

## Research Article

# Crash Prediction and Risk Evaluation Based on Traffic Analysis Zones

Cuiping Zhang,<sup>1</sup> Xuedong Yan,<sup>1</sup> Lu Ma,<sup>1</sup> and Meiwu An<sup>2</sup>

<sup>1</sup>MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

<sup>2</sup>Transportation Modeler, Saint Louis County Departments of Highways & Traffic and Public Works, 1050 N. Lindbergh, Creve Coeur, MO 63132, USA

Correspondence should be addressed to Xuedong Yan; [xdyan@bjtu.edu.cn](mailto:xdyan@bjtu.edu.cn)

Received 5 December 2013; Revised 1 January 2014; Accepted 12 January 2014; Published 25 February 2014

Academic Editor: Wuhong Wang

Copyright © 2014 Cuiping Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Traffic safety evaluation for traffic analysis zones (TAZs) plays an important role in transportation safety planning and long-range transportation plan development. This paper aims to present a comprehensive analysis of zonal safety evaluation. First, several criteria are proposed to measure the crash risk at zonal level. Then these criteria are integrated into one measure-average hazard index (AHI), which is used to identify unsafe zones. In addition, the study develops a negative binomial regression model to statistically estimate significant factors for the unsafe zones. The model results indicate that the zonal crash frequency can be associated with several social-economic, demographic, and transportation system factors. The impact of these significant factors on zonal crash is also discussed. The finding of this study suggests that safety evaluation and estimation might benefit engineers and decision makers in identifying high crash locations for potential safety improvements.

## 1. Introduction

Traffic crashes have caused tremendous losses in terms of death, injury, lost productivity, and property damage in our society. In order to investigate factors which have an impact on crashes, extensive researches have been conducted and recognized in the literature [1–4]. According to different research purposes, crashes can be analyzed individually or aggregately for road segments, intersections, or TAZs. Recent studies suggested that the TAZ level crash estimate is important for safety planning purposes and also useful in identifying and diagnosing zonal safety issues in an earlier planning stage [5–7]. In summary, zonal crash analysis is crucial in risk evaluation as well as crash prediction for TAZs. Since crash frequency is the simplest measure to assess the degree of safety, modeling crash frequency attracts more attention from researchers; most of the recent studies use it as the major target. However, modeling crash frequency is inadequate to make a comprehensive evaluation on traffic safety, especially for the safety analysis of TAZs in which safety risk may be also affected by many other factors, for

example, area, population, total roadway length, vehicle miles travelled (VMT), and so forth. Further, the safety status for each TAZ also depends on the severity of crashes. For example, crashes which involve fatality or injury should receive more attention comparing to the propriety-damage-only (PDO) crashes. By the motivation of comprehensive evaluations, it is worth formulating safety risk measurements in the consideration of crash frequency, crash severity, and zonal characteristics. Toward this end, one of the objects of this paper is to provide a more comprehensive safety evaluation analysis based on several criteria, including total number of crashes, total number of fatal and injury crashes, total crash exposure rate, injury/fatality crash exposure rate, weight hazard index (WHI), and average hazard index (AHI). These criteria are capable of capturing the degree of safety risk for each zone in different aspects; the detailed definition of these criteria will be presented in the session of methodology.

In order to develop independent and dependent variables for each TAZ, data from different sources such as socioeconomic and transportation network will be aggregated into each TAZ. First, all datasets are geocoded into the GIS

platform. With the topology relationship among accidents, roads, and TAZs, each crash will be assigned to a specified TAZ. During this process, one practical issue is to determine a TAZ for the crashes which occurred on boundaries shared by two or more adjacent TAZs. What should be pointed out is that such issue is not encountered in crash analysis for road segments or intersections. In addition, the method for allocating crashes on zonal boundaries was not clearly discussed in current literature. So this study also seeks to propose a method for assigning crashes on boundaries as well as presenting the process of multisource data integration.

## 2. Literature Review

In literature, traffic safety studies are typically conducted on road segments and intersections. Most of these studies consider crash frequency as their primary subjects. Empirical results clearly indicated the possible association between crash frequency and road characteristics. Tarko et al. [8] indicated that attributes such as traffic lane width, location of roads (rural/urban), and shoulder width are associated with crash frequency. Ye et al. [9] found that the terrain and the geometry of the roadway as well as visibility provided by lighting also significantly affect the crash frequency. For zonal crash frequency estimation, which recently received researchers' attention, can also be evaluated by the overall road characteristics within a zone as expected. Guevara et al. [5] stated that the safety predication model at planning level is feasible and the model is helpful in developing incentive programs for safety improvements. This type of research connects the crash count in a TAZ not only to the transportation characteristics but also to several social-economic and demographic characteristics, for example, average household size and zonal population.

One of the earliest models regarding the zonal level crash prediction was developed by Levine et al. [10]. They tried to relate the motor vehicle accidents to zonal population, employment, and road characteristics for the City and County of Honolulu. It established an analysis framework for zonal level accident analysis as well as safety planning. However, the model in this study is linear and hence inappropriate for crash count data analysis [6]. Exponential-type nonlinear models are preferred for handling crash count data.

