6-2017

Multivariate Poisson-lognormal model for analysis of crashes on urban signalized intersections approach

Mo Zhao
Iowa State University

Chenhui Liu
Iowa State University, cliu9@iastate.edu

Wei Li
University of Alabama

See next page for additional authors

Multivariate Poisson-lognormal model for analysis of crashes on urban signalized intersections approach

Abstract
Many studies investigate contributing factors of intersection crashes, but very limited studies focus on cashes on the intersection approach. It is important to address the characteristics of intersection-approach crashes to better understand intersection safety. This article analyzes the crashes on signalized intersection approach on urban arterials with a focus on traffic and geometric elements. The intersection approach is defined as the segment between stop bar and the location 200 ft upstream from the stop bar. The multivariate Poisson log-normal (MVPLN) model is used to model crash counts by severity. Ten-year crash data collected from 643 signalized intersections in Nebraska are used for analysis. One-way road is found to be negatively related to all three severity levels (light crash, moderate crash, and severe crash) of crashes. Compared to the 12 ft lane width, narrower lane widths generally lead to fewer crashes. The intersection approaches on urban arterials are expected to have more crashes than collector roads. The numbers of right-turn, left-turn, and through lanes, as well as the annual average daily traffic on the intersection approach and its crossing approach are statistically significant factors increasing crash frequency. The MVPLN model is compared to univariate and zero-inflated Poisson models. The results reveal that the MVPLN model provides a superior fit over the univariate Poisson model.

Keywords
Multivariate Poisson log-normal, signalized intersection, crash analysis, traffic safety, intersection approach

Disciplines
Civil Engineering | Multivariate Analysis | Transportation Engineering

Comments
This is an Accepted Manuscript of an article published by Taylor & Francis as Zhao, Mo, Chenhui Liu, Wei Li, and Anuj Sharma. "Multivariate Poisson-lognormal model for analysis of crashes on urban signalized intersections approach." *Journal of Transportation Safety & Security* (2017): 1-15. Available online DOI: 10.1080/19439962.2017.1323059. Posted with permission.

Authors
Mo Zhao, Chenhui Liu, Wei Li, and Anuj Sharma

Multivariate Poisson-Lognormal Model for Analysis of Crashes on Urban Signalized Intersections Approach

ABSTRACT

Many studies investigate contributing factors of intersection crashes, but few focus on crashes on intersection approaches. It is important to address the characteristics of intersection approach crashes to better understand intersection safety. This paper analyzes the crashes on signalized intersection approaches on urban arterials with a focus on traffic and geometric elements. The intersection approach is defined as the segment between the stop bar and 200 ft upstream, and the multivariate Poisson-lognormal (MVPLN) model is used to model crash counts by severity. Ten-year crash data, collected from 643 signalized intersections in Nebraska, are analyzed. It was found that one-way roads negatively impact all three crash severity levels (light, moderate, and severe), and compared to the 12 ft lane width, narrower lane widths generally lead to fewer crashes. The intersection approaches on urban arterials are expected to have more crashes than collector roads. The amount of right-turn, left-turn, and through lanes, and the annual average daily traffic (AADT) on the intersection approach and its crossing approach are statistically significant factors increasing crash frequency. The MVPLN model is compared to univariate and zero-inflated Poisson models. The results reveal that the MVPLN model provides a superior fit compared to the univariate Poisson model.

Keywords: Multivariate Poisson-lognormal, signalized intersection, crash analysis, traffic safety, intersection approach

1. INTRODUCTION

Intersection-related crashes account for over 50% of traffic crashes in urban areas. In particular, signalized intersection crashes are a major road safety problem in urban areas in the United States (Kuciemba and Cirillo 1992; Antonucci et al. 2004). Examining the charecteristics of intersection-related crashes and the knowledge of crash-influencing factors can be useful to develop intersection construction projects and implement countmeasures to improve safety (Poch and Mannering 1996; Tay and Rufaat 2007). Extensive studies have analyzed the factors contributing to intersection crashes, including traffic flow characteristics, geometric design, traffic control measures, and human factors (Worsey 1985; Lau et al. 1998; Abdel-Aty et al. 2005; Savolainen and Tarko 2005; Bao and Boyle 2009). Most of the literature on intersection-related crashes analyze a mix of rural and urban intersections; however, studies have shown that crash characteristics at urban and rural intersections are significantly different due to the variances in roadway, traffic, and environmental elements (Tay and Riffa 2007; Tay 2015). Moreover, the definition of intersection-related crashes is not consistent in literature. Intersection crashes and intersection-related crashes are used interchangeably in many studies. Generally, intersection-related crashes constitute the crashes that occur within the physical area of the intesection, i.e., the center area wrapped by the stop bars, and areas located on the approaches in close proximity to the center of the intersection. The definition of an intersection crash varies across the crash reporting systems by transportation agencies. For example, the Texas Department of Transportation (2015) defines an intersection crash as the crash in which the first harmful event occurs within the limits of an intersection, and an intersection-related crash is a crash in which the first harmful event occurs on an approach to or exit from an intersection and results from an activity, behavior, or control related to the movement of traffic units through the intersection. In Florida, the physical area of an intersection is considered to be the area within 50 ft of the center of the intersection (Das et al. 2008). In Australia, the Transport for New South Wales (2015) defines an intersection crash as “a crash where the first impact occurs at or within 10 m of an intersection.” Most commonly in the United States, crashes occurring within 250 ft (76 m) of the center of an intersection along major and minor roads are classified as intersection crashes. This classification criterion is non-arbitrary and easily repeatable and generalizable across jurisdictions (Ye et al. 2009).

