Statistical modeling of total crash frequency at highway intersections

Arash M. Roshandeh a,*, Bismark R. D. K. Agbelie b, Yongdo Lee c

a Department of Engineering and Public Works, City of Alpharetta, GA 30009, USA
b School of Civil Engineering, Purdue University, West Lafayette, IN 47907, USA
c Department of Civil and Environmental Engineering, The Pennsylvania State University, University Park, PA 16802, USA

A R T I C L E   I N F O

Article history:
Available online 15 March 2016

Keywords:
Total crash frequency
Random-parameter count model
Intersection
Defected pavement

A B S T R A C T

Intersection-related crashes are associated with high proportion of accidents involving drivers, occupants, pedestrians, and cyclists. In general, the purpose of intersection safety analysis is to determine the impact of safety-related variables on pedestrians, cyclists and vehicles, so as to facilitate the design of effective and efficient countermeasure strategies to improve safety at intersections. This study investigates the effects of traffic, environmental, intersection geometric and pavement-related characteristics on total crash frequencies at intersections. A random-parameter Poisson model was used with crash data from 357 signalized intersections in Chicago from 2004 to 2010. The results indicate that out of the identified factors, evening peak period traffic volume, pavement condition, and unlighted intersections have the greatest effects on crash frequencies. Overall, the results seek to suggest that, in order to improve effective highway-related safety countermeasures at intersections, significant attention must be focused on ensuring that pavements are adequately maintained and intersections should be well lighted. It needs to be mentioned that, projects could be implemented at and around the study intersections during the study period (7 years), which could affect the crash frequency over the time. This is an important variable which could be a part of the future studies to investigate the impacts of safety-related works at intersections and their marginal effects on crash frequency at signalized intersections.

© 2016 Periodical Offices of Chang’an University. Production and hosting by Elsevier B.V. on behalf of Owner. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Intersection-related crashes are associated with high proportion of accidents involving drivers, occupants, pedestrians, and cyclists. A significant portion of total fatal crashes usually occur at intersections. In order to enhance the safety of intersections, significant attention is needed to ensure safe movement of road users. In general, the purpose of
interaction safety analysis is to determine the impact of safety-related variables on pedestrians, cyclists and vehicles, so as to facilitate the design of effective and efficient countermeasure strategies to improve safety at intersections.

In the U.S., Illinois is one of the states with considerable number of crashes. For example, in 2010, 927 persons were killed, and 88,937 people were injured in crashes. Out of the total fatal crashes, 25.6% occurred at intersections. Besides, in the previous three years (2007, 2008, and 2009), an average of over 26% fatalities occurred at intersections on average (IDOT, 2013).

Although, urban intersections have been received significant attention in terms of signals’ timing optimization to minimize vehicles’ delay (Dong et al., 2014c; Nesheli et al., 2009) or to simultaneously minimize the delay of both vehicles and pedestrians (Roshandeh et al., 2014). Ignoring the expected safety effects may not yield the overall performance-based benefits. Compared with vehicle drivers, pedestrians and cyclists are much more vulnerable in intersection crashes due to their interactions with vehicles (Dong et al., 2014a; Zhou et al., 2014). Evidently, pedestrians comprise more than 22% of the 1.24 million people killed in traffic accidents worldwide (World Health Organization, 2014).

In order to investigate contributing factors on total crash frequency at intersections, this study intends to analyze effects of traffic, environmental, intersection geometric and pavement-related characteristics using a 7-year crash data from 357 randomly selected intersections in Chicago.

