Do Motor Vehicle Crashes Arise from Single or Multiple Unique Risk Processes? An Inquiry into Crash Causes and Modelling

Amir Pooyan Afghari[a], Simon Washington[a], Md. Mazharul Haque[b] and Tanya Smyth[b]

[a] Civil Engineering and Built Environment School, Queensland University of Technology
[b] Centre for Accident Research and Road Safety - Queensland (CARRS-Q), Queensland University of Technology

Abstract

Crashes at any particular transport network location consist of a chain of events arising from a multitude of potential causes and/or contributing factors whose nature is likely to reflect geometric characteristics of the road, spatial effects of the surrounding environment, and human behavioural factors. It is postulated that these potential contributing factors do not arise from the same underlying risk process, and thus should be explicitly modelled and understood. The state of the practice in road safety network management applies a safety performance function that represents a single risk process to explain crash variability across network sites. This study aims to elucidate the importance of differentiating among various underlying risk processes contributing to the observed crash count at any particular network location. To demonstrate the principle of this theoretical and corresponding methodological approach, the study explores engineering (e.g. segment length, speed limit) and unobserved spatial factors (e.g. climatic factors, presence of schools) as two explicit sources of crash contributing factors. A Bayesian Latent Class (BLC) analysis is used to explore these two sources and to incorporate prior information about their contribution to crash occurrence. The methodology is applied to the state controlled roads in Queensland, Australia and the results are compared with the traditional Negative Binomial (NB) model. A comparison of goodness of fit measures indicates that the model with a double risk process outperforms the single risk process NB model, and thus indicating the need for further research to capture all the three crash generation processes into the SPFs.

Introduction

Efficient management of resources allocated to reduce dramatic costs of vehicular crashes requires an in-depth understanding of crash causation process. It is widely accepted that crashes at any particular location of the transport network are the results of a single chain of events arising from a multitude of potential causes and/or contributing factors (Washington & Haque, 2013). In such a chain, however, different causes may not necessarily originate from the same sources and they may also have varied contributions to crash occurrence. The nature of crash contributing factors is likely to reflect geometric characteristics of the road, spatial effects arising from features of the surrounding environment, and human behavioural factors. These three sources can influence crash occurrence via unique yet interrelated underlying avenues or risk processes. Thus, a primary consequence of postulating a single unique risk process is that the influence of each separated risk process on the final outcome (crash) is not differentiated and may be mistakenly associated to the incorrect sources of crash causal factors. Nevertheless, the state of the practice in road safety network management applies a Safety Performance Function (SPF) that represents the single risk process to explain crash variability across network sites and thus is incapable of linking various portions of total observed crashes caused by separate sources of causal factors (Washington & Haque, 2013).

The first attempts in the literature to understand crash causation process emerged by modelling crash risk based on several independent explanatory variables (Hauer, 1986). Crash Prediction Models (CPM) or Safety Performance Functions (SPFs) were developed to correlate crash contributing factors as explanatory variables with the total observed crash count to identify crash risk (Hauer, 1986, 1992, 1997; Joshua & Garber, 1990; Miaou, Hu, Wright, Rathi, & Davis, 1992; Miaou & Lum, 1993a, 1993b). However, this process was based on the fundamental assumption that a multitude of crash contributing factors operate in a single chain consisting of a series of events which ultimately lead to crash occurrence.

Accordingly, researchers tried to explore the relationship between crashes and variety of roadway geometric characteristics following a single crash generating process (Ardekani, Hauer, & Jamei, 1992; Lyon, Oh, Persaud, Washington, & Bared, 2003; Oh, Lyon, Washington, Persaud, & Bared, 2003; Vogt & Bared, 1998). Although it was recognised very early in the literature that crash contributing factors may originate from different sources such as climate conditions in addition to roadway geometric features (Hauer, 1986, 1997), a separation of such sources was largely ignored in crash risk modelling. Later, it was confirmed that spatial effects arising from features of the surrounding environment contribute to crash occurrence as well (Aguero-Valverde & Jovanis, 2006; Huang & Abdel-Aty, 2010; Mitra & Washington, 2012; Qin & Reyes, 2011; Yasmin & Eluru). However, SPFs still followed a single crash generating process. The evolution of crash risk modelling continued to progress mainly in refining the statistical shortcomings of the models. Such vast advances have been structured around the underlying assumption that crashes arise from a single unique risk process.