The traffic patterns on intersection approaches and within the center area of intersections can be different. The crashes on intersection approaches have some unique features. In this paper, the crashes located on intersection approaches outside the intersection center area are defined as intersection approach crashes. Although numerous intersection crash analysis articles are available, few studies focus on the crash characteristics of intersection approach crashes and their influencing factors. Das et al. (2008) studied the effect of distance from the center of the intersection on characteristics of urban arterial crashes. They suggested that severity of arterial crashes (intersection crashes and midblock segment crashes) might be modeled independently of crash location if the intersection crash occurred within 100 ft of the center of intersection. This result was drawn by modeling the injury severity of a 9.72 mile arterial with 11.32 intersections per mile. However, the findings applied only to urban arterials with similar intersection densities and traffic patterns. For the arterials with low intersection densities, the injury severity model needs to take into account the difference between intersection crashes and segment crashes. Poch and Mannering (1996) explored the effect of geometric and traffic characteristics of intersection approaches on intersection approach crashes using 7-year crash data from 63 intersections in Bellevue, Washington. The found that left-turn volume, right-turn volume, and the total volume on opposing approaches increased the likelihood of intersection approach crashes. Their study provided the first comprehensive analysis of intersection approach crashes and identified effects of potential countermeasures.

In contrast to prevailing studies that examined crashes within 250 ft of the center of intersections, this study will focus on crashes that occurred on signalized intersection approaches of urban arterials and collector roads. Here, an intersection approach is defined as the road segment between the stop bar of a signalized intersection and 200 ft upstream from the stop bar. This study aims to identify the important relationships between geometric and traffic-related elements and crash count by injury severity. To account for the correlations among crash counts of different crash severity levels, the study used the multivariate Poisson-lognormal (MVPLN) model. Crash, traffic, and geometric data from 643 signalized intersections on urban arterials and collector roads in Nebraska are used in this analysis. The Bayesian framework via the Markov chain Monte Carlo (MCMC) method is used to estimate the parameters of the MVPLN model. The effects of influencing factors are evaluated. Finally, the MVPLN model is compared to univariate Poisson (UP) model, the univariate zero-inflated Poisson model (UZIP), and multivariate zero-inflated Poisson (MZIP) model in crash prediction performance.

2. METHODOLOGY

The MVPLN model can handle correlated data and address the over-dispersion (Chib and Winkelmann 2001; El-Basyouny and Sayed 2009; El-Basyounyetal 2014; Ma et al. 2008; Park and Lord 2007). Assume there are total *n* crash records. Each record contains the crash count of J crash types. Let y_{ij} denote the crash count of crash type j (j=1,2,…,J) of record i (i=1,2,..n). Assume m covariates are collected at the same time. Let $x_{i.} = (x_{i1}, x_{i2}, \dots, x_{im})$ denote the covariate vector of observation i. Let $\beta = (\beta_{1.}, \beta_{2.}, \dots, \beta_{J.})$ denote the J*m regression coefficient matrix, where $\beta_{j.} = (\beta_{j1}, \beta_{j2}, \dots, \beta_{jm})'$, is the regression coefficient vector of crash type *j*. The MVPLN model is built as follows (Park and Lord 2008):

$$y_{ij}|x_i, \beta_j, \beta_0, b_i \sim Poisson(\lambda_{ij}) \quad (1)$$

$$\ln(\lambda_{ij}) = \beta_{0j} + x_{i.} * \beta_{j.} + b_{ij} \quad (2)$$

$$b_{i.} = \begin{pmatrix} b_{i1} \\ b_{i2} \\ \dots \\ b_{iJ} \end{pmatrix} \sim N_J\left(\begin{pmatrix} 0 \\ 0 \\ \dots \\ 0 \end{pmatrix}, \Sigma\right) \quad (3)$$

where
n is the number of total records;
J is the number of crash types;
m is the number of covariates;
$i = 1,2, \dots, n$;
$j = 1,2, \dots, J$;
y_{ij}, the count of crash type j of observation i;
λ_{ij}, the mean count of crash type j of record i;
$x_{i.} = (x_{i1} \quad x_{i2} \quad \cdots \quad x_{im})$, the covariate vector of record i;
β_{0j}, the constant term of crash type j; $\beta_j = (\beta_{j1}, \beta_{j2}, \dots, \beta_{jm})'$, the regression coefficient vector of crash type j;
$b_{i.} = (b_{i1}, b_{i2}, \dots, b_{iJ})$, the error part, which is used to model the correlations between the crash counts of J types of severities of observation i. $b_{i.}$ is assumed to follow the multivariate normal distribution;
Σ, an unrestricted covariance matrix of b_i, where

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1J} \\ \sigma_{21} & \sigma_{22} & .. & \sigma_{2J} \\ \dots & \dots & \dots & \dots \\ \sigma_{J1} & \sigma_{J2} & \dots & \sigma_{JJ} \end{pmatrix}$$

In the MVPLN model, y_{ij} is assumed to independently follow the Poisson distribution when conditional on λ_{ij}. In the covariance matrix, the diagonal elements can be larger than the mean to accommodate the over-dispersion. The off-diagonal elements would take the correlation of different crash types of the same record into account (El-Basyouny et al. 2014).

