

## Multi-Level Crash Prediction Models Considering Influence of Adjacent Zonal Attributes

Nemat Soltani <sup>a</sup>, Mahmoud Saffarzadeh <sup>b\*</sup>, Ali Naderan <sup>c</sup>

<sup>a</sup> Ph.D. Candidate, Tarahan Parseh Transportation Research Institute, Tehran, Iran.

<sup>b</sup> Professor, Department of Civil & Environmental Engineering, Tarbiat Modares University, Tehran, Iran.

<sup>c</sup> Assistant Professor, Department of Civil Engineering, Islamic Azad University, Science and Research Branch, Tehran, Iran.

Received 20 November 2018; Accepted 12 February 2019

### Abstract

This study investigates factors affecting accidents across transport facilities and modes, using micro and macro levels variables simultaneously while accounting for the influence of adjacent zones on the accidents occurrence in a zone. To this end, 15968 accidents in 96 traffic analysis zones of Tehran were analyzed. Adverting to the multi-level structure of accidents data, the present study adopts a multilevel model for its modeling processes. The effects of the adjacent zones on the accidents which have occurred in one zone were assessed using the independent variables obtained from the zones adjacent to that specific zone. A Negative Binomial (NB) model was also developed, and results show that the multilevel model that considers the effect of adjacent zones shows a better performance compared to the multilevel model that does not consider the adjacent zones' effect and NB model. Moreover, the final models show that at intersections and road segments, the significant independent variables are different for each mode of transport. Adopting a comprehensive approach to incorporate a multi-level, multi-resolution (micro/macro) model accounting for adjacent zones' influence on multi-mode, multi-segment accidents is the contribution of this paper to accident studies.

*Keywords:* Multi-Level Model; Adjacent Zone; Crash Frequency; Micro/Macro Variable.

### 1. Introduction

Accidents are and have always been regarded as one of the sad consequences of transportation systems. In 2015, 19.9 people out of each 100000 have died in accidents in Iran; compared to 5.1 casualties in Europe [1, 2]. Hence, the necessity of paying more attention to transportation safety and carrying out pertinent investigations seems unavoidable. Since the nature and mechanism of accidents varies across diverse transportation facilities, it is essential to run separate investigations on the accidents of each mode. The studies done over the recent years have mostly addressed accidents of all modes together or have considered only a single mode (vehicle, motorcycle and pedestrian) [3]. This has led to an inaccurate understanding of the factors affecting accidents since one factor might increase accidents in one mode while decreasing them in the other. Therefore, the significance of reaching an accurate understanding of the factors leading to accidents necessitates separate investigation of accidents across different transportation modes and facilities.

Most of the previous studies on transportation safety have addressed micro-level factors which are related to accidents such as road geometry or the road lighting quality. Most of them are carried out at the operation time of transport

\* Corresponding author: [saffar\\_m@modares.ac.ir](mailto:saffar_m@modares.ac.ir)

 <https://dx.doi.org/10.28991/cej-2019-03091276>

➤ This is an open access article under the CC-BY license (<https://creativecommons.org/licenses/by/4.0/>).

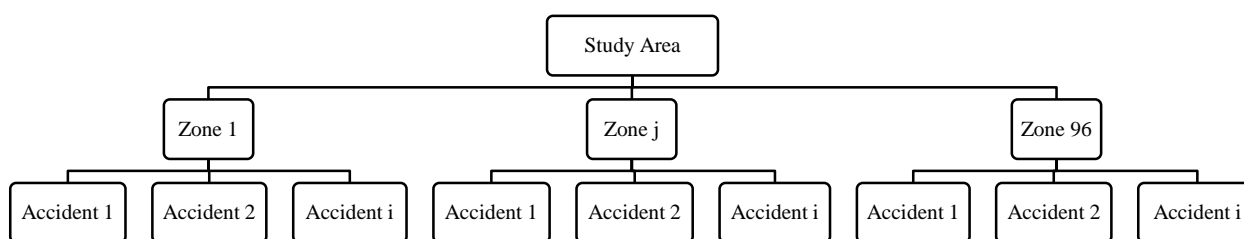
© Authors retain all copyrights.

facilities and are based on the data related to the existing accidents. Then, some engineering solutions are proposed based on interpretation of model results [4]. On the other hand, some researchers have recently investigated the factors affecting accidents at a macro level. These factors, for example economic and demographic variables, have been considered at different geographical levels such as Traffic Analysis Zones (TAZs) and census blocks. The development of accident prediction models based on macro-level variables has increased the amount of attention paid to safety in studies related to road network planning. These models result in safety enhancements prior to the operation of transportation facilities.

To get an accurate understanding of the factors leading to accidents, both micro-level and macro-level factors need to be attended to simultaneously. That's why a number of studies have recently addressed the factors affecting accidents while considering the factors at both micro and macro levels simultaneously [5]. The present research has investigated both micro and macro variables across different types of facilities (intersections and road segments) and modes of transport (vehicle, motorcycle and pedestrian). The macro-level variables used in this study were collected at TAZ-level. The use of TAZs is more common than other geographical levels (e.g. census block) because zone divisions are more in line with the studies related to transport planning models and variables (e.g. trip generation and trip distribution) are more readily accessible.

Since in the present study macro-level variables are extracted at TAZ-level, they are the same for the accidents which have occurred at the intersections and road segments in zones. Therefore, the present study has adopted a multilevel model for investigating the amount of intra-zonal correlation resulted from similar macro variables. Multilevel models are more suitable for sets of data which are multilevel and in which low-level data are nested in higher-level data [6].

In this study, the data were categorized in two levels. The first level accommodated the micro-level variables related to each accident and the second level included the macro-level variables related to TAZs. The hierarchical structure of the database is shown in Figure 1.



**Figure 1. Hierarchical structure of the variables**

On the other hand, considering the same macro-level variables for the accidents which have occurred in a single TAZ but in different situations might lead to an inaccurate understanding of the factors affecting accidents. This is because the occurrence of some of them might have been affected by the factors available in adjacent TAZs. For instance, an accident which has occurred at the borders of a TAZ can be more affected by the factors available in adjacent TAZs than the factors of its own zone.

There are two options available for measuring the influence of adjacent TAZs on the accidents which have occurred in a TAZ. The first option is through Spatial Error Correlation Effects in which the unobserved exogenous variables available in one zone exert influence on the dependent variables which exist in the target as well as adjacent zones [7, 8]. The second method is carried out by means of Spatial Spillover Effects in which the observed exogenous variables which are available in one zone affect the dependent variables existing in both the target zone as well as the adjacent zones [9, 10, and 11].

In this study, to evaluate the influence of neighboring zones on the accidents of one zone through spatial spillover effects procedure, the adjacent TAZs of one zone were identified. Then, some new variables were determined based on the amounts of this variable in the neighboring TAZs. The final models were developed using these variables.

Considering the above-mentioned, the objectives of the present study can be summarized as follows:

1. Investigating the factors affecting accidents across transport facilities and modes;
2. Considering the factors affecting accidents at both micro and macro levels simultaneously and measuring the influence of each;
3. Investigating and measuring the intra-zone correlation effect using a multilevel model;
4. Studying the influence of zones adjacent to a TAZ on the accidents which have occurred in that TAZ.

Figure 2 shows different steps followed in this research. The next section reviews related studies in this regard. Section 3 addresses data collection and Section 4 explains the methodology of the study. Next, the results obtained from the final models are presented followed by the conclusion.