Since crash counts are usually assumed to be nonnegative distributed integer numbers, Poisson or negative binomial distribution would be a nature way to modeling the data. Poisson regression models can be accepted in condition where the traffic accidents are considered as a standard Poisson process and have been applied in a number of safety researches [11, 12]. However, crash frequency is usually observed to be overdispersed [13]. In order to accommodate such situations, negative binomial regression models can be used [6]. In the study of Hadayeghi et al. [6], negative binomial regression models were developed to estimate the zonal traffic crash using variables of zonal attributes of socioeconomic and demographic, network, and traffic demand. Two categories of crash counts, total crash and severe crash, are examined in this analysis for the city of Toronto, Ontario, Canada. It is a pilot research for macrolevel accident

prediction analysis, suggesting that the model is preferred to function as a forecasting tool for zonal crash frequency rather than to infer the causes of crashes [6]. Along with the development of zonal crash prediction models, it is also necessary to investigate the statistical relationship between crash frequency and descriptive variables. Lovegrove and Sayed [7] developed zonal level models to enhance traditional reactive road safety improvement programs. In this research, they developed negative binomial regression models and applied the models for black-spot identification as a case study. Other than these parametric statistical models, a tree based model is recently developed by Siddiqui et al. [14] and it intends to make prediction on the aggregated zonal crash frequency. The tree based model provides a nonparametric approach in this area; yet it is more designed to make crash forecasting rather than inferences. Specifically, at the zonal level, negative binomial regression models were deployed and have been deducted by Abdel-Aty et al. [15, 16]. Although there are many other regression models adopted in zonal level safety analysis, for example, Bayesian based models and tree based models, negative binomial models are typically used to capture the key and basic points in transportation safety analysis [17].

Even literature on zonal crashes is emphasized recently, there still lacks analysis for zonal safety evaluations. Toward this end, this study seeks to conduct a comprehensive analysis for zonal safety evaluations.

## 3. Data

This research uses data provided by Pikes Peak Area Council of Governments, Colorado. The data includes three types of datasets: the TAZ dataset, the traffic and roadway dataset, and the accident datasets. Specifically, the TAZ dataset contains geographic boundary information for each zone, zonal attributes, for example, population, number of housing units and number of employments in a zone, and so forth. The traffic and roadway data include number of lanes, road length, and traffic characteristics such as traffic volumes, free flow speed, speed limit, functional classification, and capacity. The accident dataset is obtained from the Department of Revenue (DOR) and geocoded into GIS database by the Pikes Peak Area Council of Government (PPACG), which include location and type for each accident that occurred on the roads of the study area.

For accident data analysis, GIS technique is a crucial tool to visualize and process traffic accident data [18, 19]. It is also able to integrate different datasets from multiple sources [20]. Hence, the current study will import all datasets into GIS platform for purposes of data processing and the subsequent safety evaluations. The accident dataset consists of all reported accidents from July 2006 to December 2010 and each accident recorded in the dataset is categorized as fatal, injury, or property-damage-only. The data also includes several road environment factors such as accident location, lighting condition, road surface, and weather condition. All variables of the multisource transportation data were aggregated at the TAZ level using ArcMap 10. After data integration, the detailed TAZ level variables are illustrated in Table 1. In the study area, there are total 733 TAZs. Among them, the smallest

TABLE 1: Descriptions of TAZ level variables.

Variable name	Definition	N	Mean	Standard deviation	Min	Max
Variables related to roadway characteristics						
LENGTH	Total roadway segment length within a TAZ	733	6.45	7.19	0.30	53.51
ADT	Total roadway average daily traffic within a TAZ	733	212.94	242.17	0.53	2330.21
FFS	Average total roadway free flow speed within a TAZ	733	42.41	9.65	20	76
VMT	Total daily vehicle travelled in a TAZ	733	102	912686	59324.95	72833.07
LANE	Average number of lanes in a TAZ	733	2.86	0.88	2	6
Independent variables related to demographic and economic characteristics						
LOW	Number of low income households in proportion to the total number of households	733	0.16	0.15	0	1
LOWMID	Number of middle-low income households in proportion to the total number of households	733	0.22	0.13	0	0.55
MID	Number of middle income households in proportion to the total number of households	733	0.20	0.10	0	0.50
MIDHIGH	Number of middle-high income households in proportion to the total number of households	733	0.22	0.14	0	1
HIGH	Number of high income households in proportion to the total number of households	733	0.13	.13	0	1
BASIC	Number of workers in basic category in proportion to total number of workers in all categories	733	0.21	0.23	0	1
RETAIL	Number of workers in retail category in proportion to total number of workers in all categories	733	0.17	0.20	0	1
TOTALSERVI	Total number of workers in service category in proportion to total number of workers in all categories	733	0.60	0.27	0	1
EMENROLL	Number of students in elementary school in proportion to total number of population	733	0.12	0.33	0	4.34
HSENROLL	Number of students in high school in proportion to total number of population	733	0.05	0.26	0	2.26
COLLENROLL	Number of college students in proportion to total number of population	733	0.05	0.43	0	7.39
POP	Total number of population in a TAZ (in thousands)	733	0.88	0.93	0	8.63
AREA	Area of a TAZ	733	0.01	123.04	3.66	10.65

TAZ is 0.01 square mile and the largest is 123.04 square mile, which represents a large variation in area. Table 1 summarizes definition and descriptive statistics of these variables.