3. EMPIRICAL SETTING

3.1. Data Description

The data used in this study were collected from 643 signalized intersections in Lincoln and Omaha, Nebraska. Each intersection is broken into approaches by traffic direction. The annual crash count on an intersection approach is used as an observation. A four-leg intersection can provide four observations for each year if both traffic directions are studied. In this study, only the intersection approaches on major traffic direction are considered, involving 1,126 intersection approaches from 643 signalized intersections. These intersection approaches can be classified into four types based on road functional classification (Federal Highway Administration 2013): national functional classification (NFC)14, urban principal arterial–other connecting link; NFC15, urban principal arterial–other non-connecting link; NFC16, urban minor arterial; and NFC17, urban collector. Police-reported crash data were collected over a 10-year (2003-2012) period. Since this study mainly focuses on analyzing the effects of geometric and traffic characteristics on traffic crashes, heavy vehicle crashes, animal- and alcohol-related crashes, and crashes caused by road surface conditions were excluded from the dataset. All crashes were classified into three categories by crash severity: light crash, moderate crash, and severe crash. Light crashes included property damage only (PDO) and possible injury crashes, moderate crashes included visible injury crashes, and severe crashes included disabling injury and fatal crashes. One reason for this classification is that the very low fatal crash frequency compared to crashes in other injury severity levels could create an unbalanced dataset and adversely affect the modeling. Light crashes account for 87% of all crashes. Only 3.2% of all crashes are severe crashes.

The crash data are integrated with traffic characteristics and road geometry features on the intersection approaches where the at-fault vehicle was located. Traffic-related data include AADT and speed limit on the intersection approach and its crossing road. The AADT was obtained from the Nebraska Department of Roads, and the roadway geometric data were collected from both field measurements and measurements in Google Earth. The lane widths of these intersection approaches include 9 ft, 10 ft, 11 ft, and 12 ft. The 12 ft lane width is used as the standard lane width (Transportation Research Board 2010), and three dummy variables were created for 9 ft, 10 ft, and 11 ft, respectively, in the model. The speed limits on studied approaches range from 30 mph to 45 mph, and 30 mph is taken as the baseline speed limit. Three dummy variables were created for 35 mph, 40 mph, and 45 mph, respectively. The speed limit on the crossing road is categorized as low speed (≤25mph), high speed (≥45mph), and baseline speed (30 mph and 35 mph). The low speed category accounts for more than 50% of all crossing approaches. The descriptive statistics of the variables used in forthcoming crash severity models are provided in Table 1.

3.2. Bayesian Estimation

As is shown in Equations 1 to 3, since it is very difficult to directly derive the marginal distribution of crash counts by numerical computation due to the existence of the unrestricted covariance matrix Σ, the likelihood-based methods cannot be used here for estimation. Thus, the Bayesian method with the MCMC simulation is employed to estimate parameters (Ma et al. 2008; Park and Lord 2008; El-Basyouny et al. 2014). The JAGS (Just Another Gibbs Sampler) is a software program for analyzing Bayesian hierarchical models using MCMC simulation (Plummer 2013). In JAGS, when conjugate priors are available, the Gibbs sampling is used. Otherwise, slicing sampling is used. R is a programming language and software environment for statistical computing and graphics. Package “rjags” is an interface program to run JAGS from R (Plummer 2015), and is used to estimate the parameters of the proposed MVPLN model in this study.

3.2.1 Prior Distribution

Prior distributions are needed to estimate parameters in Bayesian analysis. They usually are chosen based on the prior knowledge of the data. However, not much knowledge is available for reference in crash frequency modeling. Thus, the uninformative priors are used. Referring to the past studies (Ma et al. 2008; Park and Lord 2008; El-Basyouny et al. 2014), each element of $\beta_{j.}$ is assumed to independently follow the $N(0,1000)$ distribution, where the variance of 1000 is helpful to find the real distributions of regression coefficients in a big range. The covariance matrix is assumed to have the inverse Wishart prior, $\Sigma \sim inverse - Wishart(P, K)$, with an identity scale matrix and (J) degrees of freedom (El-Basyouny and Sayed 2009).