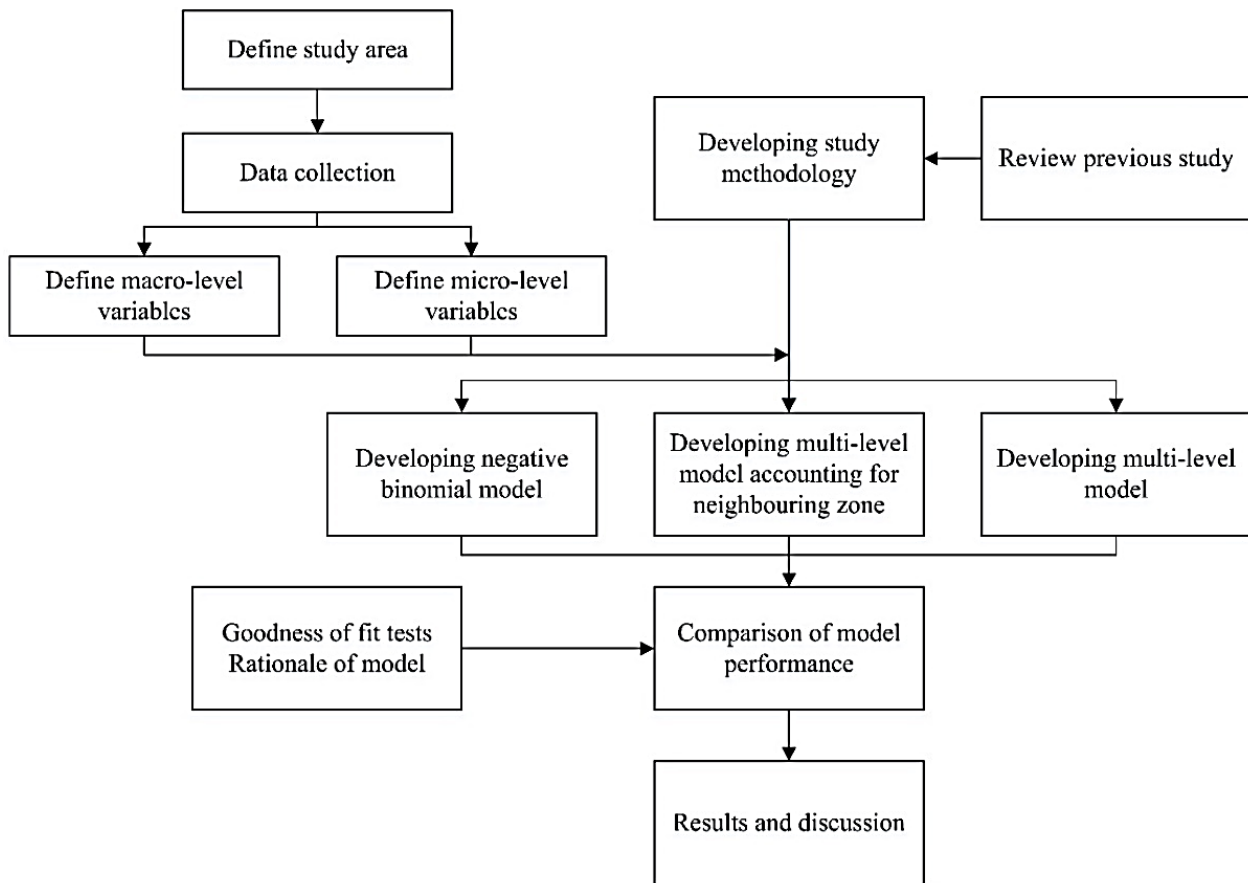


Figure 2. Research methodology

## 2. Literature Review

Finding the suitable method of analysis and selecting influential independent variables are two factors that affect the development of safety models. In the past years, researchers have proposed numerous methods for developing accident prediction models using different variables. The details of these methods and their results are presented in review papers [12, 13]

The accident prediction models are mostly developed using micro-level variables [14-16]. These studies have helped to identify the factors affecting the accidents at this level and determine solutions for decreasing the number of accidents in different transport facilities like intersections or road segments. On the other hand, extensive efforts have been made in recent years to develop accident prediction models using macro-level variables. These variables include roads length with different functional classification in a zone [17, 18], and trip generation and trip distribution data for each TAZ [19]. They also include environmental conditions like land use specifications [20], and socioeconomic factors like household income [21]. The results of these researches have led to the consideration of safety indices in road network planning.

Cai et al, investigated the influence of macro-level variables at TAZ-level on pedestrians and cyclists' accidents using Dual-State models. In their study, they were also trying to measure the influence of the neighboring zones on the accidents of one zone. According to their results, some factors like population density, employment rate, and the number of public transport users in one TAZ increase the number of accidents. Moreover, the influence of adjacent zones on the accidents of one zone turned out to be significant and Dual-State models, especially Zero-Inflated Negative Binomial model, showed a better performance in comparison to Single-State models [9].

To reach an accurate understanding of the factors affecting accidents, it seems necessary to consider suitable variables at both micro and macro levels simultaneously and develop appropriate models. Although extensive researches have been carried out on investigating the factors affecting accidents at micro and macro levels, few studies have so far merged the data at both levels to develop accident prediction models.

Huang et al, developed accident prediction models at micro and macro levels and compared the performance of these models in predicting hot zones. Based on this study, accident prediction models were developed for TAZs using macro-level variables, and for intersections and road segments using micro-level variables. The results of this research indicated

that models have a better performance at micro level and present a better picture of the micro variables affecting traffic accidents. Whereas, for investigating safety at TAZs, using accident models at macro level is more fruitful because less detailed data are needed here [22].

Guo et al, developed accident prediction models for signalized intersections based on variables at micro and macro levels. Based on the results of this study, the researchers found that at corridor level, Poisson models have a better performance in comparison to other models [23]. Mitra and Washington, investigated the effect of variables at micro level like Annual Average Daily Traffic (AADT), and at macro level like population across different age groups, total population and the number of schools in zones adjacent to intersections on the number of accidents. Based on the results of this study, the amount of population in 16-64 age groups and the annual average rainy days have a significant effect on the occurrence of accidents. Moreover, through comparing the models developed based on traffic parameters and the models developed based on all variables, the researchers found out that the omission of macro-level variables can significantly increase the effect of other factors like AADT [24].

In this study, the variables are considered at both micro and macro levels. Since the macro-level variables available for the accidents in a TAZ are the same, the structure of data in this study is multilevel. A multi-level data constitutes of correlation among observations and inter-group independence where lower-level data are nested in the higher-level data. When accident data is multilevel, using multilevel models, which account for intra-group correlation of accidents data, is beneficial [6]. Detailed information about multilevel data and the adoption of multilevel models in studies related to safety can be found in [25].