This study aims to elucidate the importance of differentiating among various underlying risk processes contributing to observed crash counts at any particular network location. It is postulated that potential crash contributing factors arise from three different underlying processes, including roadway geometric, spatial, and human behavioural factors. To demonstrate the principle of this theoretical and corresponding methodological approach, this study attempts to model two crash generating processes to initiate the multiple unique risk process models. In particular, this study explores engineering and unobserved spatial factors as two explicit sources of crash contributing factors, leaving the human behavioural factors as the next step of this research to further increase model complexities and improves models performances. A Bayesian Latent Class (BLC) analysis is used to investigate these two sources and to incorporate prior information about their contribution to crash occurrence. The methodology is applied to the state controlled roads in Queensland, Australia and the results are compared with the traditional Negative Binomial model.

Methodology

In order to explicitly assess the separated underlying risk processes, it is required to establish two separate SPFs correlating the predicted means of crash counts in each process (μ_1 and μ_2) with two different sets of covariates:

$$\mu_1 = {F_1}^{\alpha_1} e^{(\alpha_3 X_3 + \alpha_4 X_4 + \alpha_5 X_5 + \cdots)} e^{(\varepsilon_{i1})} \quad \text{Equation [1]}$$

$$\mu_2 = {F_1}^{\beta_1} e^{(\beta_3 Z_3 + \beta_4 Z_4 + \beta_5 Z_5 + \cdots)}\ e^{(\varepsilon_{i2})} \quad \text{Equation [2]}$$

where F_1 is the measure of exposure, X_i and Z_i are explanatory variables for each distinct risk process and α_i and β_i are unknown regression parameters. To incorporate randomness into the models, random terms (ε_{i1} and ε_{i2}) are added to SPFs. To account for unobserved heterogeneities, these random terms are allowed to vary across observations by assigning a Multivariate Normal distribution as follows:

$$\boldsymbol{\varepsilon_i} \sim \mathrm{MN}\,(\boldsymbol{\xi}, \Sigma) \quad \text{where} \quad \boldsymbol{\varepsilon_i} = \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \end{bmatrix}, \quad \boldsymbol{\xi} = \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

where $\boldsymbol{\xi}$ is the vector of mean values and ξ_1 and ξ_2 are the mean values for random terms respectively. It should be mentioned that the above specification of multivariate distribution accounts for possible correlations between the two risk processes.

Each of the abovementioned predicated means accounts for a proportion (w_1 and w_2 respectively) of the total predicted mean of crash counts (μ):

$$\mu_1 = w_1 \mu$$

$$\mu_2 = w_2 \mu$$

$$w_1 + w_2 = 1$$

In other words, the total predicted mean of crash counts is a weighted sum of the two aforementioned means:

$$\mu = \frac{1}{2\,w1} \mu_1 + \frac{1}{2\,w2} \mu_2 \quad \text{Equation [3]}$$

where $\frac{1}{2\,w1}$ and $\frac{1}{2\,w2}$ are the predicted weights associated to each distinct risk process. Accordingly, the total observed crash counts (Y) which follows a Poisson distribution with the mean of μ is a weighted sum of two latent (underlying) crash counts (Y_1 and Y_2); each one representing a proportion (W_1 and W_2 respectively) of the total observed crash counts:

$$Y_1 = W_1 Y$$

$$Y_2 = W_2 Y$$

$$W_1 + W_2 = 1$$

Or equivalently:

$$Y = \frac{1}{2\,W1}Y_1 + \frac{1}{2\,W2}Y_2 \qquad \text{Equation [4]}$$

where $\frac{1}{2\,W1}$ and $\frac{1}{2\,W2}$ are the observed weights associated with each distinct risk process. Indeed, the proposed methodology utilises the BLC analysis and prior knowledge in order to determine w_i as an estimate of W_i which illustrates the contribution of each risk process to the total observed crash counts. Having been assigned a distribution with known parameters (known from the prior knowledge), these weights are allowed to vary across observations. The proposed model is calibrated in a Bayesian framework where the posterior is equal to the product of likelihood: $P(Y|\ \mu)$ and prior: $\pi(\mu)$. Markov Chain Monte Carlo (MCMC) simulation is used to estimate the entire unknown parameters including w_i, α_i, β_i, ς_i and σ_{ij} and to make inferences about the posterior.

Finally, the Bayesian Information Criterion (BIC) is used to compare the performance of models:

$$BIC = -2 \text{ Log Likelihood} + p \text{ Log } (n) \qquad \text{Equation [5]}$$

where *p* and *n* are the number of estimated parameters and the number of observations respectively and the model with a lower DIC and BIC values outperforms the other models.