Deleted: f

Deleted: -1

Table 1. Descriptive statistics of model variables

Variables	**Description**	**Mean**	**Std. err.**	**Min.**	**Max.**
Dependent variables					
Light	Number of property damage only and possible injury crashes per year	1.224	1.811	0	26
Moderate	Number of visible injury crashes per year	0.146	0.421	0	5
Severe	Number of disabling injury and fatal crashes per year	0.037	0.197	0	2
Continuous independent variables					
AADTPL	AADT per lane on studied approach (1,000 vehicles)	4.335	0.018	0.070	16.75
AADTM	AADT per lane on crossing approach (1,000 vehicles)	12.327	0.012	0.154	181.575
Categorical independent variables					
City	1, Omaha (74.8%); 0, Lincoln (25.2%).				
NFC	The functional classification of target intersection approach: NFC14 - urban principal arterial–other connecting link, (14.0%); NFC15 - urban principal arterial–other non-connecting link, (30.7%); NFC16 - urban minor arterial, (46.3%); NFC17 - major collector, (8.9%).				
LLns	Number of left-turn lanes: no lane (14.4%); one lane (74.8%); two lanes (10.8%).				
ThLns	Number of through lanes: one lane (26.1%); two lanes (69.7%); three lanes (4.2%).				
RLns	Number of right-turn lanes: no lane (72.2%); one lane (27.5%); two lanes (0.35%).				
LnWd	The through lane width of target intersection approach: 9 ft (1.7%); 10 ft (12.3%); 11 ft (31.6%); 12 ft (54.5%).				
One-way	One-way road indicator: 1, one-way (1.1%); 0, two-way (98.9%).				
Legs	Number of intersection legs: three legs (6.8%); four legs (93.0%); five legs (0.3%).				
Shoulder	Shoulder indicator: 1, shoulder exists (21.2%); 0, no shoulder (78.8%).				
Median	Median indicator: 1, median exists (62.2%); 0, no median (37.8%).				
MajorSpd	The speed limit on studied approach: 30 mph (11.6%); 35 mph (31.6%); 40 mph (31.3%); 45 mph (25.5%).				

MinorSpd	The speed limit on crossing approach: ≤25 mph (29.4%); (25 mph to 45 mph) (50.8%); ≥45 mph (19.8%).
Skewangle	Skew angle indicator of the intersection: 1, 90° (81.7%); 0, < 90° (18.3%)

3.2.2 MCMC Setting

Theoretically, the prediction accuracy of parameters would increase correspondingly with the increase of the posterior sampling data, although the computing time would also increase. As a trade-off, two simulation chains are used with 4,000 iterations for each chain. The first 2,000 iterations are discarded as warmup. The second 2,000 iterations are used for parameter estimation and inference. The initial values are randomly produced by JAGS. The trace plots of estimated parameters are checked to see if the posterior samples tend to converge after the warmup iterations.

3.3 Model Comparison and Goodness of Fit

Deviance information criteria (DIC) is a generalized version of Akaike information criterion (AIC) for evaluating the hierarchical models (Spiegelhalter et al. 2002). It is often used in evaluating the goodness of fit in Bayesian analysis. The definition of DIC is (Spiegelhalter et al. 2002; Spiegelhalter et al. 2003):

$$DIC = D(\bar{\theta}) + 2pD$$

where $\bar{\theta}$ is the posterior mean of the estimated parameters, $D(\bar{\theta})$ is the deviance of the posterior mean of the parameters, and pD is defined as the difference between $\bar{D}$ and $D(\bar{\theta})$.

The small DIC is desired in model comparison. The models are thought to have significant differences when the DIC difference is larger than 5; otherwise, the difference is insignificant (Spiegelhalter et al. 2003).

In addition, although DIC can be used for comparison of different models, it cannot evaluate effectively whether a model matches the observation data. The out-of-sample validation is one of the most important goodness of fit measures. Referring to Dong et al. (2014a) and Dong et al. (2014b), the out-of-sample prediction accuracy is adopted here. In this study, nine-years (2003–2011) of crash data are used to estimate the parameters, whereas the 2012 crash data are used for out-of-sample prediction. The differences of the observed crash frequency and predicted crash frequency in 2012 are used for goodness of fit test.

4. MODEL RESULTS

4.1. Discussion of Model Variables

The estimated parameters of the MVPLN model are shown in Table 2. The mean, standard deviation, and 95% credible intervals of the posterior sampled parameters are constructed based on the posterior distributions. The 95% credible interval contains the sampled data values from the 2.5% percentile to the 97.5% percentile of the posterior distributions. Similar to the confidence interval, when the 95% credible interval contains zero, it means the variable is statistically insignificant (Gelman 2004). A negative value of the 97.5% percentile of the

Deleted: Bayesian

posterior distributions means the variable has a negative effect. That is, with the increase of the variable, the crash frequency decreases. When the 2.5% percentile value is positive, the variable has a positive effect; that is, with the increase of the variable, the crash frequency increases. Based on the mean values of estimated parameters, the expected changes in crash frequencies due to the changes of independent variables are summarized in Table 3. The calculation of expected changes did not consider the impact of interaction among the three levels of studied crashes because this impact was handled by the MVPLN model, but it was not obtainable from the model outputs.

From Table 2, the City variable shows significantly negative influence for all three levels of crashes. That is, the intersection approaches in Omaha have fewer light, moderate, and severe crashes than in Lincoln, holding all other variables constant. The annual light, moderate, and injury crashes are expected to decrease by 41.7%, 41.4%, and 45.2%, respectively, in Omaha. Many factors may be attributable for this, such as population, urban layout, and so on. Further study is needed to determine why this difference exists.