#### 4. Methodology

4.1. *Evaluation Measures.* Zonal level crash analysis and safety evaluation should use the criteria which will not only

capture the safety condition of each zone, but also compare the crash risk and severity among zones. In the light of comprehensive evaluation, it is useful to identify the unsafe zones at the stage of planning level. This study introduces six measures for evaluation of zonal traffic safety risks. The first five measurements are total number of crashes, total number of injury and fatal crashes, total crash exposure rate, injury/fatal exposure rate, and weighted hazard index (WHI).

And the average hazard index (AHI) is the last one which comprehensively combines the first five measurements by a scoring system.

Before introducing the six criteria, it is necessary to assign crashes to a zone which occurred at the zonal boundaries. This study proposed a method to assign these crashes; the concept is to allocate these crashes to the surrounding zones in proportional to the number of crashes which occurred insider these zones. Consider the following:

$$C_i = C_i^\alpha + \sum_{j \in J_i} \frac{C_j^\alpha}{B_j^\alpha}. \quad (1)$$

Equation (1) illustrates the proposed split mechanism, and  $i$  indexes the TAZs in the entire research area;  $J_i$  is the set which contains all crashes that occurred at the boundary of zone  $i$ .  $C_i$  is the total number of crashes assigned to each zone, which is the sum of crashes that occurred inside this zone, represented by  $C_i^\alpha$  and the scaled crashes at the boundaries.  $B_j^\alpha$  is the summation of the number of insider crashes of TAZs that share the crash  $j$ . So  $C_i^\alpha$  constitutes a part of  $B_j^\alpha$ . To be noted, (1) can be used to define the total number of crashes for each zone, and it is also capable of defining the number of injures crashes, number of fatal crashes, and so forth. Then the six safety risk evaluation measurements are introduced as follows.

**4.1.1. Total Number of Crashes.** It is the sum of all crashes including fatal ( $F$ ), injury ( $I$ ), and property damage only ( $O$ ). It can be mathematically expressed by (2), where  $M_{1i}$  is the total number of crashes in TAZ  $i$  and  $F_i, I_i$ , and  $O_i$  are the number of fatal crashes, injury crashes, and property-damage-only crashes, respectively. Consider the following:

$$M_{1i} = F_i + I_i + O_i. \quad (2)$$

**4.1.2. Total Number of Fatal and Injury Crashes.** It is the sum of only fatal ( $F$ ) and injury ( $I$ ) type of crashes. It provides an indication of crash severity in each zone; it can be illustrated in

$$M_{2i} = F_i + I_i. \quad (3)$$

**4.1.3. Total Crash Exposure Rate.** This exposure rate reflects relatively safety at a zone. It can be expected that the number of crashes would naturally increase if vehicle miles travelled increase in a particular TAZ. The exposure rate is calculated by (4) and it is reported on the basis of one million vehicle miles traveled due to the small chance of accidents. Here,  $VMT_i$  is the total vehicle miles travelled during the study period in TAZ  $i$ . Consider the following:

$$M_{3i} = \frac{F_i + I_i + O_i}{VMT_i} \times 10^6. \quad (4)$$

**4.1.4. Injury/Fatality Crash Exposure Rate.** Similar to the total crash exposure rate, it measures the exposure rate for injury and fatal crashes in each TAZ. Since there are even fewer

injury and fatal crashes, this rate is calculated on the basis of 100 million vehicle miles traveled in

$$M_{4i} = \frac{F_i + I_i}{VMT_i} \times 10^8. \quad (5)$$

**4.1.5. Weighted Hazard Index (WHI).** This measure takes into consideration the weighted severity of the crashes. It is also a type of exposure rate. In (6), the weight for each type of crashes is represented by  $A, B$ , and  $C$ , respectively, with  $A$  for fatal crashes,  $B$  for injury crashes, and  $C$  for property-damage-only crashes. In this study,  $A = 12, B = 5$ , and  $C = 1$  are adopted [21]. Consider the following:

$$M_{5i} = \frac{F_i \times A + I_i \times B + O_i \times C}{VMT_i} \times 10^6. \quad (6)$$

**4.1.6. Average Hazard Index (AHI).** Since each of the five criteria reflects the traffic safety condition from different aspects, it is useful to integrate these five criteria into one which can reflect a more comprehensive performance of traffic safety for each TAZ. First, the five criteria are transferred to dimensionless scores for each TAZ. The scores are integers ranging from 0 to 4 to represent five levels of safety evaluation. For each criterion, 5%, 50%, and 95% percentile values are calculated for all TAZs in the study area, and then the scores are assigned to each TAZ under each criterion by the following definition.