Over the past years, several studies have been conducted on intersection safety analysis. Agbelie and Roshandeh (2015) applied a random-parameters negative binomial model to investigate the impacts of signal-related factors on crash frequency. The results showed that increasing the number of signal phases and traffic volume at an intersection would increase crash frequency, whereas, increasing the number of approach lanes and the maximum green time would also lead to an increase in crash frequency at many intersections. Wu et al. (2013) used intersections crash data, all with intersection approaches having signal-warning flashers, to estimate a random-parameter negative binomial model of crash frequency. The estimation results revealed that a 5 mi/h speed-limit reduction decreases the frequency of crashes in some cases, whereas in other cases, it increases crash frequency. Oh et al. (2010) used a traffic conflict technique to evaluate traffic safety at signalized intersections and analyzed traffic conflicts that occurred at the time of a signal violation. By using image procedure technology, traffic images from two intersections in South Korea were analyzed and found that serious and dangerous conflicts took place at the time of signal violation. Gomes et al. (2012) applied Poisson-gamma modeling framework to develop predictive models for estimating safety performance of signalized and unsignalized intersections in Lisbon, Portugal. They put forward that highway geometric characteristics affect crash severity occurring at urban three- and four-leg intersections. Zhou et al. (2013) proposed a root cause degree procedure to measure intersection safety in Shanghai and found that clearance time, safety education, enforcement, trajectory inside the intersection, crossing a refuge island, speeding, and the right turn control pattern affect crash frequency. Wang and Abdel-Aty (2006) used the generalized estimating equations with the negative binomial link function to model rear-end crash frequencies at signalized intersections to investigate the crash temporal or spatial correlation among the data. Results showed that intersections with heavy traffic, more right and left-turn lanes, large number of phases per cycle, high speed limits, and in high population areas are more likely to have higher rear-end crashes. Das and Abdel-Aty (2011) applied genetic programming technique to analyze rear-end crash counts and found that crashes decreased with an increase in skid resistance during morning peak hours, whereas the crashes increased in the afternoon peak period. Dong et al. (2014a) investigated the contributing factors on crash frequency at urban signalized intersections and found that compared with the univariate Poisson-lognormal (UVPLN) and multivariate Poisson (MVP) models, the multivariate Poisson-lognormal (MVPLN) model better identifies significant factors and predicts crash frequencies. Their analysis suggests that traffic volume, truck percentage, lighting condition, and intersection angle significantly affect intersection safety. Park and Lord (2007) employed a new multivariate approach for modeling data on crash counts by severity based on MVPLN models. The method was applied to the multivariate crash counts from 451 intersections in California obtained in 10 years. It showed that the new MVPLN regression approach could address both overdispersion and a fully general correlation structure in the data. El-Basyouny and Sayed (2013) used a dataset corresponding to 51 signalized intersections in British Columbia to investigate the relationship between conflicts and collisions. A lognormal model was employed to predict conflicts and a conflicts-based negative binomial (NB) safety performance function was then used to predict collisions. Data on collision frequency, average hourly conflicts, average hourly volumes, area type (urban/suburban), the number of through lanes and the presence of right and left turn lanes were used as explanatory variables. The results showed that the effects of conflicts on collisions are non-linear with decreasing rates. Dong et al. (2014b) employed a multivariate random-parameters zero-inflated negative binomial model for crash frequency modeling, which showed that this method performed better than Poisson, negative binomial, and Poisson-lognormal models. In another study, Dong et al. (2014c) used a Bayesian a multivariate zero-inflated Poisson model and proved that it would address correlations among various crash severity levels and properly handles observations with zero crashes.

The existing literature have extensively accounted for impacts of different variables on crash frequency at intersections. However, in order to investigate contributing effects of pavement condition and intersection’s work zones on crash frequency, further research is still required. As such, this study endeavors to fill this gap and analyze impacts of pavement condition, intersection’s work zone, traffic and environmental characteristics on total crash frequency at intersections.

The dataset available for the present study includes crash data of 357 intersections in Chicago, collected from 2004 to 2010. For the purpose of investigating safety impacts at these
intersections, each intersection was considered as an observation, with the preservation of distinct geometric, pavement and traffic-flow characteristics for each approach. Meanwhile, the crash frequency of each intersection for each year was also considered as an observation for the 357 intersections, generating 2499 records since each intersection has 7 years of crash data. The primary variables of interest were traffic, environmental, intersection geometric and pavement-related characteristics. The variables selected in the final model were those observed to be statistically significant and were provided in Table 1.

<table>
<thead>
<tr>
<th>Variable description</th>
<th>Mean</th>
<th>Std. dev.</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work zone (1 if yes, 0 otherwise)</td>
<td>0.163</td>
<td>0.369</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Average annual traffic on major road (in thousands)</td>
<td>65.824</td>
<td>28.289</td>
<td>23.821</td>
<td>298.861</td>
</tr>
<tr>
<td>Snowy weather condition (1 if yes, 0 otherwise)</td>
<td>0.502</td>
<td>0.500</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Average annual traffic on minor road (in thousands)</td>
<td>37.185</td>
<td>13.964</td>
<td>4.901</td>
<td>99.750</td>
</tr>
<tr>
<td>Intersection without street lights in the evening (1 if yes, 0 otherwise)</td>
<td>0.574</td>
<td>0.495</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Weekday evening peak periods (5 p.m.−7 p.m.) (1 if yes, 0 otherwise)</td>
<td>0.799</td>
<td>0.401</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Defected pavement surface (1 if yes, 0 otherwise)</td>
<td>0.805</td>
<td>0.396</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Number of crashes</td>
<td>17.40</td>
<td>4.11</td>
<td>2</td>
<td>106</td>
</tr>
</tbody>
</table>