Table 2. Estimated coefficients of the MVP model

Variables	**Light crashes**	**Moderate crashes**	**Severe crashes**
City	-0.540 (-0.620, -0.453)*	-0.534 (-0.685, -0.386)*	-0.602 (-0.884, -0.323)*
NFC14	0.800 (0.630, 0.974)*	1.315 (0.959, 1.653)*	1.837 (1.062, 2.734)*
NFC15	0.220 (0.050, 0.371)*	0.571 (0.211, 0.900)*	1.078 (0.322, 1.972)*
NFC16	0.330 (0.195, 0.464)*	0.802 (0.473, 1.121)*	0.932 (0.240, 1.761)*
LLns	0.342 (0.268, 0.407)*	0.147 (0.018, 0.282)*	0.158 (-0.116, 0.442)
ThLns	0.549 (0.476, 0.613)*	0.370 (0.206, 0.530)*	0.436 (0.171, 0.726)*
RLns	0.251 (0.188, 0.313)*	0.261 (0.132, 0.384)*	0.183 (-0.062, 0.438)
LnWd9	-0.503 (-0.756, -0.262)*	-0.410 (-0.935, 0.075)	-1.241 (-2.700, -0.094)*
LnWd10	-0.311 (-0.416, -0.277)*	-0.385 (-0.588,- 0.186)*	-0.437 (-0.815, -0.055)*
LnWd11	-0.347 (-0.416, 0.277)	-0.239 (-0.382, -0.100)*	-0.167 (-0.442, 0. 096)
One-way	-0.535 (-0.786, -0.286)*	-0.575 (-1.108, -0.092)*	-1.596 (-2.855, -0.502)*
AADTPL	0.215 (0.155, 0.324)*	0.174 (0.147, 0.199)*	0.135 (0.079, 0.183)*
Legs	0.245 (0.135, 0.303)*	0.218 (-0.143, 0.522)	0.572 (0.195, 0.906)*
Shoulder	-0.013 (-0.086, 0.060)	0.102 (-0.041, 0.244)	-0.181 (-0.479, 0.105)
Med	-0.159 (-0.245, -0.073)*	-0.053 (-0.229, 0.113)	-0.205 (-0.536, 0.107)
MajorSpd35	-0.040 (-0.155, 0.075)	-0.089 (-0.322, 0.153)	-0.243 (-0.671, 0.241)
MajorSpd40	-0.037 (-0.150, 0.089)	-0.054 (-0.288, 0.194)	-0.334 (-0.779, 0.183)
MajorSpd45	-0.417 (-0.549, 0.285)	-0.633 (-0.895, -0.360)*	-0.766 (-1.258, -0.213)*
AADTM	0.008 (0.006, 0.010)*	0.007 (0.003, 0.011)*	0.001 (-0.008, 0.009)
MinorSpdLow	-0.596 (-0.668, -0.520)*	-0.456 (-0.600, -0.308)*	-0.360 (-0.652, -0.074)*
MinorSpdHigh	-0.311 (-0.389, -0.237)*	-0.434 (-0.595, -0.269)*	-0.493 (-0.830, -0.168)*
Skewangle	0.145 (0.081, 0.213)*	0.064 (-0.070, 0.200)	0.034 (-0.223, 0.308)
Constant	-3.121 (-3.419, -2.703)*	-4.629 (-5.917, -2.909)*	-7.115 (-8.548, -5.327)*

Note: the values are the posterior means. The values in parentheses show the 95% credible intervals. * means the variables are significant at 95% credible level.

Compared with the major collector (NFC17), urban principal arterial–other connecting link (NFC14), urban principal arterial–other non-connecting link (NFC15), and urban minor arterial (NFC16) show significantly positive influences on all three crash levels. That is, the approaches on these roads are expected to have more crashes. In addition, the intersection approaches located in the urban principal arterial–other connecting link (NFC14) are expected to have the most crashes. It is understandable since these roads are supposed to have larger and more complex traffic than other roads.

Both the number of the left-turn lanes and right-turn lanes show significantly positive influences on the light and moderate crashes, but not the severe crashes. The number of through lanes shows significantly positive influences on all three crash types.

Compared to the 12 ft through lane width, a narrower lane would generally lead to fewer crashes. In addition, it appears that the narrower the lane width, the less likely crashes will occur. Similar results have been found in the study of the impacts of lane width on segments in Nebraska (Wood et al. 2015). Also, another study of urban Nebraska suggested that the combination of no left-turn lanes and narrowed through lanes reduced the crash frequency compared to the combination of no left-turn lanes and standard through lanes on roadways with a speed limit of 35 mph outside CBD (Sharma et al. 2015). A possible explanation is that the drivers may drive more cautiously on the narrow lanes. However, the influences of the narrow lane widths on crashes still show slight differences. The 9 ft lane width has significantly negative influences on light and severe injury crashes, but the 11 ft lane width only has a significantly negative influence on the moderate crashes. The 10 ft lane width shows significantly negative influences on all three crash levels.