Score = 0 if zero crash was presented.

Score = 1 if  $0 < \text{safety criterion} \leq 5\%$  percentile.

Score = 2 if  $5\% \text{ percentile} < \text{safety criterion} \leq 50\%$  percentile.

Score = 3 if  $50\% \text{ percentile} < \text{safety criterion} \leq 95\%$  percentile.

Score = 4 if  $95\% \text{ percentile} < \text{safety criterion}$ .

To be noted, scores under the five safety measurements are calculated for each TAZ, respectively. So the AHI is calculated by averaging the five scores, and then it is rounded to the closed integers. Equation (7) gives the formulation of AHI, where  $S_{1i}, S_{2i}, S_{3i}, S_{4i}$ , and  $S_{5i}$  are the scores of  $M_{1i}, M_{2i}, M_{3i}, M_{4i}$ , and  $M_{5i}$ , respectively. Consider the following:

$$A_i = \text{Round} \left( \frac{S_{1i} + S_{2i} + S_{3i} + S_{4i} + S_{5i}}{5} \right). \quad (7)$$

**4.2. Crash Frequency Predication Model.** As discussed in Section 2, traffic crash frequency is overdispersed, so this study would adopt negative binomial regression model to analyze and estimate crash frequency. Traffic crashes are assumed to be independently distributed in the form of negative binomial with positive mean parameter  $\mu_i$  and positive scale parameter  $\alpha$  for zone  $i$  shown as below:

$$Y_i \sim \text{Negative Binomial}(\mu_i, \alpha). \quad (8)$$

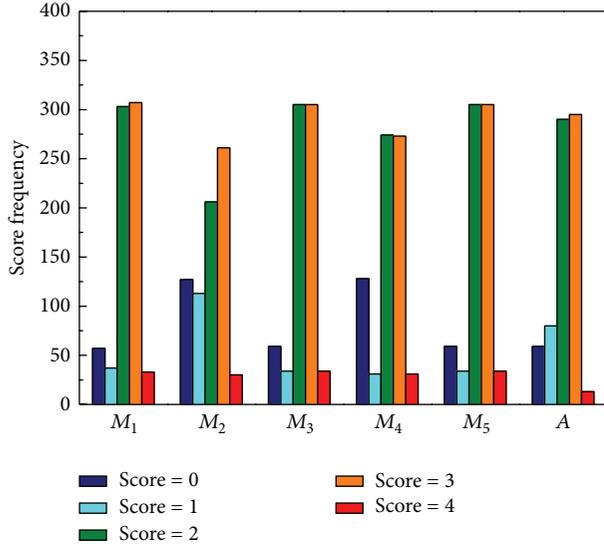


FIGURE 1: Score frequency of the six evaluation measurements.

Therefore, the probability mass function for crash frequency of zone  $i$  is presented by (9), where  $\Gamma(\cdot)$  is gamma function. Consider the following:

$$P(Y_i = y | \mu_i, \alpha) = \frac{\Gamma(y + \alpha)}{y! \Gamma(\alpha)} \left( \frac{\mu_i}{\mu_i + \alpha} \right)^y \left( \frac{\alpha}{\mu_i + \alpha} \right)^\alpha, \quad (9)$$

$$y \in \{0, 1, 2, \dots\}.$$

With the specification above, it can be found that the expected value of  $Y_i$ ,  $E(Y_i) = \mu_i$  and its variance  $\text{Var}(Y_i) = \mu + \mu^2/\alpha > E(Y_i)$ . The scale parameter  $\alpha$  is assumed to be the same for all TAZ samples while the mean parameter  $\mu_i$  is varying across TAZs. In a form of generalized linear model, a logarithm link function is used as follows:

$$\log(\mu_i) = \mathbf{X}_i^T \boldsymbol{\beta}, \quad (10)$$

where  $\mathbf{X}_i$  is the vector of explanatory variables and  $\boldsymbol{\beta}$  is the coefficients for the corresponding vectors.

## 5. Analysis and Results

**5.1. Descriptive Analysis.** The total number of crashes for the 733 TAZs is 46948, among which 0.32% are fatal crashes. The distribution of the scores under the six safety evaluation criteria is presented in Figure 1. According to the definition, the proportion of TAZs with scores of 2, 3, and 4 should be roughly 45%, 45%, and 5%, respectively, while the remaining 5% is consisted of zero-crash zones and the zones with score one. There are also connections between the six measures. For example, the number of TAZs with score zero is identical for the measurements of total number of fatal and injury crashes and injury/fatality crash exposure rate since those TAZs have no injury or fatal crashes presented. A safe TAZ should have relatively small score under all the six criteria. There should not be too much variation between these scores

TABLE 2: Score differences between the AHI and the first five criteria.