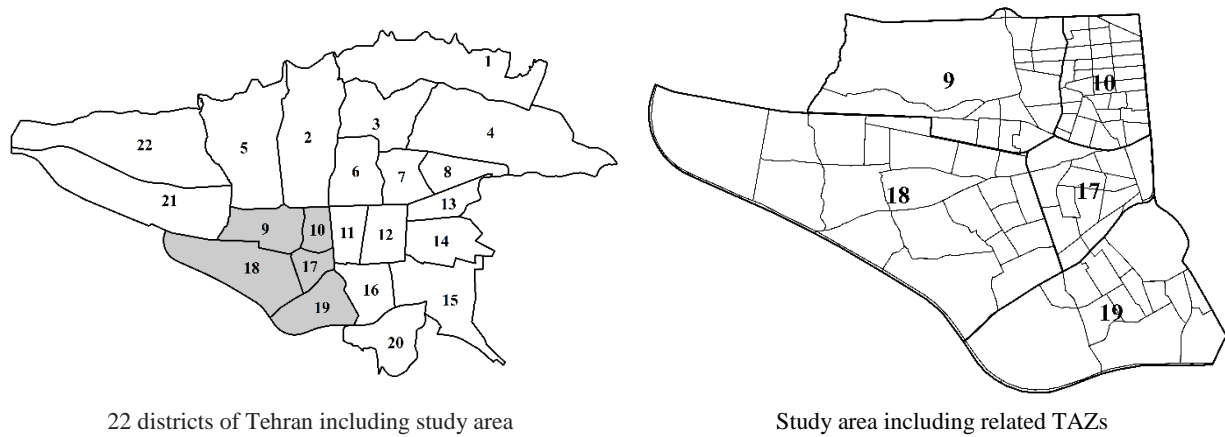
Huang and Abdel-Aty, adopted a five-level structure (geographic region level, traffic site level, traffic crash level, driver-vehicle unite level and occupant level) as the general structure of accidents data. In this study, macro-level analysis is considered based on the three high levels of geographic region level, traffic site level, and traffic crash level; and micro-level analysis has considered the three low levels of traffic crash level, driver-vehicle unit level, and occupant level. The authors in this study have proposed different methods for investigating multilevel data including analysis of accidents data at intersections and time level [26].

Shi et al, investigated the number of highway accidents using multilevel and Negative Binomial (NB) models. In their research, the highway was divided into 196 segments based on its geometrical specifications. The traffic data were obtained through Automotive Vehicle Identification (AVI) systems installed in the highway. Since the output data of AVI systems divided the highway into 43 segments, each AVI system represented the data related to some segments. Due to the dual-level structure of the data, a multilevel model was used for investigating traffic accidents. Results show that the multilevel model had a better performance than NB model. Moreover, some factors such as the increase of speed or the increase of the horizontal degree of curvature decrease the number of accidents [27].

Considering the information presented above, in this study, both micro-level and macro-level independent variables for the accidents which have occurred at intersections and road segments are collected across transport modes (vehicle, motorcycle, and pedestrian) so that a comprehensive investigation can be carried out. Besides, the performance of multilevel models in estimating the number of traffic accidents was evaluated as well. Finally, the influence of the adjacent zones on the accidents were investigated.

### 3. Data Collection

In the present study, for developing accident prediction models using micro and macro variables, accident data over the years 2014 and 2015, were collected for the west and the south west of Tehran, Iran. In general, data related to 15968 accidents (1231 accidents occurring at intersections and 14737 in road segments) which have occurred in 360 intersections and 892 road segments were collected. Tehran, as the capital city of Iran, has 5 main areas which in total comprise 22 districts. The west and the south west main areas are composed of districts no. 9, 10, 17, 18, and 19 that include 96 TAZs. The accident data was obtained through the database available in Tehran Traffic Police Center, and the demographic data, i.e. population, education and employment were obtained through Iran National Census Center. Besides, the traffic data was collected through Transportation and Traffic Organization of Tehran Municipality and based on the results obtained from running Tehran traffic model. After collecting the required data, all information was imported to the GIS application. Then the traffic, social, and demographic data related to each accident was calculated. Figure 3 shows the districts under study along with the TAZs in those districts.

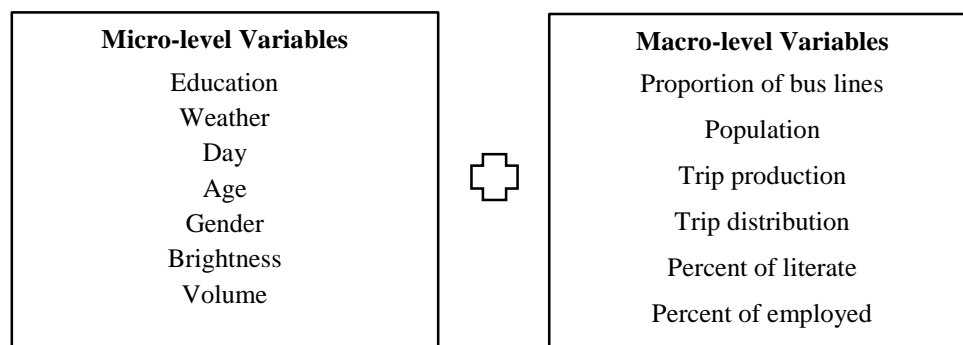


**Figure 3. Location of study area**

The independent variables are shown in figure 4 and table 1 and table 2 list the variables used in this study at micro and macro levels at intersections and in road segments respectively across modes of transport along with their descriptive statistics.

Since in regression models there is usually a logarithmic relation between independent variables and the response variable, using the logarithm of independent variables in the modeling process makes interpretation of the results much easier. This is also very common in the previous studies [28, 29]. Moreover, this method also decreases variance among variables [17, 30]. Hence, the present study uses logarithmic conversion of the variables related to the population and trip generation, trip distribution of TAZs and traffic volume in road segment.

As already mentioned, to evaluate the effect of adjacent zones on the accidents occurring in a TAZ, all TAZs adjacent to that TAZ were identified. Then, a new variable based on the value of the each independent variable from surrounding TAZs was obtained. These variables capture the effect of neighboring TAZs on crash frequency in one TAZ. In Table 3 you can find a descriptive summary of the variables extracted from adjacent TAZs.



**Figure 4. Independent variables at micro and macro level**

**Table 1. Descriptive summary of variables for the intersection models**

Mode of Transportation		Vehicle				Motorcycle				Pedestrian			
Variable	Definition	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD
Edu	Education of driver; with university education= 0, without university education= 1	0.00	1.00	0.83	0.38	0.00	1.00	0.63	0.48	0.00	1.00	0.79	0.41
Weather	Sunny days= 0, rainy and snowy= 1	0.00	1.00	0.02	0.15	0.00	1.00	0.02	0.13	0.00	1.00	0.03	0.17
Day	Weekday= 0, weekend day= 1	0.00	1.00	0.30	0.46	0.00	1.00	0.29	0.45	0.00	1.00	0.25	0.44
Age	Age of driver	17.00	75.00	35.77	13.47	16.00	81.00	31.51	15.27	11.00	75.00	30.51	16.02
Gender	Man= 0, woman= 1	0.00	1.00	0.10	0.30	0.00	1.00	0.06	0.23	0.00	1.00	0.11	0.31
Bri	Daylight= 0, night= 1	0.00	1.00	0.28	0.45	0.00	1.00	0.29	0.45	0.00	1.00	0.27	0.44
Pro-bus	Proportion of bus lines per road lines in TAZ	0.00	10.04	1.63	1.58	0.00	10.04	1.87	1.79	0.00	10.04	1.59	1.51
Pop	Population of TAZ	0.00	39688.00	15761.33	7467.20	0.00	39688.00	16473.84	7970.85	0.00	33856.00	16383.91	7009.66
T-dis	Trip distribution of TAZ	2469.08	45787.48	13732.53	6828.73	2469.08	45787.48	13540.96	6361.23	5066.68	30838.77	14134.54	6597.04
T-gen	Trip generation of TAZ	2368.48	29483.32	14677.16	5536.39	2368.48	29483.32	15111.75	6117.71	2368.48	29483.32	15418.33	5716.89
Lit	Percent of literate of TAZ	0.00	97.61	88.76	14.19	0.00	97.61	88.07	13.83	0.00	94.97	88.84	12.66
Emp	Percent of employed of TAZ	0.00	96.16	84.16	13.32	0.00	96.16	84.23	13.11	0.00	93.20	84.43	11.84

