in terms of the reactive approach. Also, some studies have paid due proactive attention to designing safe transportation facilities in order to address safety issue and reduce future crashes. Unfortunately, researchers and transportation planners have paid less attention to the issue of safety at the planning stage.

As such, the main objective of this research was to make use of crash prediction models in the urban transportation planning process. The results of this research can be used to determine the status of safety planning in the transportation planning process so that the safety issues may be approached proactively and, then, urban transportation systems may be improved and safer thereof.

2. Literature Review

The ratification of Transportation Equity Act for the 21st Century in the United States of America (1998) played an important role in developing macro crash prediction models. According to this Act, which was passed as a general rule, organizations concerned with urban transportation planning were bounded to observe the safety aspects of long-term transportation planning studies. Accordingly, this led researchers to pay more attention to the link between transportation safety and urban transportation planning.

Some researchers consider a combination of planning models and transportation safety models as an ideal solution in this regard. Having reviewed several studies in this regard, it was observed that there are three major issues in the process of macro crash modelling: aggregation level of data, selection of variables and selection of modelling approach. Table 1 summarizes the different studies with an emphasis on the aforementioned three major issues.

Table 1. Different studies with an emphasis on the three major issues in the process of macro crash modelling

Researcher	Study Area	Agg. Level	Variables	Modelling Approach	Comment
(Wang *et al.*, 2017)	Hillsborough County, Florida, USA	TAZ*	traffic volume and geometric design	negative binomial models & random parameters negative binomial	-
(Huang *et al.*, 2016)	Hillsborough County, Florida, U.S.	TAZ	road and traffic characteristics, trip generation, and demographic and socioeconomic factors	macro and micro level Bayesian spatial model	Macro-level crash analysis has the advantage of requirement of less detailed data.
(Tasic and Porter, 2016)	Chicago	Census tracts	socio-economic, land use, road network, travel demand	negative-binomial regression with fixed and random effects	Strong association between the variables related to multimodal transportation availability and usage, and crashes.
(Wei and Lovegrove, 2013)	British Columbia, Canada	TAZ	socio-demographics, transportation demand management, rural or urban, exposure data	negative binomial	-
(Naderan, 2010)	Mashhad, Iran	TAZ	Trip production/ attraction	negative binomial	Strong association between each aim trips and crashes.
(Wier *et al.*, 2009)	San Francisco, California, US	Census tracts	street, land use, and population characteristics	simple bivariate models	-
(Quddus, 2008)	London, UK	census wards	road infrastructure, socioeconomic and traffic conditions.	negative binomial and spatial model	Bayesian hierarchical models are more appropriate in developing a relationship between area-wide traffic crashes and the contributing factors.
(Hadayeghi *et al.*, 2007)	Toronto, Canada	TAZ	traffic intensity, socioeconomic and demographic factors, land use, and traffic demand measures	negative binomial	Good planning-level forecasting models should be parsimonious (i.e. with few variables).
(de Guevara *et al.*, 2004)	Tucson, Arizona, U.S.	TAZ	intersection density, population density, number of employees, intersections density, percentage of miles of principal arterial, percentage of miles of minor arterials, and percentage of miles of urban collectors	negative binomial	Planning-level safety models are feasible and may play a role in future planning activities.
(Hadayeghi *et al.*, 2003)	Toronto, Canada	TAZ	demographic and socioeconomic factors, traffic demand, transport network parameters	negative binomial	-

* Traffic Analysis Zone (TAZ)

3. Research Methodology

According to the studies conducted by (Cameron and Trivedi, 1998), it was indicated that the process of formulating the aggregate crash models was composed of five major steps, as shown in the Figure 1.