Difference between AHI and the first five measures	Frequency of the values				
	-2	-1	0	1	2
$A_i - M_{1i}$	0	175	480	78	0
$A_i - M_{2i}$	0	103	369	249	12
$A_i - M_{3i}$	10	104	593	26	0
$A_i - M_{4i}$	0	66	530	133	4
$A_i - M_{5i}$	2	104	617	10	0

numbers as positive correlations exist. In order to understand the discrepancy between the first five scores and the AHI, Table 2 provides the distribution of the difference between the first five scores and the AHI. For the 733 TAZ samples, this difference is varying from  $-2$  to  $2$ , in which the negative values indicate that the AHI leads to less TAZs than the corresponding measurement under this score. From Table 2, there are only mismatch of less one category between the AHI and other scores. Therefore, AHI is reliable and capable of presenting a comprehensive criterion to measure the safety risk.

**5.2. Risk Evaluation Analysis.** In this study, each TAZ is first evaluated using the AHI score. Figure 2 shows the geographic distribution of AHI scores for the 733 TAZs. TAZs with the highest risk ( $A_i = 4$ ) should be paid more attentions, especially at the planning stage. To be noted, most of the zones belong to categories of score = 3 and score = 2, which account for 90% of zones. One objective of this analysis is to identify the zones with the highest risk. It can be seen that almost all the high-risk ( $A_i = 4$ ) TAZs are located in high-density areas and the right part of Figure 2 presents the enlarged image of this area. It is also interesting that a great proportion of the high-risk TAZs are connected, which may indicate the geographic correlation between adjacent zones in terms of risks.

Besides AHI, another method which also comprehensively combines the six measures is undertaken to identify high-risk TAZs. For each TAZ, the number of the six criteria whose score reaches 4 is used to identify these unsafe TAZs. This approach is designed to identify unsafe zones with more high-risk features. So this number ranges from zero to six for each TAZ. Zero represents that none of these six scores is greater than 4 whereas six represents that zones may be under high safety risk. Figure 3 illustrates this approach as well as its distribution for the 733 TAZs. Similarly to the distribution of AHI, unsafe zones are also concentrated in the high-density area.

**5.3. The Crash Prediction Model.** The negative binomial regression structure is used to model the total number of crashes. The 733 TAZ samples are assumed to be distributed as independent negative binomial with the same scale parameter  $\alpha$ . As mentioned before, the mean parameter is modeled as the logarithm of linear combination of explanatory variables. Table 3 illustrates the original model with total 20 variables including the variables of roadway characteristics, for example, total roadway segment length within a TAZ and average free flow speed within a TAZ, demographic and

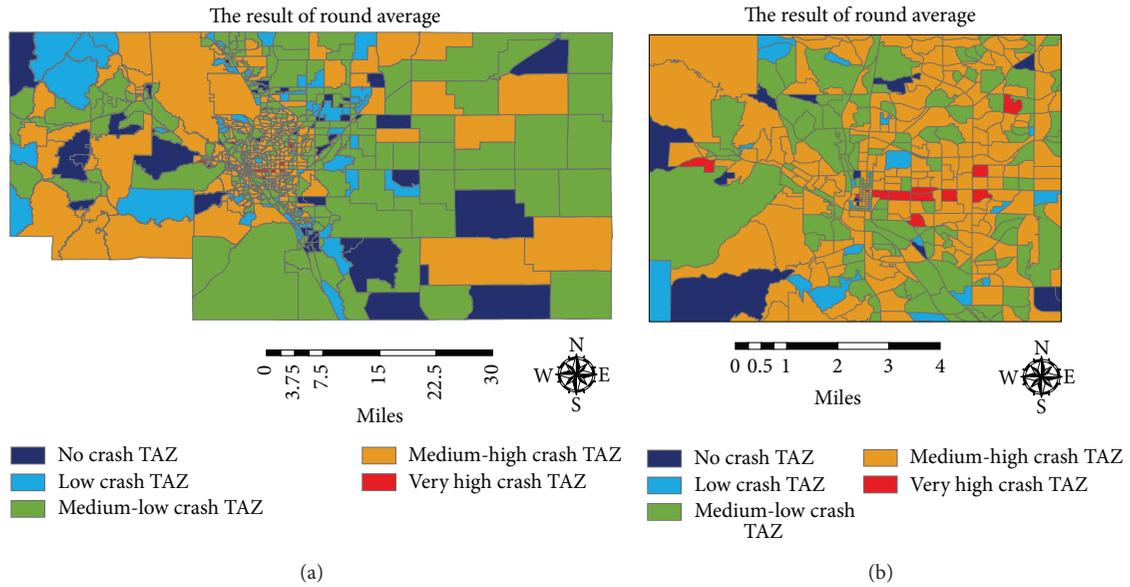


FIGURE 2: Geographic distribution of AHI for 733 TAZs.

TABLE 3: Estimation results of the original regression model.