Deleted: in

Deleted: ,

Deleted: the total crash frequencies on intersection approaches are modeled by groups classified by lane width, where speed limit and left-turn lane, and the results

Table 3. Expected changes in crash frequency corresponding to exploratory variables

Variables	**Light crashes**	**Moderate crashes**	**Severe crashes**
City	-41.7%	-41.4%	-45.2%
NFC14	122.6%	272.5%	528.0%
NFC15	24.5%	77.1%	194.0%
NFC16	39.1%	123.0%	153.9%
LLns	40.8%	15.8%	-
ThLns	73.1%	44.7%	54.7%
RLns	28.6%	29.8%	-
LnWd9	-39.6%	-	-71.1%
LnWd10	-26.7%	-32.0%	-35.4%
LnWd11	-	-21.2%	-
One-way	-41.4%	-43.7%	-79.7%
AADTPL	23.9%	19.1%	14.5%
Legs	27.7%	-	77.3%
Shoulder	-	-	-
Med	-14.7%	-	-
MajorSpd35	-	-	-
MajorSpd40	-	-	-
MajorSpd45	-	-46.9%	-53.5%

AADTM	0.8%	0.7%	-
MinorSpdLow	-44.9%	-36.6%	-30.3%
MinorSpdHigh	-26.8%	-35.2%	-38.9%
Skewangle	15.6%	-	-

Note: "-" means the variable is not significant at 95% confidence level.

The one-way street shows significantly negative influences on all three crash levels. That is, fewer crashes are expected on the one-way streets compared to two-way traffic.

The AADT on the studied approach shows significantly positive influences on all three crash levels, and the AADT of the crossing approach appears to increase the amount of light and moderate crashes. This is consistent with many studies on intersection crashes. The number of intersection legs also shows significantly positive influences on all three crash levels. It is thought that with an increase in the number of intersection legs, the traffic condition is expected to be more complex and, thus, the collision risk is expected to increase.

The presence of a shoulder does not show any significant influences on intersection approach crashes. This might be because the shoulder, which is usually narrower on urban streets, may not have as much of a safety effect. The median only shows significantly negative influences on the light crashes. It is thought that the median barrier would reduce sideswipe and head-on crashes.

The speed limit is usually positively related to the operating speeds. Speed limits less than 45 mph do not show any significant influence on intersection approach crashes. Only the speed limit of 45 mph has significantly negative influences on all three crash levels. A possible explanation is that the drivers may drive with more caution when approaching higher speed intersections. Both the low speed limit (⩽25 mph) and high speed limit (⩾45 mph) on the crossing road show significantly negative influences on all three crash levels. When the crossing approach speed limit is ⩽25 mph, the light, moderate, and severe crashes decreased by 44.9%, 36.6%, and 30.3%, respectively. When the crossing approach speed limit is ⩾45 mph, the light, moderate, and severe crashes decreased by 26.8%, 35.2%, and 38.9%, respectively. The study by Poch and Mannering (1996) found the higher the approach speed limit, the higher the total crash frequency, and the crossing approach speed limit is negatively related to total intersection approach frequency.

The skew angle of 90° has significantly positive influences on light crashes. Although this result seems counterintuitive, some past studies have similar findings (Wang 2006). A possible explanation is that longer all-red signal durations are set for these skewed intersections. However, we cannot say that the skew intersections are generally safer. Tay (2015) modeled crash frequency using five-year data collected from 4,201 urban intersections in Alberta, Canada, and found crashes at skewed intersections were more likely to occur in urban than rural areas. Actually, the skew intersection should always be avoided in field design (Antoucci et al. 2004).

4.2. Comparison of model performance

The crashes in 2012 predicted by the MVPLN model are shown in Table 4. For comparison, the 2012 crashes predicted by three univariate Poisson (UP) models (Lord and Mannering 2010) using the same variables are also included. In addition, since the zero crash proportions of all three crash levels are high (51.0%, 89.7%, and 96.6% for light, moderate, and severe crashes, respectively), the univariate zero-inflated Poisson model (UZIP) (Dong et al. 2014b) and the multivariate zero-inflated Poisson model (MVZIP) (Li et al. 1999) are also used to predict the crashes for comparison. The UP is the most commonly used model for crash frequency modeling, while the UZIP model is a variant of the UP model to account for the over-dispersion due to the excess zero data (Lord and Mannering 2010; Aguero-Valverde 2013; Mannering and Bhat 2014). The MVZIP model is the multivariate version of the UZIP model (Dong et al. 2014a; Dong et al. 2014b). The parameters of the UP, UZIP, and MVZIP models are also estimated with the Bayesian method. The DIC values and prediction results are shown in Table 4:

- The UZIP model shows better prediction in the severe crashes than the UP model, but it has no significant differences in predicting the light and moderate crashes. This is reasonable since only the severe crashes have extremely high zero-crash proportion.
- The MVZIP model appears to not perform as well as the UZIP model. A possible explanation is that the MVZIP is excessively complex and may over fit the data in this study. The two zero-inflated models do not perform better than the MVPLN model for light and moderate crashes. The zero-crash proportions in this dataset may not be high enough to utilize the zero-inflation structure for modeling light and moderate crashes.
- Compared to the other three models, the prediction accuracy of the MVPLN model is superior for the light and moderate crashes, but slightly worse for the severe crashes. This might be caused by the large amount of zero-crash counts for server crashes. The total prediction accuracy is highest for the MVPLN model. In addition, the DIC of the MVPLN model is also the smallest. Generally speaking, the MVPLN performs better than all other models, which is consistent with previous studies (Ma et al. 2008; Park and Lord 2008; Aguero-Valverde and Jovanis 2010). The superiority of the MVPLN model lies in that it addresses the unobserved heterogeneity and correlation across the three types of crashes.