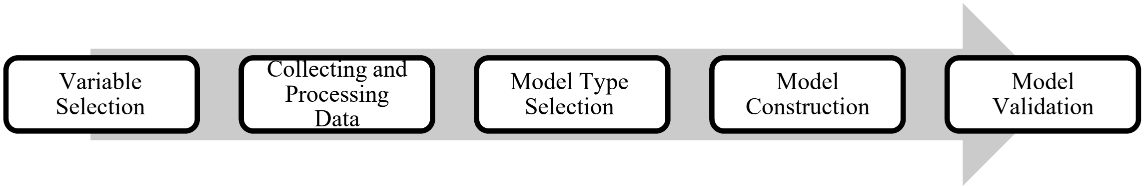


Figure 1. Process of formulating the aggregate crash models

The identification of initial variables was based on modelling objective and theoretical foundations of research. The main objective of this research was directed towards application of crash prediction models in urban transportation planning process. The modal share of different vehicles can be considered as an important variable which plays an important role in urban transportation planning. However, little attention has been paid to this notion. Thus, the development of mathematical models to predict the volume of traffic crashes in Traffic Analysis Zones (TAZs) based on the number of trips taken with different vehicles can be used to improve urban transportation planning processes. Accordingly, some independent accessible variables should be selected in such studies so that the impact of the number of trips taken with different vehicles on the occurrence of crashes may be depicted vividly. Therefore, the result of mode split models as the number of trips taken in each traffic analysis zone separated by trip vehicle was selected as the independent variable in this study. The research data were gathered on the basis of results of Mashhad transportation master plan updating study (MTTO, 2009). All data were initially entered into the GIS environment in the form of different layers and, then, they were assigned to 253 TAZs. TAZs Selected as aggregation level for this study based on two reasons; (i) Most of the literature suggest that TAZs level can make better results in macro-modelling process, (ii) Mashhad transportation master plan had done in TAZs level and all recorded data was in TAZs level. Regarding the next phase (selecting the form of model) and in accordance with previous studies, it was attempted to make use of generalized linear model with negative binomial distribution and geographically weighted regression as research methods in this study.

The negative Binomial (NB) formulation with log-link function is used for estimation of Aggregate Crash Prediction Models (ACPMs) in this study. In the negative binomial model, the difference between variance and mean in presented by $V[u] = u+\alpha u^2$, where V[u] is the estimated variance of the crash frequency; u is the estimated mean crash frequency and α is the NB over-dispersion parameter. The Poisson regression model is regarded as a limiting model of the NB regression model as α approaches zero, which means that the selection between these two models is dependent on the value of α. The negative binomial distribution has the general form of Equation 1.

$$P(y_i) = \frac{\Gamma((1/\alpha)+y_i)}{\Gamma(1/\alpha)y_i!}\left[\frac{1/\alpha}{(1/\alpha)+\lambda_i}\right]^{1/\alpha}\left[\frac{\lambda_i}{(1/\alpha)+\lambda_i}\right]^{y_i} \quad (1)$$

Where Γ(.) is a gamma function. The NB model was developed mathematically by assuming that unobserved crash heterogeneity (variation) across TAZs is gamma distributed, while crashes within sites are Poisson distributed (Washington, *et al*., 2006). The model form in Equation 2 is used where y=predicted crash frequency, Xi=explanatory variables and bi=model parameters:

$$y = \exp\sum_{i=1}^{n} b_i X_i\,. \quad (2)$$

The general form of the geographically weighted regression is shown in Equation 3 (ESRI, 2011):

$$y_{(u,v)} = \beta_{0(u,v)} + \beta_{1(u,v)} + \cdots + \beta_{n(u,v)} + \varepsilon_{(u,v)}\,, \quad (3)$$

where: Y represents the dependent variable, β_0 represents the fixed amount, β_1 to β_n represent coefficients of independent variables, ε represents computational error, and (u, v) represent the coordinates of urban zones (zones in which the data were gathered).