Parameter	Estimate coefficient	Std. error	95% Wald confidence interval		Hypothesis test		
			Lower	Upper	Wald chi-square	df	P value
(Intercept)	3.179	1.075	1.071	5.286	8.736	1	.003
LENGTH	.033	.0087	.016	.050	14.248	1	<.0001
ADT	.002	.0004	.002	.003	45.815	1	<.0001
FFS	-.027	.0057	-.038	-.016	22.487	1	<.0001
VMT	3.291E - 06	1.2109E - 06	9.175E - 07	5.664E - 06	7.386	1	.007
LANE_2	-.271	1.0288	-2.287	1.745	.069	1	.792
LANE_3	.269	1.0269	-1.743	2.282	.069	1	.793
LANE_4	.429	1.0274	-1.585	2.443	.174	1	.676
LANE_5	.328	1.0374	-1.705	2.362	.100	1	.752
LANE_6	0*						
LOWMID	1.107	.4010	.321	1.893	7.621	1	.006
MID	-.508	.4799	-1.448	.433	1.119	1	.290
MIDHIGH	-.208	.3775	-.948	.532	.304	1	.581
HIGH	-.145	.3994	-.928	.638	.132	1	.716
RETAIL	.858	.2440	.380	1.337	12.373	1	.000
TOTALSERVI	.154	.1809	-.201	.508	.721	1	.396
EMENROLL	.118	.1370	-.150	.387	.743	1	.389
HSENROLL	.243	.1486	-.048	.535	2.682	1	.101
COLLENROLL	-.048	.0980	-.240	.144	.237	1	.627
POP	.326	.0589	.210	.441	30.621	1	<.0001
AREA	-.002	.0045	-.011	.007	.189	1	.664
1/α	.964	.0514	.868	1.070			

\*Reference category.

social-economic characteristics, for example, proportion of low income households, total number of population within a TAZ, and TAZ areas. However, the variables are still overestimated as some of them are not significant enough. The model in Table 3 is developed through a selection process based

on the backward elimination with a 0.05 significance level. The final model with all significant variables is presented in Table 4.

For the variables of roadway characteristics, the model results indicate that the VMT (vehicle miles travelled) and

TABLE 4: Estimation results of the final regression model.

Parameter	Estimate coefficient	Std. error	95% Wald confidence interval		Hypothesis test		
			Lower	Upper	Wald chi-square	df	P value
(Intercept)	2.952	.2543	2.454	3.450	134.805	1	<.0001
LENGTH	.019	.0068	.006	.032	7.666	1	.006
ADT	.004	.0002	.003	.004	236.843	1	<.0001
FFS	-.019	.0052	-.029	-.009	14.087	1	<.0001
HIGH	-1.009	.3018	-1.60	-.417	11.178	1	.001
RETAIL	1.226	.2351	.765	1.686	27.182	1	<.0001
TOTALSERVI	.391	.1736	.051	.731	5.081	1	.024
HSENROLL	.296	.1470	.008	.584	4.062	1	.044
POP	.372	.0488	.276	.467	57.998	1	<.0001
1/α	1.025	0.054	.925	1.136			

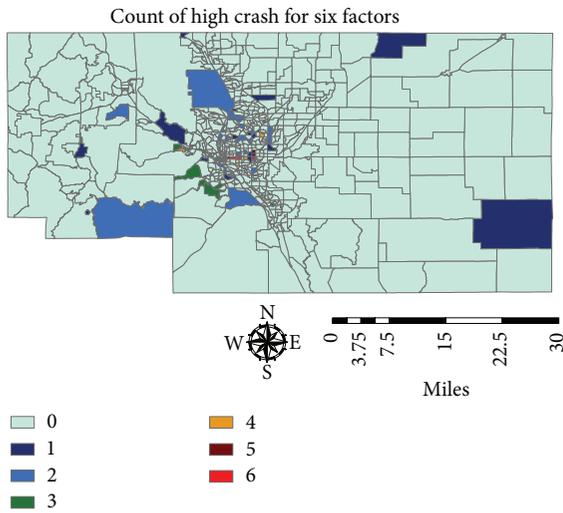


FIGURE 3: Number of measurements with large scores.

LANE (average number of lanes in a TAZ) are insignificant, whereas the LENGTH, ADT, and FFS are preserved after the backward elimination process. VMT are highly correlated with LENGTH and ADT since high VMT may indicate high LENGTH or high ADT. So the VMT are not included in the model because of collinearity. The average number of lanes (LANE) is insignificant in this model. With controlling all the other variables, LANE cannot contribute additional information to enhance the model and it may not have an influence on zonal crash.

The length of roadway within a TAZ has a positive coefficient as expected. Therefore, the zones with more roadways will likely have more crashes. The total average daily traffic within a TAZ also has a positive coefficient. With ADT and other variables being controlled, zones with more LENGTH tend to have high VMT and hence the zone is more exposed to accidents. On the other hand, with LENGTH and other variables being controlled, zones with higher ADT will be inclined to lead to high VMT.

The negative coefficient on FFS indicates that TAZs with higher average free flow speed (FFS) are less likely to experience crashes with the controlling of other variables. This finding is consistent with past research [22], in which the crash frequency is defined for road segments and intersections. One possible explanation is that FFS is highly correlated with the conditions and functional classification of road facilities; better road facilities will provide better safety conditions for driving and hence reduce the crash frequency accordingly.