**Table 2. Descriptive summary of variables for the road segments models**

Mode of Transportation		Vehicle				Motorcycle				Pedestrian			
Variable	Definition	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD
Edu	Education of driver; with university education= 0, without university education= 1	0.00	1.00	0.81	0.39	0.00	1.00	0.66	0.47	0.00	1.00	0.71	0.46
Weather	Sunny days= 0, rainy and snowy= 1	0.00	1.00	0.02	0.15	0.00	1.00	0.00	0.07	0.00	1.00	0.02	0.14
Day	Weekday= 0, weekend day= 1	0.00	1.00	0.27	0.44	0.00	1.00	0.26	0.44	0.00	1.00	0.24	0.43
Age	Age of driver	16.00	86.00	35.42	14.45	12.00	90.00	31.45	15.62	7.00	81.00	30.41	18.02
Gender	Man= 0, woman= 1	0.00	1.00	0.07	0.25	0.00	1.00	0.04	0.19	0.00	1.00	0.06	0.25
Bri	Daylight= 0, night= 1	0.00	1.00	0.23	0.42	0.00	1.00	0.26	0.44	0.00	1.00	0.29	0.45
Vol	Traffic count per segment base on peak hour	1.00	23445.00	9501.67	7082.57	1.00	23445.00	5821.36	5901.79	1.00	23445.00	5212.70	5423.44
Pro-bus	Proportion of bus lines per road lines in TAZ	0.00	15.79	2.07	2.94	0.00	15.79	2.11	2.83	0.00	15.79	1.87	2.26
Pop	Population of TAZ	0.00	39688.00	16897.35	9178.58	0.00	39688.00	15730.87	7893.03	0.00	39688.00	15826.84	7738.84
T-dis	Trip distribution of TAZ	2469.08	45787.48	13614.77	8028.87	2469.08	45787.48	13776.10	7862.80	2469.08	45787.48	14167.46	8572.22
T-gen	Trip generation of TAZ	1211.37	29483.32	15066.68	5769.00	1211.37	29483.32	14551.61	5780.43	1211.37	29483.32	14663.78	5683.04
Lit	Percent of literate of TAZ	0.00	97.61	89.10	9.76	0.00	97.61	89.24	11.62	0.00	97.61	89.72	10.29
Emp	Percent of employed of TAZ	0.00	96.16	85.97	9.21	0.00	96.16	85.33	11.05	0.00	96.16	85.79	9.77

**Table 3. Descriptive summary of the variables extracted from adjacent TAZs**

Mode of Transportation	Vehicle				Motorcycle				Pedestrian			
Variable	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD
<b>Intersection</b>												
Pro-bus	0.46	7.61	2.19	1.47	0.56	6.7	2.31	1.45	0.56	6.7	2.23	1.58
Pop	2785.5	24900.5	15511.05	4942.23	2785.5	24900.5	15993.28	4399.8	8864.75	23486.8	16031.7	4693.3
T-dis	5457.03	21660.49	12379.58	4072.01	5283.8	21660.49	12623.56	3556.65	5174.75	21660.49	12514.62	4111.84
T-gen	4597.51	20922.76	14025.84	3682.42	4597.51	20922.76	14256.96	3189.52	8472.48	20922.76	14283.63	3400.38
Lit	85.03	94.8	90.39	2.97	85.03	94.8	89.72	2.67	85.44	94.8	90.2	2.85
Emp	83.94	88.49	86.34	1.19	83.94	88.49	86.57	1.23	83.94	88.49	86.39	1.14
<b>Segment</b>												
Pro-bus	0.34	7.61	2.35	1.66	0.34	7.61	2.2	1.48	0.34	7.61	2.22	1.56
Pop	2785.5	24900.5	15572.84	5054.69	2785.5	24900.5	14925.17	4872.91	2785.5	24900.5	14774.07	4908.31
T-dis	3970.85	21660.49	12036.6	3590.86	3970.85	21660.49	12050.12	3895.83	3970.85	21660.49	12008.03	4045.11
T-gen	4597.51	20922.76	13815.47	3552.36	4597.51	20922.76	13494.19	3577.26	4597.51	20922.76	13474.49	3548.93
Lit	85.03	94.8	89.64	2.74	85.03	94.8	90.15	2.58	85.03	94.8	90.31	2.62
Emp	83.94	88.49	86.45	1.37	83.94	88.49	86.27	1.24	83.94	88.49	86.27	1.24

## 4. Methodology

Poisson model is a type of statistical model which, due to the random, non-negative and sporadic nature of accidents data, has had remarkable and successful applications. One of the fundamental assumptions of this model is equality of accidents' mean frequency and variance. To consider the over-dispersion of accidents data, an NB model would be adopted. By adding gamma-distributed error term to the average available in Poisson model, this model considers the over-dispersion available in accidents data and thus is preferred over Poisson model.

The formula for the NB model is presented in the following equations:

$$Y_i \sim \text{Poisson}(\lambda_i) \quad (1)$$

$$\ln \lambda_i = \beta_0 + \beta X_i + \varepsilon_i \quad (2)$$

Where:

$Y_i$ : Represents the crash frequency by modes at intersection  $i$  or road segment  $i$ ;

$\lambda_i$ : Shows the expectation of  $Y_i$ ;

$X_i$ : Indicates a vector of explanatory variables;

$\beta_0$ : The intercept;

$B$ : The vector of estimable parameters;

$\varepsilon_i$ : Represents the error term which is considered to be independent  $X$  and has a two-parameter gamma distribution.

One of the main assumptions in NB models is the independence of observations. However, it is hardly possible in practice to consider accidents independent from one another. For instance, the accidents occurring in one area might have unobserved common factors [12]. To enhance accidents models for intersections and road segments and to consider the correlation among accidents occurring in one zone due to their common macro-level variables, the present model adopted a multilevel modelling approach.

The general equation for the single-level model or the conventional simple regression model is as follows:

$$y_i = \beta_0 + \beta_1 X_{1i} + e_i \quad (3)$$

In the above equation, the subscript  $i$  represents an individual respondent,  $y$  and  $x$  stand for the dependent and independent variables respectively. There are also two fixed parameters ( $\beta_0$  and  $\beta_1$ ) that show the intercept and the slope, and a random part ( $e$ ) that makes it possible to have fluctuations around the fixed part. The word "random" here means "allowed to vary".

The micro-level of the individual is the sole place where this equation is specified. For developing a multilevel model, this micro-model needs to be re-specified through differentiating TAZs with the subscript  $j$ . This provides the following for the random intercept and random slope model:

$$y_{ij} = \beta_{0j} + \beta_{1j} X_{1ij} + e_{ij} \quad (4)$$

At TAZ-level, two macro models exist:

$$\beta_{0j} = \beta_0 + u_{0j} \quad (5)$$

$$\beta_{1j} = \beta_1 + u_{1j} \quad (6)$$

The first macro-model allows for different TAZ-level intercept ( $\beta_{0j}$ ) to change from one TAZ to another around the overall intercept ( $\beta_0$ ) through the addition of random component  $u_{0j}$ . The second macro-model allows for differential slope ( $\beta_{1j}$ ) to change around the overall slope ( $\beta_1$ ) through the addition of random component  $u_{1j}$  [31].