As shown in Equation 3, the geographically weighted regression may be fitted using the Least Squares Method so that the coefficients of points (*u, v*) can be estimates thereof. In this method, the weighting is done in such a way that the data closer to (*u, v*) receive greater weight compared to farther data. This leads the geographical variations of different data into their own related equation and, thus, the accuracy of model is boosted. In this method, the spatial relation among different data enters into modelling process but this does not happen in normal regression (ESRI, 2011). The estimator of geographically weighted regression can be calculated using Equation 4:

$$\hat{\beta}(u_i, v_i) = [X^T W(u_i, v_i) X]^{-1} X^T W(u_i, v_i) y \tag{4}$$

Where, $\hat{\beta}(u_i, v_i)$ represents a vector to estimate β_1 to β_n, $W(u_i, v_i)$ represents the geographical weights matrix, n represents the number of zones, and i represent points. The number of rows and columns of the matrix is equal to the number of observations. This matrix is a diagonal matrix which makes use of a weighting pattern to determine the weights.

The IBM SPSS Statistics 23 (IBM Corp, 2014), ArcMap 10 (ESRI, 2011) and Forward Modelling have been used to develop the models. Given the forward modelling, it is tried to firstly develop a model only by a fixed amount and, subsequently, independent variables are added one after another to the model. The added variable to the model remains only if it improves the model. Otherwise, it is removed and the next variable is tested thereof. This process was performed for different combinations of independent variables. It was attempted to make use of significance of coefficients, Akaike information criterion (AIC), Log Likelihood and their coefficients of determination in order to validate the generalized linear models. Besides, it was attempted to make use of significance of coefficients, AIC, condition number of models and their coefficients of determination in order to validate the geographically weighted regressions. Finally, the top models were selected from both methods.

4. Case Study

Mashhad is Iran's second city in terms of size and population and it has a good transportation infrastructures compared to other cities of Iran. In 2008, a total of 38,882 crashes occurred in Mashhad (out of 6404638 trips in that year). Details of the trips occurred during this period have been shown in Figure 2 and the concerned statistical specifications of data have been shown in Table 2.

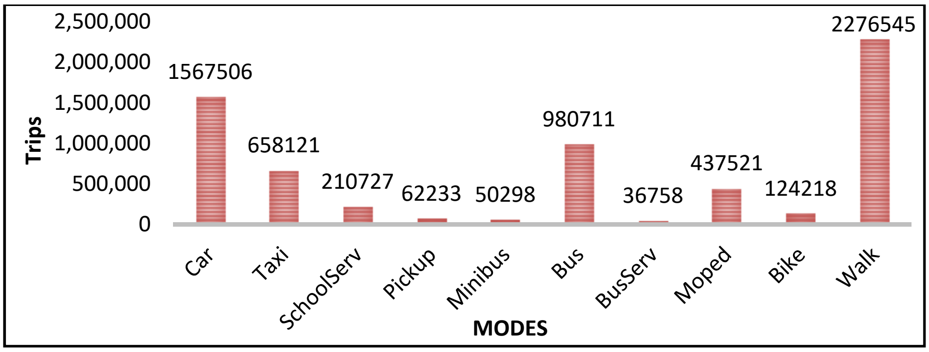


Figure 2. Mashhad Modal share of trips in 2008

Table 2. Descriptive Analysis of Variables

Category	Variable	Definition`	Mean*	Std. D.	Min	Max	Sum
Crashes	Total	Total No. of Crashes	153.68	135.15	0	1,163	38,882
Trips Generated by Mode	Car	Number of trips made by car	6,196.68	4,544.88	139	32,437	1,567,506
	Taxi	Number of trips made by taxi**	2,601.27	2,376.92	0	17,757	658,121
	SchoolServ	Number of trips made by School Services	832.91	792.71	0	4,919	210,727
	Pickup	Number of trips made by pickup	245.98	347.87	0	2,623	62,233
	MiniBus	Number of trips made by minibus**	198.81	339.52	0	3,225	50,298
	Bus	Number of trips made by bus***	3,876.33	5,593.54	0	77,798	980,711
	BusServ	Number of trips made by bus services****	145.29	230.13	0	1,513	36,758
	Moped	Number of trips made by motorbike	1,729.33	1,644.97	0	10,807	437,521
	Bike	Number of trips made by bike	490.98	613.05	0	5,263	124,218
	Walk	Nimber of trips made by walking	8,998.20	8,592.14	0	58,318	2,276,545