The economic characteristics of a TAZ also have an impact on crash frequency. In this study, the proportion of high income households within a TAZ (HIGH) is negatively associated with the total number of crashes. Therefore, TAZs with more high income households are statistically safer than other TAZs. Several reasons can lead to such result. For example, richer households may have safer driving behavior, and the condition of road facility in richer TAZs may be also better than other TAZs, which lead to a safer driving environment. Other than the impact of income attributes, the proportion of retail and service employment also affect the crash frequency in a positive manner. Therefore, the increase of RETAIL or TOTALSERVI with the other variables being fixed may cause the increase of crash frequency.

The coefficient on HSENROLL (number of students in high school in proportion to total number of population) is positive, which suggests that increase of HSENROLL may cause the increase of crash frequency with other variables being controlled. It is probably that the age of high school students is between 13 and 18, some of these students (age greater than 16) are allowed to hold driver's licenses in Colorado, and these students will account for part of the driving activities. These drivers are in fact less experienced and own high crash risk than older ones [23]. So they are more likely to make more crashes for the TAZ with other variables being fixed. POP (total number of population in a TAZ) is positively correlated with the crash frequency in this analysis. Consider the following:

$$\mu = \exp(2.592 + 0.019LENGTH + 0.004ADT - 0.019FFS - 1.009HIGH)$$

$$\begin{aligned}
 &+ 1.226\text{RETAIL} + 0.391\text{TOTALSERVI} \\
 &+ 0.296\text{HSENROLL} + 0.372\text{POP}). \quad (11)
 \end{aligned}$$

Other than behavioral inferences between crash frequency and these explanatory variables, another important function of the model is to provide a tool for crash prediction. According to the estimation of the negative binomial model, the total number of crashes in a TAZ can be predicted by its mean expression in (11).

## 6. Conclusions and Discussions

Zonal crash prediction and safety evaluation are critical in the light of transportation safety planning and diagnosis for safety issues. This paper contributes by presenting a comprehensive safety evaluation analysis with combining multiple criteria, as well as developing a crash prediction model for the diagnostic purposes. A GIS based platform for data integration and safety evaluation is introduced.

First, five criteria for measuring TAZ level safety risk are introduced. Then based on the distribution of measurements under the five criteria, a dimensionless score system is defined and a more comprehensive criterion (AHI) is calculated by the rounded average of the five scores for each TAZ. AHI is then applied to identify unsafe TAZs, most of which are located in the high-density area in this study.

For diagnosis purposes, this study developed a crash prediction model through a negative binomial framework. According to the estimation results, it is found that the crash frequency is associated with roadway and traffic characteristics, for example, average free flow speed, average daily traffic within a zone and total roadway length within a zone, and social-economic and demographic characteristics, for example, total population, proportion of high income households, and so forth.

This paper introduces first step for comprehensive evaluation analysis of zonal safety under the GIS platform, and there are still several avenues for further research. At least, the statistical modeling of other evaluation criteria such as WHI and AHI is helpful in understanding their sensitivity to the zonal change of social-economic, demographic, and transportation system characteristics.

There still exist some limitations of the paper. The safety evaluation analysis will help to identify the high-risk TAZs which may need more attention during the planning or diagnostic process. Therefore, another interesting work is to examine the connection between the safety situations and the surrounding environment, for example, zonal social-economic and demographic characteristics and road and traffic characteristics. In this analysis, the total number of crashes is of primarily interest for the following reasons. First, the analysis process and modeling technique are readily to extent for number of fatal or injury crashes. Second, the calculation of all these evaluation criteria relies on this number. So this study treats the total number of crashes as the response variable. And the modeling analysis for other criteria leaves as directions for future research.

Furthermore, in interpreting the increase of RETAIL or TOTALSERVI with other variables being fixed, it is conjectured that the pattern could include the driving behaviors and person characteristics for workers in the categories of retail and service. However, the current model does not clearly state the reasons. So it is necessary to include more explanatory variables in the model and this part is identified as an import direction for future research.

Finally, it would be more appropriate to introduce more zonal level descriptors which may affect drivers' behavior in crash prediction models for example, the gender distribution, age distribution, and so forth. With additional explanatory variables, the model will be more enriched and explainable. It is also interesting to study the safety correlation between adjacent TAZs.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

This work is supported by National Natural Science Foundation (71210001), Ph.D. Programs Foundation of Ministry of Education of China (20110009110013), the Fundamental Research Funds for the Central Universities (2013JBM008, 2014YJS084), "863" Research Project (2011AA110303), and Program for New Century Excellent Talents in University (NCET-11-0570).