Once more, the micro model is regarded as an intra-zonal equation, whereas the macro models are between zonal equations where the parameters of the intra- model are the responses.

It is worthy of notice that when the notation is used with  $e_{ij}$  as a part of the micro model as opposed to the macro model for in that case just the micro-model includes both subscripts  $i$  and  $j$ , and this demonstrates a within situation, whereas the macro-model in that case just includes subscript  $j$ , which demonstrates a between situation. The completely random two-level model includes a combination of all three equations:

$$y_{ij} = \beta_0 + \beta_1 X_{1ij} + (u_{1j} X_{1ij} + u_{0j} + e_{ij}) \quad (7)$$

The best accident model for intersections and road segments in each mode of transport was chosen based on three criteria, namely log-likelihood, Akaike's Information Criterion Corrected (AICC) and Bayesian Information Criterion (BIC). What follows are the formulae for this measure:

$$\text{AICC} = 2k - 2\text{LL}(\text{full}) + 2k(k+1) / (n - k - 1) \quad (8)$$

$$\text{BIC} = k \ln(n) - 2\text{LL}(\text{full}) \quad (9)$$

In the above formulae  $k$  represents the number of parameters,  $n$  indicates the number of observations, and  $LL(\text{full})$  shows the log-likelihood for the full model.

### 5. Results and Discussion

Accident prediction models were developed across intersections and road segments and for three modes of transport (vehicle, motorcycle, and pedestrian). After developing numerous models including 3 types of NB models, multilevel models with and without neighboring zones' effect, 18 final models were created in total. To compare the performance of the multilevel model with that of the NB model and to find out how much influence is exerted by the neighboring zones on the accidents occurring in one zone, some comparisons were made among the final models based on the criteria Model Goodness of Fit, Log-Likelihood, AICC, and BIC. The summary of model performances for models of intersections and road segments are given in Table 4 for each mode of transport (vehicle, motorcycle, and pedestrian).

Results show that multilevel models which take into account the effect of neighboring zones have a better performance than other models. Since the independent variables used in the modeling process proved significant in some modes of transport yet insignificant in some others, the investigation of factors affecting accidents across different facilities and in different modes of transport seems necessary. This result agreed with the safety research that the sets of significant variables in crash frequency analysis differed for different transportation modes [32]. Moreover, multilevel models can better estimate the number of accidents for they consider the multilevel structure of the data. Considering the effect of the neighboring zones on the accidents occurring in a TAZ has a significant effect on both model performance and the results of model goodness of fit.

Table 5 lists the coefficients and average marginal effects of the significant variables ( $P\text{-value} < 0.05$ ) in six final models. Based on the final models the following results were achieved. Figure 5 shows spatial distribution of modeled vs. observed accidents per TAZ by transportation modes.

**Table 4. Summary of model performances**

Type of Model	Multilevel			Multilevel-ENZ*			NB		
	Vehicle	Motorcycle	Pedestrian	Vehicle	Motorcycle	Pedestrian	Vehicle	Motorcycle	Pedestrian
<b>Intersection</b>									
Akaike Information Criterion Corrected (AICC)	1049.89	772.03	328.59	990.82	731.47	299.68	1071.48	822.87	388.61
Bayesian Information Criterion (BIC)	1097.28	816.42	361.62	1057.39	790.71	343.67	1115.6	864.24	419.61
- Log Likelihood	510.04	370.92	148.03	473.56	344.7	125.01	521.96	397.49	179.37
<b>Road Segment</b>									
Akaike Information Criterion Corrected (AICC)	5206.64	3320.76	2922.14	5155.32	3268.69	2872.09	5242.62	3404.84	3028.71
Bayesian Information Criterion (BIC)	5276.1	3387.78	2988.22	5252.23	3362.12	2964.18	5307.48	3467.43	3090.43
- Log Likelihood	2588.01	1645.01	1445.7	2556.06	1612.64	1414.29	2607.04	1688.1	1500.01

\*Multilevel models which take into account the Effect of Neighboring Zones

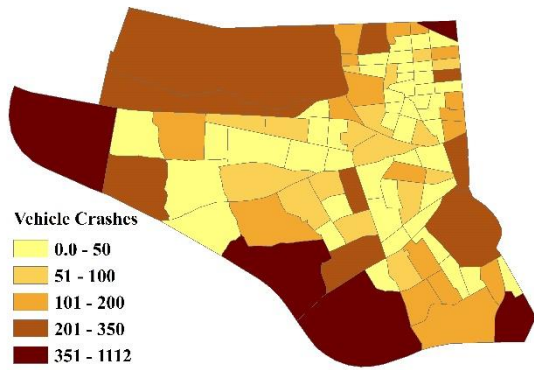
**Table 5. Accident prediction models at intersection and road segments by mode of transportation**

Modes of Transportation	Vehicle			Motorcycle			Pedestrian			
	Variable	Coefficient	P-value	Average marginal effect	Coefficient	P-value	Average marginal effect	Coefficient	P-value	Average marginal effect
<b>Intersection</b>										
Intercept	-9.7	0.00	0.00	-11.72	0.00	0.00	-7.32	0.031	0.00	
Age	-0.05	0.00	0.95	-	-	-	-	-	-	
Pro-bus	-	-	-	-1.42	0.04	0.24	-	-	-	
Log-Pop	-	-	-	-	-	-	2.3	0.01	9.97	
Lit	0.034	0.00	1.03	-	-	-	-	-	-	
Emp	0.071	0.03	1.07	-	-	-	-	-	-	
Log-Pop*	-	-	-	-	-	-	2.12	0.00	8.33	
Log-T-gen*	4.37	0.00	79.04	-	-	-	-	-	-	
Lit*	-	-	-	-	-	-	0.086	0.03	1.09	
Emp*	-	-	-	0.248	0.00	1.28	-	-	-	
<b>Road Segment</b>										
Intercept	-5.1	0.00	0.01	-	-	-	-8.87	0.00	0.00	

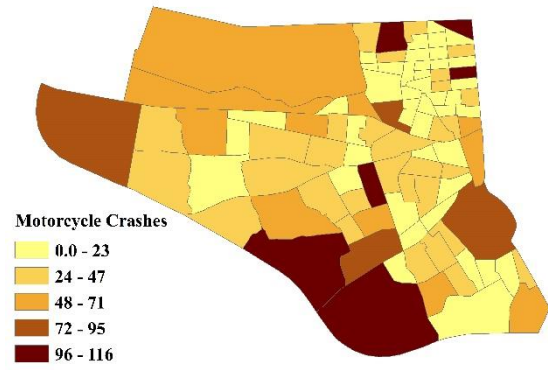


Edu	-	-	-	-	-	-	0.068	0.00	1.07
Weather	-	-	-	-0.486	0.00	0.62	-	-	-
Log-Vol	1.11	0.00	3.03	0.83	0.004	2.29	0.431	0.00	1.54
Pro-bus	-0.31	0.00	0.73	-0.018	0.00	0.98	-0.009	0.00	0.99
Log-Pop	-1.53	0.02	0.22	0.34	0.00	1.4	-	-	-
Log-T-gen	2.3	0.00	9.97	-	-	-	0.65	0.048	1.92
Lit	-	-	-	0.15	0.02	1.16	0.013	0.00	1.01
Emp	-	-	-	0.084	0.00	1.09	0.042	0.034	1.04
Log-Pop*	0.98	0.023	2.66	-	-	-	-	-	-
Lit*	0.034	0.0	1.03	0.017	0.015	1.02	0.063	0.00	1.07
Emp*	0.15	0.00	1.16	0.08	0.00	1.08	-	-	-