2. Methodology

Crash frequency analysis at intersections is a count data analysis, and count data modeling approach is generally adopted in the modeling process. It can be observed that crash numbers are nonnegative in nature, and the assigned crash frequencies are nonnegative integers, thus, using a count data modeling technique will be the most appropriate (Lord and Mannering, 2010). A number of statistical approaches, including Poisson regression, negative binomial, zero-inflated Poisson regression, and zero-inflated negative binomial, have been used to model count data (Washington et al., 2011). In the Poisson framework, the probability \( P(\gamma_k) \) of intersection \( k \) having \( \gamma_k \) crashes per year is shown as follow

\[
P(\gamma_k) = \exp(-\phi_k) \phi_k^{\gamma_k} / \gamma_k!
\]

where \( \phi_k \) is the assigned Poisson parameter for intersection \( k \), which is intersection \( k \)'s expected crash frequencies, \( E(\gamma_k) \).

Generally, the Poisson regression specifies the crash frequency parameter \( \phi_k \) as a function of independent variables using a log-linear function.

\[
\phi_k = \exp(a X_k)
\]

where \( a \) is a vector of estimable parameters, \( X_k \) is a vector of independent variables (Washington et al., 2011).

The Poisson distribution constraints the variance and mean to be equal, such that \( \text{VAR}(\gamma_k) \) equals \( E(\gamma_k) \). However, if this equality is violated, the data can be considered as either overdispersed (\( \text{VAR}(\gamma_k) > E(\gamma_k) \)) or underdispersed (\( \text{VAR}(\gamma_k) < E(\gamma_k) \)). The estimated parameter vector standard errors will be incongruous resulting in incongruous inferences. In order to account for the possibility of overdispersion, the negative binomial model is derived by modifying Eq. (2) as follow

\[
\phi_k = \exp(a X_k + \tau_k)
\]

where \( \exp(\tau_k) \) is a gamma-distributed error term with mean \( 1 \) and variance \( \theta \). The justification for adding this term allows the flexibility of the variance to vary from the mean as follow

\[
\text{VAR}(\gamma_k) = E(\gamma_k)[1 + \beta E(\gamma_k)] = E(\gamma_k) + \beta [E(\gamma_k)]^2
\]

The negative binomial probability density function has the form as follow

\[
P(\gamma_k) = \left( \frac{1}{\beta} \right)^{\frac{1}{\beta} + \gamma_k} \Gamma \left( \frac{1}{\beta} + \gamma_k \right) \left( \frac{\phi_k}{\beta + \phi_k} \right)^{\frac{1}{\beta} + \gamma_k} \left( \frac{1}{\beta + \phi_k} \right)^{\frac{1}{\beta}}
\]

where \( \Gamma(\cdot) \) is a gamma function, \( \beta \) is an overdispersion factor.

As \( \beta \) approaches zero, the Poisson regression becomes a limiting model of the negative binomial regression. Thus, if \( \beta \) is significantly different from zero, the use of the negative binomial is appropriate, and if it is not, the Poisson model is appropriate (Washington et al., 2011). In order to account for the possible heterogeneity, which may vary from one intersection to another, random-parameter model can be introduced. Greene (2007) developed an estimation procedures (using simulated maximum likelihood estimation) for incorporating random-parameters in count-data models. To develop a random-parameter model that accounts for possible unobserved heterogeneity across intersections, the individual estimable parameters are written as follow

\[
a_{ij} = a + \omega_i
\]

where \( \omega_i \) is a randomly distributed term for each signalized intersection \( i \) and it can take on a wide variety of distributions, such as Weibull, Erlang, logistic, log-normal, normal, etc.

By using Eq. (5), the crash frequency parameter (Poisson parameter), \( \phi_k \), becomes \( \phi_{k|\omega_i} = \exp(a X_k + \tau_k) \) in the negative binomial with the corresponding probabilities for Poisson or negative binomial now \( P(\phi_{k|\omega_i}) \). The log-likelihood function (LL) for the random-parameter negative binomial in this case can be written as follow

\[
LL = \sum_{i} \ln \int_{\omega_i} g(\omega_i) P(\phi_{k|\omega_i}) d\omega_i
\]

where \( g(\cdot) \) is the probability density function of the \( \omega_i \).