Table 4. Summary of model assessment

Model	**Light crashes**	**Moderate crashes**	**Severe crashes**	**DIC**
Observed	1,181	134	43	
MVPLN	1,033 (-12.5%)	149 (+11.2%)	38 (-11.6%)	34,286
UP	1,353 (+14.6%）	161 (+20.1%)	39 (-9.3%)	42,086
UZIP	1,356 (+14.8%)	162 (+20.9%)	44 (+2.3%)	34,624
MVZIP	1,391 (+17.8%)	166 (+24.1%)	40 (-6.3%)	38,624
Zero-crash (%)	50.9	89.7	96.6	-

Note: the values in parentheses represent the percentage difference between the observed and predicted values. For the UP model, the DIC is the sum of three individual DICs, which are 30,488, 8,457, and 3,118 for light, moderate, and severe crashes. The DIC of the UZIP model is also the sum of the DICs of three individual UZIP models.

5. CONCLUSIONS

Intersection safety has been a major focus of transportation safety research. Most intersection safety studies analyze the crashes that occurred in close vicinity of intersections, usually within

250 ft of the center of the intersection. The traffic operation at signalized intersections is complex, and the crash risk is associated with many traffic and geometric factors. The traffic conflict types could be different on intersection approach and the center of intersection areas, therefore, the crash characteristics on intersection approach might be unique. However, there are limited studies concentrating on crashes on intersection approach, especially for signalized intersections on urban arterials.

This study analyzes the crashes that occurred on signalized intersection approaches on urban arterials and collector roads. Crash data were collected for 634 signalized intersection approaches over 10 years in the urban areas of Lincoln and Omaha, Nebraska. These crashes are classified into light crashes, moderate crashes, and severe crashes based on injury severity. Considering the possible correlations between the three crash levels, the MVPLN model is built to quantify the influences of various factors on these three categories of crashes simultaneously. Compared with modeling the three crash levels separately, the MVPLN model produces more reliable results.

The city, road functional classification, lane number, and lane width are found to have significant influences on frequency of intersection approach crashes. The approaches with the narrow lane widths are generally found to have fewer crashes. With the trend of designing according to Complete Streets, there is increasing use of narrowed lane width, especially in urban areas. This finding could be used to further identify the safety impact of narrow lane width on urban arterials. In addition, the intersection approach crashes are found to be significantly influenced by the speed limit and AADT of the intersecting road. The speed limit of the studied approaches and the crossing approaches show different influences on crashes. The positive coefficients estimated for AADT on both approaches indicate crashes are more likely under high traffic volume than low volume conditions.

Due to the limitation of data, this study does not include traffic control variables. Traffic signal timing could be considered because it can affect drivers' decisions when approaching an intersection (Agbelie and Roshandeh 2015). For example, unreasonable signal timing may increase rear-end crashes and collision risks of passing the dilemma zone. Signal coordination decreases vehicle stop-and-go and thus decreases the potential of rear-end crashes. In addition, this study focuses little on the interactive influences of various factors, but the combinations of geometric characteristics may show better prediction results (Gross et al. 2009). However, the interactive effects of variables may make the model more complex and difficult to explain. A trade-off between the interpretation ability and complexity of the model is considered here.

Based on the findings in this paper, future studies may be conducted to analyze intersection approach crashes by crash type, e.g., rear-end and sideswipe crashes, and to compare the characteristics of intersection approach crashes with those that occurred within the physical area of signalized intersections. The crash patterns in terms of both frequency and severity may be significantly different due to different traffic conflicting mechanisms in the two areas.

The model developed in this paper can be used for screening high intersection-related crash locations. The findings could be used to further understand the traffic and geometric elements contributing to intersection-related crashes in different severities. This knowledge can then be applied to intersection design and implementation of countermeasures to improve intersection safety. However, the possible temporal correlations due to multiyear crashes on the same intersection approach and spatial correlation among nearby intersections are not considered in this paper. Future studies are needed to address the effects of spatial-temporal correlations.

REFERENCES

Abdel-Aty, M., & Keller, J. (2005). Exploring the Overall and Specific Crash Severity Levels at Signalized Intersections, *Accident Analysis & Prevention*, 37, 417-425.

Agbelie, Bismark R. D. K. & Arash M. Roshandeh. (2015). Impacts of Signal-Related Characteristics on Crash Frequency at Urban Signalized Intersections, *Journal of Transportation Safety & Security*, 7:3, 199-207.

Antonucci, N. D., Hardy, K. K., Slack, K. L., Pfefer, R., & Newman, T. R. (2004). *A guide for reducing collisions at signalized intersections*. National Co-operative Highway Research Program (NCHRP) Report 500 Volume 12.Washington D.C.: Transportation Research Board of the National Academies.

Bao, S. & Linda Boyle. (2009). Age-related differences in visual scanning at median-divided highway intersections in rural areas. *Accident Analysis & Prevention*, 41, 146-152.

Caliendo, C., & Maurizio Guida. (2014). A New Bivariate Regression Model for the Simultaneous Analysis of Total and Severe Crashes Occurrence. *Journal of Transportation Safety & Security*, Vol. 6, Issue 1, 78-92.

Chib, S., and R. Winkelmann. (2001). Markov Chain Monte Carlo Analysis of Correlated Count Data, *Journal of Business and Economic Statistics*, 19, 428–435.