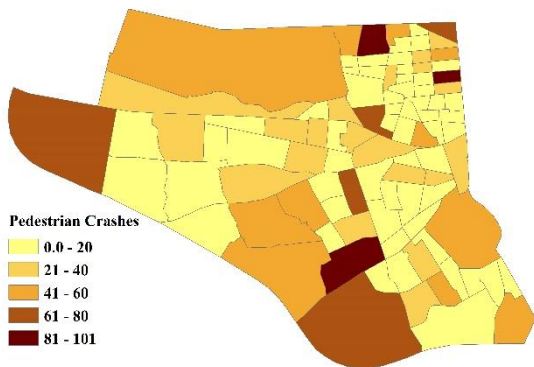
\*: exogenous variable from neighboring TAZs



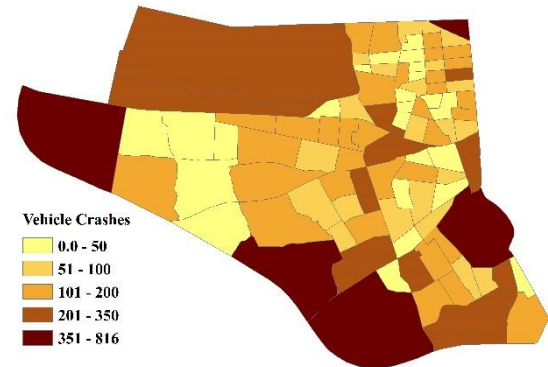
a. Spatial distribution of observed vehicle accidents



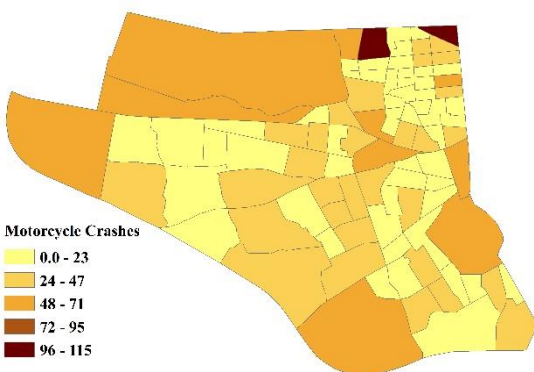
b. Spatial distribution of observed motorcycle accidents



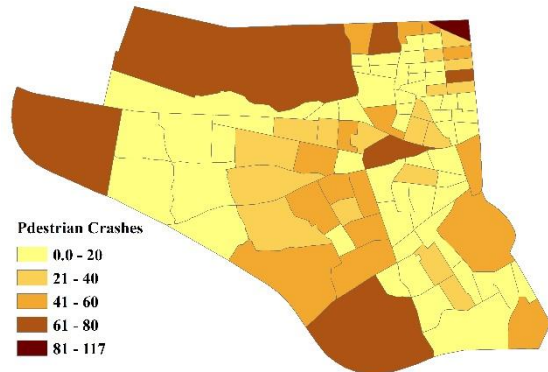
c. Spatial distribution of observed pedestrian accidents



d. Spatial distribution of modelled vehicle accidents



e. Spatial distribution of modelled motorcycle accidents



f. Spatial distribution of modelled pedestrian accidents

Figure 5. Spatial distribution of accidents by transportation modes

### 5.1. Micro Variables

Based on the final model results for vehicle accidents at intersections, higher age of drivers leads to less accidents. Since older drivers are more experienced and practice more caution in driving, the number of their accidents would decrease. This result is similar to the findings of Kazazi et al, which concluded the older drivers had fewer accidents because of their more cautious behavior [33].

Drivers' level of education is negatively correlated with the number of pedestrian accidents in road segments, i.e. as the level of education increases among drivers, fewer pedestrian accidents occur in road segments. Since more educated people observe the rules more meticulously, the number of accidents, quite naturally, decreases. It is consistent with previous study that drivers who had a lower education level had more accidents [34].

Rainy and snowy weather decreases the number of motorcycle accidents in road segments. This variable does not show a significant change in other modes of transport and for the accidents occurring in intersections. This significant change can be explained by the fact that when it rains or snows, the use of motorcycle decreases which naturally leads to fewer accidents for this mode of transport in such weather conditions. Mitra and Washington, showed that the annual average number of rainy days had negative relationships with crash occurrence at intersection. They explained that is because of decreased driving population during the rainy season in the sample data [24]. On the other hand, some previous studies found that rainy weather increases crashes [35]. In this case, further studies and consideration of all related factors such as friction factor of road surface is required.

Higher volumes of traffic leads to more accidents in road segments in all three modes of transport (vehicle, motorcycle, pedestrian), which is consistent with previous studies [36-38]. Obviously, such an increase in traffic increases the amount of activity on roads which in turn results in a higher likelihood of accidents.

### 5.2. Macro Variables

Higher ratios of bus lines to roads length in a TAZ leads to fewer motorcycle accidents at intersections and fewer accidents in road segments for all modes of transport. Since increasing the number of bus lines leads to increasing public transportation facilities in one zone which in turn encourages people to use these facilities, the traffic of vehicles and motorcycles would naturally decrease in that zone. This lower volume of traffic also results in lowering the possibility of accidents (i.e. exposure).

Some researchers found that population is statistically a significant variable to predict crashes [39, 24, 5]. Lee et al, observed that a higher population density had a propensity to increase pedestrian and bicycle crashes in intersection [5]. In this study, population increase in one TAZ also increases the number of pedestrian accidents in intersections and the number of motorcycle accidents in road segments. Moreover, population increase decreases the number of vehicle accidents in road segments. Since the study area is a densely populated urban area, the road network is mainly composed of collectors and due to the low capacity of these roads, the volume of vehicle traffic is not much. Therefore, in densely populated areas, due to their roads network structure and the composition of the passing traffic, the number of pedestrian and motorcycle accidents increases and that of vehicles decreases. Population increase in the neighboring zones of one TAZ leads to more pedestrian accidents in intersections and vehicle accidents in road segments. Since population increase leads to more people moving about in that zone as well as in the neighboring zones, the likelihood of pedestrian-vehicle accidents increases.

Trip generation increase in one TAZ leads to a higher number of pedestrian and vehicle accidents in road segments and this increase in the neighboring zones leads to more vehicle accidents in intersections. Increase of trip generation in one zone and in its neighboring zones leads to a higher volume of vehicle passing traffic in the roads network which, in turn, increases the likelihood of vehicle collisions. Several previous studies found that trip generations and attractions have a significant impact on crash frequency [40]. For example, trip generations and attractions per area are positively associated with segment crash frequency; however, this factor has no significant effect on intersection crashes [41].