* Number of observations=253 TAZs, **para-transit, ***transit, ****shuttle

Regarding the development of generalized linear model with negative binomial distribution, it was firstly attempted to formulate a database in the IBM SPSS Statistics 23 Software and the initial model was constructed only by a fixed amount. Then, the independent variables were entered into the base model one after another. Then, improvement or lack of improvement in the base model was gauged using significance of coefficients as well as statistical parameters as AIC, Log Likelihood and concerned coefficients of determination. If the introduction of an independent variable into the model improved it, the concerned variable would be maintained in the base model and the next independent variable would enter into the model and, subsequently, improvement or lack of improvement in the base model was examined. This process was repeated for all independent variables as well as different combinations of independent variables and some diverse models were developed thereof. Having compared the models' goodness of fit, the final model was chosen and specified in Table 3.

Table 3. Estimation of Generalized Linear model with Negative Binomial Distribution

Var.	**Coeff.**	**Std Err**	**Sig.**
cons.	4.302	0.0819	0.000
car	0.000111	0.000013	0.000
bus	-0.000321	0.000012	0.008
busserv	0.001	0.0002	0.026
Goodness of fit			
No. of observations	253		
Degree of freedom (df)	244		
Deviance/df	1.12		
α	0.50		
Log-likelihood (LL)	-1452.93		
LL of constant-only model	-1497.93		
LL of Poisson model	-8320.29		
R^2	0.28		

As shown in Table 3, car, bus and busserv were significant in 95% level. Deviance/df amount of the final model is 1.12 which is in the 0.8-1.12 acceptable range. Also the parameter of α has more difference with zero; it means that Negative Binomial distribution is true selection comparing Poisson distribution.

Regarding the development of geographically weighted regression, it was indicated that a similar process was adopted but this time the forenamed procedures were run in ArcMap 10 Software. Similarly, it was attempted to make use of significance of coefficients, AIC, condition number of models and their coefficients of determination in order to determine improvement or lack of improvement in the base model across different phases and compare different models in this regard. The specifications of the final model are depicted in Table 4.

Table 4. Estimation of Geographically Weighted Regression

Var.	**Mean**	**Std Err**	**Min**	**Max**
cons.	57.270	19.928	15.483	125.92
car	0.0131	0.0052	0.0045	0.0230
bus	-0.0022	0.0028	-0.0140	0.0098
Schoolserv	0.0232	0.0171	0.0039	0.0804
Goodness of fit				
Condition No.	6.492	0.637	5.134	7.998
No. of observations	253			
Adjusted R^2	0.40			

As shown in Table 4, car, bus and schoolserv were significant in 95% level. Mean condition number amount of the final model is 6.492 which is very low than goodness limitation. Also minimum and maximum amount of condition number has more difference from that limitation. Most literature such as (ESRI, 2011) suggested that the condition number should be lower than 30 in geographically weighted regression.

5. Discussion

As mentioned in Table 3 and Table 4, it was observed that the trips taken by private cars, bus, bus service and school service were significant at 95 percent across all the finalized models. The fixed amount was purposefully included in this study to represent the impact of trips taken by other vehicles which were not included in the model. According to Table 3 and Table 4, the fixed amount was highlighted by a positive sign. The positive sign indicated that the trips taken by other vehicles had a positive impact on the number of crashes.

It was observed that the trips taken by private cars were highlighted by a positive sign in both finalized models. That meant that more trips by private cars would increase the number of traffic crashes in the study area. Put it simply, increased number of trips taken by private cars may mean increased level of exposure to crash and it is in line with the aforementioned statement in that each trip may be the source of a crash. Factors such as the density of private cars on the streets, the driving skills of drivers of private cars and their physical and mental conditions can be enumerated among the reasons for the positive impact of trips taken by private cars on increased number of crashes.