Because maximum likelihood estimation of the random-parameter negative binomial models is computationally cumbersome (due to the required numerical integration of the
negative binomial function over the distribution of the random-parameters), a simulation-based maximum likelihood method is used (the estimated parameters are those that maximize the simulated log-likelihood function while allowing for the possibility that the variance of \( \omega_i \) for intersection-level parameters is significantly greater than zero). The most popular simulation approach uses Halton draws, which has been shown to provide a more efficient distribution of draws for numerical integration than purely random draws (Greene, 2007). Finally, to assess the impact of specific variables on the mean number of crashes, marginal effects are computed (Washington et al., 2011). Marginal effects are computed for each observation and then averaged across all observations. The marginal effects give the effect that a one-unit change in \( x \) has on the expected number of crashes at each approach, \( \gamma_k \).

3. Estimation results

In order to investigate the appropriate model, the random-parameter Poisson and negative binomial models were developed. The models were estimated using simulation-based maximum likelihood with 200 Halton draws. The number of Halton draws was selected because it has been proven to produce consistent and accurate parameter estimates (Agbelie, 2014; Bhat, 2003; Mannering and Bhat, 2014; Milton et al., 2008). Halton draws were used for the simulation instead of random draws, because it has been shown that fewer Halton draws are required to attain convergence compared to random draws (Train, 2003). Also, the efficiency of Halton draws is generally significant relative to random draws. In order to select the random-parameter density functional forms, the following distributions were investigated: uniform, triangular, lognormal and normal distributions. However, the normal distributions were found to yield the best statistical fit for all the parameters. Nonetheless, future studies could be conducted to further investigate and compare the different distributions for random-parameter model in accident analysis.

Out of the two models, the random-parameter Poisson model was found to be the most appropriate, because the overdispersion factor \( \rho \) (8.155 \times 10^3) was found to be statistically insignificant (t-statistic of 0.00001). Also, when the fixed-parameter Poisson model was initially investigated, the McFadden \( \rho^2 \) statistic was 0.11, while a random-effect Poisson model yielded a McFadden \( \rho^2 \) statistic of 0.90. However, the results from the random-parameter Poisson model indicate that the model has the best statistical fit (McFadden \( \rho^2 \) statistic of 0.93) and the estimated parameters are of plausible sign and magnitude. The estimated results and marginal effects from the random-parameter Poisson model are presented in Table 2. In order to determine if a parameter is random, the standard deviation of the parameter density has to be statistically significant. However, if the estimated standard deviation is not statistically significant (statistically different from zero), the parameter is considered to be fixed across the population of intersections. The estimation results presented in Table 2 reveal that eight parameters are statistically significant (including constant term), and six of them (constant, average annual traffic on major road (in thousands), snowy weather condition, intersection without street lights in the evening, weekend evening peak periods, and defected pavement surface), are found to yield statistically significant random-parameters (estimated standard deviation for each parameter distribution is found to be significantly different from zero), while two parameters (work zone location, and average annual traffic on minor road) are fixed across the population of intersections.

Examination of the results shows that increasing the number of intersections located at work zones would increase crash frequency. This variable produced a positive fixed parameter. A unit increase in the number of intersections located at work zones would increase the mean crash rate by 1.184 (as shown by the marginal effect value in Table 2). This indicates that intersections located at work zones are more