Considering the results of the final models, a higher percentage of literate people in one TAZ leads to more vehicle accidents in intersections and more motorcycle and pedestrian accidents in road segments. Moreover, a higher percentage of literate people in the neighboring zones of one zone leads to more pedestrian traffic accidents in intersections and generally accidents in road segments in all three modes of transport. Since higher levels of education usually leads to higher levels of public welfare, when the percentage of literate people increases in one zone, in fact the level of public welfare increases which normally leads to the possession of more personal vehicles. Therefore the volume of vehicle traffic increases which, in turn, results in more accidents in the mentioned transportation modes and facilities.

Number of employment plays a significant role in the probability of an accident [42, 17]. According to the results in this study, as the number of employed people increases in a TAZ, the number of vehicle accidents in intersections and the number of motorcycle and pedestrian accidents in road segments increase as well. Also, as the number of employed people increases in the neighboring zones of one TAZ, the number of motorcycle accidents in intersections and road segments and the number of vehicle accidents in road segments increase. When the number of employment increases in

a TAZ, not only does the vehicle traffic volume increase in that zone and its neighboring zones, but also the level of public welfare and, in turn, the number of personal vehicles increases, too. This leads to a higher volume of passing motor vehicle traffic and more accidents in those zones.

### 5.3. Sensitivity Analysis of Variable

To run a quantitative comparison of the effects of different variables on the accidents occurring in road segments, a sensitivity analysis of variables was carried out. This would help in understanding how important a variable is in the accidents of each mode of transport (vehicle, motorcycle, and pedestrian). In fact, the sensitivity analysis shows how much of the variance in the response variable is accounted for by one unit of change in the independent variable. The results are presented in Table 5.

As can be seen, in intersection accidents, trip generation in the neighboring zones has the greatest impact on the number of accident, so that adding 1 unit to the logarithm of trip generation in the neighboring zones of a zone increases vehicle accidents in that zone by a factor of 79.04. On the other words, when the trip generation is 10 times, vehicle crashes is 79.04 times. The percentage of literate people in one TAZ has the least effect, such that for each percentage of literate people in one TAZ increase, the expected number of accident is increased by a factor of 1.03.

Regarding crashes in road segments, the logarithm of traffic volume in a road segment has different effects on accidents based on transportation modes, such that adding 1 unit to this variable increases vehicle, motorcycle and pedestrian accidents by factors of 3.03, 2.29 and 1.54 respectively. Therefore, the influence of this variable on vehicle accidents is approximately 2 times stronger than that on pedestrian accidents. Regarding the variable population's logarithm of a TAZ, adding 1 unit to this variable decreases the number of vehicle accidents by a factor of 0.22. Whereas, this variable increases motorcycle accidents by a weight of 1.4.

The same interpretation is true for the other variables available in Table 5. Based on the results obtained from analyzing the sensitivity of the significant variables of the final models, it is observable that a variable influences the accidents in various modes of transport differently. Therefore, the necessity of investigating the factors influencing accidents across different modes of transport becomes evident.

## 6. Conclusion

Since the factors affecting accidents are different across transport facilities and modes, the accident prediction model of the present study was developed across different modes of transport (vehicle, motorcycle, and pedestrian). The independent variables were considered at both micro and macro levels. To consider the intra-zone correlation due to common macro variables for the accidents occurring in one TAZ, a multilevel model was adopted in the modeling process. Since the accidents occurring in one TAZ might be affected by the variables available in neighboring zones, a multilevel model was also developed using the variables extracted from the neighboring zones of one TAZ. To this end, the data related to 15968 accidents occurring in 96 TAZs of Tehran were collected. Next, the traffic, social and demographic data related to the area under study were also collected and a database was created under GIS application platform.

Multilevel models were developed with/without considering the effect of neighboring zones, and for comparison purposes, an NB model was also developed and its results were compared with those of the multilevel models.

The final models were developed for the accidents occurring in intersections and road segments for each mode of transport based on the criteria model fit, Log-Likelihood, AICC, and BIC.

According to the results, the multilevel model which considered the influence of neighboring zones had a better performance in comparison to the other two models, namely the multilevel model that did not take into account the effect of neighboring zones and the NB model. Considering the results obtained from the final models, some variables like higher driver age, higher level of driver education, rainy and snowy weather, and higher ratio of bus lines to the roads length available in a TAZ lead to decreasing the number of accidents; while some variables like higher vehicle traffic and higher percentage of employed and literate people in a TAZ lead to a higher number of accidents. Since the variables used in the modeling process proved significant in intersections and road segments for some modes of transport but insignificant for some other modes, and also Based on the results obtained from analyzing the sensitivity of the significant variables of the final models, because the effects of these variables were of different types and amounts in intersections and road segments for different modes, the necessity of investigating the factors affecting accidents across different modes and different transport facilities seems unavoidable.

## 7. Acknowledgements

The authors would like to thank from Transportation and Traffic Organization of Tehran Municipality for providing the traffic data.

## 8. Conflict of Interest

The authors declare no conflict of interest.