Trips taken by bus were another variable that significantly appeared in the final models. Interestingly, this variable was embedded with negative coefficient in the both finalized models. The latter finding represented the negative impact of trips taken by bus on the number and volume of traffic crashes in the study area. Assuming all other factors being constant, it was indicated that as the ratio of trips taken by bus increased in a traffic area (in relation to total trips), the number of traffic crashes would decrease in that area.

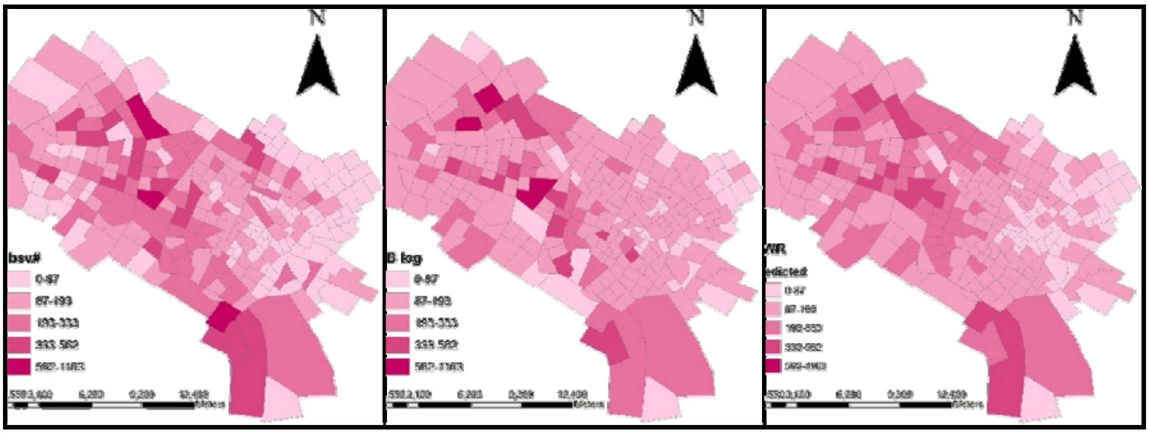
In this study, trips taken by bus services were those trips taken by buses other than those dedicated to urban public transportation system. Therefore, those trips included trips related to agencies, organizations and factories. According to Table 3, trips taken by bus services were embedded with a positive sign in the model. The positive significance of this variable in the final model could be caused by different factors such as the time of the trips, which occur mainly during rush hours, numerous stops and movements of these types of vehicles and (in some cases) potential delays on the part of passengers and the rush of passengers during boarding or landing.

Regarding Table 4, it was indicated that the trips taken by school services lasted a positive impact on the number of crashes occurred in the study area. The positive significance of this variable might be examined from different aspects. Firstly, the passengers of these types of vehicles were mainly students whose age and mental conditions kept them among those individuals who may experience high crash risk. Secondly, the trip time should be considered duly because most of these trips occurred early morning and at noon (exactly in pick hours). Besides, the school services are usually very crowded and it can result in the driver's lack of concentration and, thus, they may be categorized as high-risk trips.

6. Conclusion

The main objective of this research was to make use of crash prediction models in the urban transportation planning process. To do this end, it was attempted to make use of generalized linear model with negative binomial distribution and geographically weighted regression on the basis of results of Mashhad transportation master plan to develop crash prediction model in this city. Accordingly, it was indicated that both models could be efficiently used to accurately predict the number of crashes in Mashhad urban areas (at 95% confidence level).

Figure 3 depicts the number of observed crashes in 2006 in Mashhad as well as the number of predicted crashes on the basis of aforementioned methods. According to this Figure, it is observed that the models obtained in this study have properly predicted the number of crashes in most urban traffic areas in Mashhad.