Das, A., Anurag Pande, Mohamed Abdel-Aty, & Joseph B. Santos. (2008). Characteristics of Urban Arterial Crashes Relative to Proximity to Intersections and Injury Severity. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2083, 137–144.

Dong, C., Clarke, D.B., Yan, X., et al. (2014a). Multivariate random parameters zero-inflated negative binomial regression model: an application to estimate crash frequencies at intersections. *Accident Analysis & Prevention*, 70, 320-329.

Dong, C., Richards, S.H., Clarke, D.B., et al. (2014b). Examining signalized intersection crash frequency using multivariate zero-inflated Poisson regression. *Safety Science*, 70, 63-69.

Dong, C., Clarke, D.B., Richards, S.H., et al. (2014c). Differences in passenger car and large truck involved crash frequencies at urban signalized intersections: an exploratory analysis. *Accident Analysis & Prevention*, 62, 87-94.

El-Basyouny, K., S. Barua, and M. T. Islam. (2014). Investigation of time and weather effects on crash types using full Bayesian multivariate Poisson lognormal models, *Accident Analysis & Prevention*, 73, 91–99.

El-Basyouny, K. and T. Sayed. (2009). Collision prediction models using multivariate Poisson-lognormal regression. *Accident Analysis & Prevention*, 41, 820–828.

Gelman, A. (2004). *Bayesian Data Analysis*, 2nd ed., Chapman & Hall/CRC, Boca Raton, FL.

Gross, F., P. P. Jovanis, and K. Eccles. (2009). Safety Effectiveness of Lane and Shoulder Width Combinations on Rural, Two-Lane, Undivided Roads. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2103, 42–49.

Kuciemba, S. R., and J. A. Cirillo. (1992). *Safety Effectiveness of Highway Design Features, Volume V: Intersections.* Report No. FHWA-RD-91-048. Federal Highway Administration. Washington, D.C.

Lau, M.Y, & May, A.D. (1988). *Accident Prediction Model Development: Signalized Interections Final Report.* Institute of Transportion Studies, University of California, Berkeley, Berkeley, CA.
Li, C., Lu, J., Park, J., Kim, K., Brinkley, P. a, & Peterson, J. P. (1999). Multivariate Zero-Inflated Poisson Models and Their Applications. *Technometrics*, *41*(1), 29–38.
Ma, J., K. M. Kockelman, and P. Damien. (2008). A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods, *Accident Analysis & Prevention*, 40, 964–975.
Park, E. S. and D. Lord. (2007). Multivariate Poisson-Lognormal Models for Jointly Modeling Crash Frequency by Severity. *Transportation Research Record: Journal of the Transportation Research Board,* No. 2019, 1–6.
Plummer, M. (2013). JAGS Version 3.4.0 user manual.
Martyn Plummer (2016). rjags: Bayesian Graphical Models using MCMC. R package version 4-6. https://CRAN.R-project.org/package=rjags

Poch, M. & F. Mannering. (1996) Negative Binomial Analysis of Intersection-Accident Frequencies. *Journal of Transportation Engineering*, 122(2): 105-113.
Savolainen, P. & Tarko, A. (2005). Safety Impacts at Intersections on Curved Segments, *Transportation Research Record: Journal of the Transportation Research Board,* No. 1908, 130–140.
Sharma, A., Li, W., Zhao, M., & Rilett, L. R. (2015). *Safety and operational analysis of mid-block segment lane width and intersection approach in the Urban Environment in Nebraska*. Nebraska Transportation Center, Lincoln, NE
Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde. (2002). Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society: Series B (Statistical Methodology*, 64(4), 583–639.
Spiegelhalter, D., A. Thomas, N. Best, and D. Lunn. (2003). *WinBUGS User Manual*. Cambridge, England: Medical Research Council Biostatistics Unit, 1–60.
Tay, Richard. (2015). A random parameters probit model of urban and rural intersection crashes. *Accident Analysis and Prevention*, 84, 38–40.
Tay, R., & Rifaat, S. M. (2007). Factors contributing to the severity of intersection crashes. *Journal of Advanced Transportation*, 41(3), 245–265.
Transport for NSW. (2015). Definitions and notes to support road crash data. NSW Center for Road Safety.
Texas Department of Transportation. (2015). Classification of Motor Vehicle Traffic Crash in Texas. 2015 Edition. Texas Department of Transportation Traffic Operations Division.
Wang, Xuesosng. (2006). Safety analyses at signalized intersections considering spatial, temporal and site correlation. Ph.D. Thesis. University of Central Florida, Orlando, FL.
Wood, J. S., J. P. Gooch, and E. T. Donnell. (2015). Estimating the safety effects of lane widths on urban streets in Nebraska using the propensity scores-potential outcomes framework, *Accident Analysis & Prevention*, 82, 180–191.
Worsey, G. (1985). Predicting Urban Accident Rates from Road and Traffic Characteristics. *ITE Journal*, 55(12), 37-40.
Ye, X., R. M. Pendyala, S. P. Washington, K. Konduri, & J. Oh. (2009). A simultaneous equations model of crash frequency by collision type for rural intersections. *Safety Science*, 47 (3), 443–452.

Deleted: Plummer, M. (2015). Package “rjags”.