## References

- [1] V. Shankar, F. Mannering, and W. Barfield, "Effect of roadway geometrics and environmental factors on rural freeway accident frequencies," *Accident Analysis and Prevention*, vol. 27, no. 3, pp. 371–389, 1995.
- [2] A. T. McCart, V. S. Northrup, and R. A. Retting, "Types and characteristics of ramp-related motor vehicle crashes on urban interstate roadways in Northern Virginia," *Journal of Safety Research*, vol. 35, no. 1, pp. 107–114, 2004.
- [3] X. Yan, E. Radwan, and M. Abdel-Aty, "Characteristics of rear-end accidents at signalized intersections using multiple logistic regression model," *Accident Analysis and Prevention*, vol. 37, no. 6, pp. 983–995, 2005.
- [4] S. Mitra and S. Washington, "On the significance of omitted variables in intersection crash modeling," *Accident Analysis and Prevention*, vol. 49, pp. 439–448, 2012.
- [5] F. L. de Guevara, S. P. Washington, and J. Oh, "Forecasting crashes at the planning level: simultaneous negative binomial crash model applied in Tucson, Arizona," *Transportation Research Record*, no. 1897, pp. 191–199, 2004.
- [6] A. Hadayeghi, A. S. Shalaby, and B. N. Persaud, "Macrolevel accident prediction models for evaluating safety of urban transportation systems," *Transportation Research Record*, no. 1840, pp. 87–95, 2003.
- [7] G. R. Lovegrove and T. Sayed, "Macrolevel collision prediction models to enhance traditional reactive road safety improvement programs," *Transportation Research Record*, no. 2019, pp. 65–73, 2007.

- [8] A. P. Tarko, M. Inerowicz, J. Ramos, and W. Li, "Tool with road-level crash prediction for transportation safety planning," *Transportation Research Record*, no. 2083, pp. 16–25, 2008.
- [9] X. Ye, R. M. Pendyala, S. P. Washington, K. Konduri, and J. Oh, "A simultaneous equations model of crash frequency by collision type for rural intersections," *Safety Science*, vol. 47, no. 3, pp. 443–452, 2009.
- [10] N. Levine, K. E. Kim, and L. H. Nitz, "Spatial analysis of Honolulu motor vehicle crashes: II. Zonal generators," *Accident Analysis and Prevention*, vol. 27, no. 5, pp. 675–685, 1995.
- [11] M. Poch and F. Mannering, "Negative binomial analysis of intersection-accident frequencies," *Journal of Transportation Engineering*, vol. 122, no. 2, pp. 105–113, 1996.
- [12] A. Tarko and M. Tracz, "Accident prediction models for signalized crosswalks," *Safety Science*, vol. 19, no. 2-3, pp. 109–118, 1995.
- [13] S. C. Joshua and N. J. Garber, "Estimating truck accident rate and involvements using linear and Poisson regression models," *Transportation Planning and Technology*, vol. 15, no. 1, pp. 41–58, 1990.
- [14] C. Siddiqui, M. Abdel-Aty, and H. Huang, "Aggregate nonparametric safety analysis of traffic zones," *Accident Analysis and Prevention*, vol. 45, pp. 317–325, 2012.
- [15] M. Abdel-Aty, C. Siddiqui, and H. Huang, "Managing roadway safety at the traffic analysis zones level," in *Proceedings of the 1st International Conference on Transportation Information and Safety (ICTIS '11)*, pp. 230–236, ASCE, Wuhan, China, June-July 2011.
- [16] M. Abdel-Aty, C. Siddiqui, H. Huang, and X. Wang, "Integrating trip and roadway characteristics to manage safety in traffic analysis zones," *Transportation Research Record*, no. 2213, pp. 20–28, 2011.
- [17] G. R. Lovegrove and T. Sayed, "Macro-level collision prediction models for evaluating neighbourhood traffic safety," *Canadian Journal of Civil Engineering*, vol. 33, no. 5, pp. 609–621, 2006.
- [18] T. Steenberghen, T. Dufays, I. Thomas, and B. Flahaut, "Intra-urban location and clustering of road accidents using GIS: a Belgian example," *International Journal of Geographical Information Science*, vol. 18, no. 2, pp. 169–181, 2004.
- [19] S. Erdogan, I. Yilmaz, T. Baybura, and M. Gullu, "Geographical information systems aided traffic accident analysis system case study: city of Afyonkarahisar," *Accident Analysis and Prevention*, vol. 40, no. 1, pp. 174–181, 2008.
- [20] B. P. Y. Loo, "Validating crash locations for quantitative spatial analysis: a GIS-based approach," *Accident Analysis and Prevention*, vol. 38, no. 5, pp. 879–886, 2006.
- [21] Felsburg Holt & Ullevig, "2035 Regional transportation plan update, safety analysis-procedural overview," Denver Regional Council of Governments, 2008.
- [22] X. Yan, B. Wang, M. An, and C. Zhang, "Distinguishing between rural and urban road segment traffic safety based on zero-inflated negative binomial regression models," *Discrete Dynamics in Nature and Society*, vol. 2012, Article ID 789140, 11 pages, 2012.
- [23] C. G. Prato, T. Toledo, T. Lotan, and O. Taubman-Ben-Ari, "Modeling the behavior of novice young drivers during the first year after licensure," *Accident Analysis and Prevention*, vol. 42, no. 2, pp. 480–486, 2010.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