## 9. References

- [1] Road Maintenance and Transportation Organization (RMTO). "RMTO-Statistical yearbook, 2015." Ministry of Road and Urban Development, Iran, (2015).
- [2] European Commission. "Road safety in the European Union: Trends, statistics and main challenges." (November 2016).
- [3] Huang, Helai, Hanchu Zhou, Jie Wang, Fangrong Chang, and Ming Ma. "A Multivariate Spatial Model of Crash Frequency by Transportation Modes for Urban Intersections." *Analytic Methods in Accident Research* 14 (June 2017): 10–21. doi:10.1016/j.amar.2017.01.001.
- [4] Petzold, R. "Proactive approach to safety planning." *Public roads* 66(6) (2003): 6-9.
- [5] Lee, J., Abdel-Aty, M., Cai, Q. "Intersection crash prediction modeling with macro-level data from various geographic units." *Accident Analysis & Prevention* 102 (2017): 213-226. doi:10.1016/j.aap.2017.03.009
- [6] Gelman, A., Hill, J. "Data analysis using regression and multilevel/hierarchical models." Cambridge University Press New York, NY, USA, (2006).
- [7] Siddiqui, C., Abdel-Aty, M., Choi, K. "Macroscopic spatial analysis of pedestrian and bicycle crashes." *Accident Analysis & Prevention* 45 (2012): 382-391. doi:10.1016/j.aap.2011.08.003.
- [8] Lee, J., Abdel-Aty, M., Choi, K., Huang, H. "Multi-level hot zone identification for pedestrian safety." *Accident Analysis & Prevention* 76 (2015): 64-73. doi:10.1016/j.aap.2015.01.006.
- [9] Cai, Q., Lee, J., Eluru, N., Abdel-Aty, M. "Macro-level pedestrian and bicycle crash analysis: Incorporating spatial spillover effects in dual state count models." *Accident Analysis & Prevention* 93 (2016): 14-22. doi:10.1016/j.aap.2016.04.018.
- [10] Lascala, E.A., Gerber, D., Gruenewald, P.J. "Demographic and environmental correlates of pedestrian injury collisions: A spatial analysis." *Accident Analysis & Prevention* 32(5) (2000): 651-658. doi:10.1016/s0001-4575(99)00100-1.
- [11] Ha, H.-H., Thill, J.-C. "Analysis of traffic hazard intensity: A spatial epidemiology case study of urban pedestrians." *Computers, Environment and Urban Systems* 35(3) (2011): 230-240. doi:10.1016/j.compenvurbsys.2010.12.004.
- [12] Lord, D., Mannering, F. "The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives." *Transportation Research Part A: Policy and Practice* 44(5) (2010): 291-305. doi:10.1016/j.tra.2010.02.001.
- [13] Mannering, F.L., Bhat, C.R. "Analytic methods in accident research: Methodological frontier and future directions." *Analytic methods in accident research* 1(2014): 1-22. doi:10.1016/j.amar.2013.09.001.
- [14] Anastasopoulos, P.C., Mannering, F.L. "A note on modeling vehicle accident frequencies with random-parameters count models." *Accident Analysis & Prevention* 41(1) (2009): 153-159. doi:10.1016/j.aap.2008.10.005.
- [15] Couto, A., Ferreira, S. "A note on modeling road accident frequency: A flexible elasticity model." *Accident Analysis & Prevention* 43(6) (2011): 2104-2111. doi:10.1016/j.aap.2011.05.033.
- [16] Lord, D., Guikema, S.D., Geedipally, S.R. "Application of the Conway–Maxwell–Poisson generalized linear model for analyzing motor vehicle crashes." *Accident Analysis & Prevention* 40(3) (2008): 1123-1134. doi:10.1016/j.aap.2007.12.003.
- [17] Quddus, M.A. "Modelling area-wide count outcomes with spatial correlation and heterogeneity: An analysis of London crash data." *Accident Analysis & Prevention* 40(4) (2008): 1486-1497. doi:10.1016/j.aap.2008.03.009.
- [18] Hadayeghi, A., Shalaby, A.S., Persaud, B.N. "Development of planning level transportation safety tools using geographically weighted Poisson regression." *Accident Analysis & Prevention* 42(2) (2010): 676-688. doi:10.1016/j.aap.2009.10.016.
- [19] Naderan, A., Shahi, J. "Aggregate crash prediction models: Introducing crash generation concept." *Accident Analysis & Prevention* 42(1) (2010): 339-346. doi:10.1016/j.aap.2009.08.020.
- [20] Pulugurtha, S.S., Duddu, V.R., Kotagiri, Y. "Traffic analysis zone level crash estimation models based on land use characteristics." *Accident Analysis & Prevention* 50 (2013): 678-687. doi:10.1016/j.aap.2012.06.016.
- [21] Xu, P., Huang, H. "Modeling crash spatial heterogeneity: Random parameter versus geographically weighting." *Accident Analysis & Prevention* 75 (2015): 16-25. doi:10.1016/j.aap.2014.10.020.
- [22] Huang, H., Song, B., Xu, P., Zeng, Q., Lee, J., Abdel-Aty, M. "Macro and micro models for zonal crash prediction with application in hot zones identification." *Journal of Transport Geography* 54 (2016): 248-256. doi:10.1016/j.jtrangeo.2016.06.012.

- [23] Guo, F., Wang, X., Abdel-Aty, M.A. "Modeling signalized intersection safety with corridor-level spatial correlations." *Accident Analysis & Prevention* 42(1) (2010): 84-92. doi:10.1016/j.aap.2009.07.005.
- [24] Mitra, S., Washington, S. "On the significance of omitted variables in intersection crash modeling." *Accident Analysis & Prevention* 49 (2012): 439-448. doi:10.1016/j.aap.2012.03.014.
- [25] Dupont, E., Papadimitriou, E., Martensen, H., Yannis, G. "Multilevel analysis in road safety research." *Accident Analysis & Prevention* 60 (2013): 402-411. doi:10.1016/j.aap.2013.04.035.
- [26] Huang, H., Abdel-Aty, M. "Multilevel data and Bayesian analysis in traffic safety." *Accident Analysis & Prevention* 42(6) (2010): 1556-1565. doi:10.1016/j.aap.2010.03.013.
- [27] Shi, Q., Abdel-Aty, M., Yu, R. "Multi-level Bayesian safety analysis with unprocessed automatic vehicle identification data for an urban expressway." *Accident Analysis & Prevention* 88 (2016): 68-76. doi:10.1016/j.aap.2015.12.007.
- [28] Greibe, P. "Accident prediction models for urban roads." *Accident Analysis & Prevention* 35(2) (2003): 273-285. doi:10.1016/s0001-4575(02)00005-2.
- [29] Abbas, K.A. "Traffic safety assessment and development of predictive models for accidents on rural roads in Egypt." *Accident Analysis & Prevention* 36(2) (2004): 149-163. doi:10.1016/s0001-4575(02)00145-8.
- [30] Gujarati, D.N. "Basic Econometrics." Tata McGraw-Hill Education, (2009).
- [31] Jones, K. "Using multilevel models for survey analysis." *Journal of the Market Research Society* 35 (1993): 249-265.
- [32] Wang, J., Huang, H., Zeng, Q. "The effect of zonal factors in estimating crash risks by transportation modes: Motor vehicle, bicycle and pedestrian." *Accident Analysis & Prevention* 98 (2017): 223-231. doi:10.1016/j.aap.2016.10.018.
- [33] Kazazi, J., Winkler, S., Vollrath, M. "The influence of attention allocation and age on intersection accidents." *Transportation research part F: traffic psychology and behaviour* 43 (2016): 1-14. doi:10.1016/j.trf.2016.09.010.
- [34] Mehdizadeh, M., Shariat-Mohaymany, A., Nordfjaern, T. "Accident involvement among Iranian lorry drivers: Direct and indirect effects of background variables and aberrant driving behaviour." *Transportation research part F: traffic psychology and behaviour* 58 (2018): 39-55. doi:10.1016/j.trf.2018.05.029.
- [35] Qiu, L., Nixon, W.A. "Effects of adverse weather on traffic crashes: Systematic review and meta-analysis." *Transportation Research Record* 2055(1) (2008): 139-146. doi:10.3141/2055-16.
- [36] Abdel-Aty, M., Wang, X. "Crash estimation at signalized intersections along corridors: Analyzing spatial effect and identifying significant factors." *Transportation Research Record: Journal of the Transportation Research Board* 1953 (2006): 98-111. doi:10.3141/1953-12.
- [37] Xie, K., Wang, X., Huang, H., Chen, X. "Corridor-level signalized intersection safety analysis in shanghai, china using Bayesian hierarchical models." *Accident Analysis & Prevention* 50 (2013): 25-33. doi:10.1016/j.aap.2012.10.003.
- [38] Wang, X., Song, Y., Yu, R., Schultz, G.G. "Safety modeling of suburban arterials in shanghai, china." *Accident Analysis & Prevention* 70 (2014): 215-224. doi:10.1016/j.aap.2014.04.005.
- [39] Wier, M., Weintraub, J., Humphreys, E.H., Seto, E., Bhatia, R. "An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning." *Accident Analysis & Prevention* 41(1) (2009): 137-145. doi:10.1016/j.aap.2008.10.001.
- [40] Abdel-Aty, M., Siddiqui, C., Huang, H., Wang, X. "Integrating trip and roadway characteristics to manage safety in traffic analysis zones." *Transportation Research Record* 2213(1) (2011): 20-28. doi:10.3141/2213-04.
- [41] Wang, J., Huang, H. "Road network safety evaluation using Bayesian hierarchical joint model." *Accident Analysis & Prevention* 90 (2016): 152-158. doi:10.1016/j.aap.2016.02.018.
- [42] Noland, R.B., Quddus, M.A. "A spatially disaggregate analysis of road casualties in England." *Accident Analysis & Prevention* 36(6) (2004): 973-984. doi:10.1016/j.aap.2003.11.001.