<table>
<thead>
<tr>
<th>Variable description</th>
<th>Estimated parameter</th>
<th>t-statistic</th>
<th>Marginal effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant (standard deviation of parameter distribution)</td>
<td>2.176 (0.129)</td>
<td>105.510 (28.102)</td>
<td></td>
</tr>
<tr>
<td>Work zone location (1 if yes, 0 otherwise)</td>
<td>0.064</td>
<td>5.222</td>
<td>1.184</td>
</tr>
<tr>
<td>Average annual traffic on major road (in thousands) (standard deviation of parameter distribution)</td>
<td>7.681 \times 10^{-4} (8.096 \times 10^{-4})</td>
<td>4.624 (12.696)</td>
<td>0.014</td>
</tr>
<tr>
<td>Snowy weather condition (1 if yes, 0 otherwise) (standard deviation of parameter distribution)</td>
<td>0.106 (0.052)</td>
<td>11.540 (8.411)</td>
<td>1.971</td>
</tr>
<tr>
<td>Average annual traffic on minor road (in thousands)</td>
<td>0.003</td>
<td>9.524</td>
<td>0.058</td>
</tr>
<tr>
<td>Intersection without street lights in the evening (1 if yes, 0 otherwise) (standard deviation of parameter distribution)</td>
<td>0.125 (0.040)</td>
<td>13.230 (6.992)</td>
<td>2.319</td>
</tr>
<tr>
<td>Weekday evening peak periods (5 p.m.–7 p.m.) (1 if yes, 0 otherwise) (standard deviation of parameter distribution)</td>
<td>0.328 (0.084)</td>
<td>24.077 (12.951)</td>
<td>6.085</td>
</tr>
<tr>
<td>Defected pavement surface (1 if yes, 0 otherwise) (standard deviation of parameter distribution)</td>
<td>0.225 (0.064)</td>
<td>17.482 (12.981)</td>
<td>4.164</td>
</tr>
<tr>
<td>Number of observations</td>
<td>2499</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log-likelihood at zero LL(0)</td>
<td>−111,205.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log-likelihood at convergence LL(\rho)</td>
<td>−7752.62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \rho^2 ) [1−LL(0)/LL(\rho)]</td>
<td>0.93</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
likely to experience crashes compared with those without work zones. Thus, the results indicate that work zones around intersection locations should have adequate signage and enforcement in order to reduce crash frequency at such locations.

Average annual traffic on major road was found to have a normally distributed parameter with mean $7.681 \times 10^{-4}$ and standard deviation of $8.096 \times 10^{-4}$, signifying that an increase in traffic volume on the major road would positively increase accident frequency for 82.86% of the intersections, and would decrease accident frequency for the remaining 17.14% of intersections. The result reveals that for most of the intersections, a unit increase (in thousands) in major road traffic volume would increase accident frequency by 0.014 (in thousands). From the results, it can be observed that traffic on the major road at an intersection plays a significant role in crashes.

The snowfall indicator variable produced a statistically significant normally distributed positive parameter with mean 0.106 and standard deviation of 0.052, indicating for 97.92% of intersections, an increase in snowfall would increase the probability of experiencing crashes, while reducing the crash frequency for 2.08% of the intersections. A unit increase in snowfall at an intersection would increase crash frequencies by 1.971. This shows that as snowfall increases at an intersection, drivers may find it difficult to stop when approaching an intersection thereby resulting in a crash.

Finally, intersections with defected pavement surfaces produced a normally distributed positive parameter with a mean 0.225 and standard deviation 0.064, suggesting that for 99.98% of intersections, an increase in the number of intersections with defected pavement surfaces would increase accident frequency, while decreasing crash frequency for the remaining 0.02% of intersections. A unit increase in the number of defected surfaces at an intersection would increase crash frequencies by 4.164. The marginal effect value is high and suggests that agencies should ensure defected surfaces at intersections are prioritized in maintenance works, and pavement conditions should be regularly monitored to ensure defected surfaces are timely repaired to prevent crashes.

4. Summary and conclusions

This paper investigates effects of a number of factors on the safety of highway intersections. Random-parameter Poisson and negative binomial models were initially tested, however, the dispersion factor was found to be statistically insignificant. Thus, the random-parameter Poisson model was used to analyze crash frequency data from 2004 to 2010, in order to examine the effects of a number of variables on crashes at intersections. Out of the eight estimated parameters (including the constant term) found to be statistically significant, six produced normally distributed parameters while the remaining were observed to be fixed across intersections. The marginal effects were used to estimate the effects of explanatory variables on crash frequency. The results indicate that out of the identified factors, evening peak period traffic volume, pavement condition, and unlighted intersections have the greatest effects on crash frequencies. In relation to pavement condition, the marginal effect value for defected pavement surface at intersections was high, and the countermeasure for this will be for agencies to ensure defected surfaces at intersections are prioritized in maintenance works so that defected surfaces are timely repaired to prevent crashes. Overall, the results seek to suggest that, in order to ensure effective highway-related safety countermeasures at intersections, significant attention must be focused on ensuring that pavements are adequately maintained and intersections should be well lighted.

It needs to be mentioned that, projects could be implemented at and around the study intersections during the study period (7 years), which could affect the crash frequency over the time. This is an important variable, which could be a part of the future studies to investigate the impacts of safety-related works at intersections and their marginal effects on crash frequency at signalized intersections.

Acknowledgments

The authors are grateful for the assistance of transportation agencies in the Chicago metropolitan area for data collection as part of methodology application.

References