(a) (b) (c)

Figure 3. 2006 Crash, (a) Observed Crash, (b) NB-log Predicted Crash, (c) Geographically Weighted Regression Predicted Crash

Besides, it has observed that any change in the modal share of different types of urban trips can revolutionize the number of crashes in urban areas. For example, Table 3 and Table 4 indicate that if the number of trips taken by private cars decreases, one can expect to observe decreased number of crashes in different urban areas. The latter achievement may increase transportation safety in cities. Conversely, if the number of trips taken by buses increases, the condition and status of urban transportation safety will be significantly improved. In addition, preparing a detailed plan to change the pattern of use of vehicles (fewer private cars and more buses) will certainly lead to better results.

The application of the process presented in this study can improve the urban transportation safety management processes and lead to more accurate prediction in terms of crashes across urban traffic areas. The application of macro-level modelling in order to predict urban crashes is considered as a major feature of this study. Besides, independent variables have been gauged and urban officials and planners can make use of these parameters accordingly. This process may be developed if the generalized linear models with random distribution and geographically weighted regression models move towards geographically weighted models with Poisson and negative binomial distributions.

Acknowledgment

This paper is part of an ongoing research on PhD dissertation at Faculty of Civil Engineering, Semnan University, Semnan, Iran. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. Cameron, A.C. and Trivedi, P. K. (1998) *Regression Analysis of Count Data.* New York: Cambridge University Press.
2. de Guevara, F.L., Washington, S.P., Oh, J. (2004) Forecasting Crashes at the Planning Level: A Simultaneous Negative Binomial Crash Model Applied in Tucson, Arizona. *Transportation Research Record 1897, TRB,* National Research Council, Washington, D.C.
3. ESRI. (2011) ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Instiude.
4. Hadayeghi, A., Shalaby, A. S., Persaud, B. (2003) Macro-Level Accident Prediction Models for Evaluating Safety of Urban Transportation Systems. *Transportation Research Record 1840, TRB,* National Research Council, Washington D.C.
5. Hadayeghi, A., Shalaby, A.S., Pesuad, B. (2007) Safety Prediction Models: A Proactive Tool for Safety Evaluation in Urban Transportation Planning Applications. *Transportation Research Record, TRB,* National Research Council, Washington D.C.
6. Huang, H., Song, B., Xu, P., Jae, Q., Lee, J., Abdel-Aty, M. (2016) Macro and micro models for zonal crash prediction with application in hot zones identification, *Journal of Transport Geography*, 54, pp. 248–256.
7. IBM Corp. (2014) IBM SPSS Statistics 23. - www.vsni.co.uk.
8. MTTO. (2009) *Comprehensive Transportation Studies of Mashhad City, update.* Mashhad, Iran: Mashhad Transportation and Traffic Organization (MTTO).
9. Naderan, A. (2010) *Estimating Crash Frequencies in Urban Districts using Aggregate (Macro) Models.* Tehran, Iran: Ph.D. Thesis in Iran University of Science and Technology.
10. Naderan, A., Shahi, J. (2010) Aggregate crash prediction models: Introducing crash generation concept. *Accident Analysis and Prevention,* 42 (1), 339-346.
11. Quddus, M.A. (2008) Modelling area-wide count outcomes with spatial correlation and heterogeneity: An analysis of London crash data. *Accident Analysis and Prevention 40 (2008)* 1486–1497.
12. Tasic, I., Porter, R.J. (2016) Modeling spatial relationships between multimodal transportation infrastructure and traffic safety outcomes in urban environments. *Safety Science,* 82 (2016), 325-337.
13. Wang, J., Huang, H., Zeng, Q. (2017) The effect of zonal factors in estimating crash risks by transportationmodes: Motor vehicle, bicycle and pedestrian. *Accident Analysis and Prevention*, 98, pp. 223-231.
14. Washington, S., Schalwyk, I.V., Meyer, M., Dumbaugh, E., Zoll, M. (2006) *Incorporating Safety into Long-Range Transportation Planning.* NCHRP Report 546, TRB, National Cooperative Highway Research Program,Washington D.C.
15. Wier, M., Weintraub, J., Humphreys, E.H., Seto, E., Bhatia, R. (2009) An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. *Accident Analysis and Prevention,* 41 (1), 137–145.