Data

The methodology is applied to the network of state controlled roads in Queensland, Australia consisting of 4,913 roadway segments and approximately 33,510 kilometres in length. Five years of crash data (2010 to 2014) with a total count of 18,484 crashes associated to the network were analysed. All crash severities were included (fatal, hospitalisation, medical treatment, minor injury and Property Damage Only).

Roadway geometrical characteristics were collected from the Queensland Transport and Main Roads Department in GIS formats. The database included segments length, number of lanes, Average Annual Daily Traffic (AADT), percentage of Heavy Vehicle (HV) traffic, Level Of Service (LOS), segments terrain (horizontal and vertical alignment), pavement seal conditions, speed limit, pavement rutting, roughness, longitudinal and alligator cracking conditions. The number of lanes did not vary significantly across observations and thus was not included in the model. The Average Annual Daily Traffic (AADT) of road segments was employed as the exposure variable. LOS is defined as a qualitative measure of traffic service in the road which scales from A to F (Garber & Hoel, 2014) as in the following:

LOS A: Highest quality of service; motor vehicles drive at their desired speed
LOS B: Lower quality of service; the passing demand and passing capacity are almost equal
LOS C: Formation of platoons and platoon size; passing opportunities are severely decreased.
LOS D: Unstable flow and incomplete passing manoeuvres
LOS E: Impossible passing; longer and more frequent platoons; unstable operating conditions
LOS F: Full congestion; demand exceeding capacity.

Moreover, it was found that pavement rutting, roughness, longitudinal and alligator cracking conditions were all highly correlated and so to avoid multicollinearity, only pavement rutting conditions were included in the model. Pavement rutting is defined as permanent deformations of the pavement in the wheel paths. The maximum allowable rutting is 12mm and thus segments rutting was defined as between 0 (the smoothest pavement surface) and 12mm (the roughest surface). Dummy values were assigned to the speed limit, general terrain and pavement seal conditions of road segments to create associated categorical variables. Speed limit was categorised into three groups including low speed limit (Speed Limit<50Km/hr), medium speed limit (50 <Speed Limit<100Km/hr) and high speed limit (Speed Limit>100Km/hr). Terrain condition includes two categories: level and mountainous/rolling. Surface seal condition also includes two categories: sealed and unsealed.

Many studies have emphasised the influence of spatial features of the transport network, such as precipitation, number of rainy days, number of snowy days, presence of college or university within a certain distance of road segments, on crash occurrence (Aguero-Valverde & Jovanis, 2006; Mitra & Washington, 2012). To investigate the effects of such factors, climate data were collected from Australian Bureau of Meteorology for the associated network. Climatic factors included average yearly rainfall (over 5mm), average rainy days per year, average daily solar exposure as well as average sunshine hours (to capture glare effects of sunshine on drivers), average monthly wind speed, and average thunder days per year. To better capture the effects of rain and solar conditions and facilitate the interpretation of rainfall and days of rain as well as solar exposure and sunshine hours, two new variables were established to capture the combined effects of these variables. The former was achieved via dividing rainfall by number of rainy days per year and the latter by dividing solar exposure by number of sunshine hours per day, and these new variables were named as ‘rain conditions’ and ‘solar conditions’, respectively.

To incorporate the effects of adjacent land use patterns, the geographic locations of schools and population centres were collected from the Queensland Spatial Catalogue in GIS formats. As there are many vulnerable road users (pedestrians) in the vicinity of such centres, their proximity to road segments may increase the risk of crash occurrence. Moreover, the intensity of bridges and culverts (number of bridges and culverts per kilometre) were also derived from the geometrical database. Such factors can influence the concentration and cautiousness of drivers which can be interpreted as unobserved spatial effects of the surrounding environment. Eventually, engineering and spatial data were merged using a GIS platform based on spatial coordinates of roadway segments. Table 1 presents descriptive statistics of the study variables.

Table 1. Descriptive Information of study variables

Variable	Minimum	Maximum	Mean	Standard Deviation
Crash	0	150	4	8
Length (Km)	0	63.4	6.8	7.7
AADT (Vehicles/Day)	0	72405	7594	11753
Percent of HV Traffic	1	92	16.7	10.9
Rutting	0	11.2	3.7	1.6
Rainfall (mm)	0	8000	1276.4	976.5
Number of Rainy Days per Year	0	75	36	11
Solar Exposure (MJ/m^2)	0	24	20.8	3.2
Sunshine Hours per Day	0	10	8.3	0.5
Number of Thunder Days per Year	0	80	25	6
Wind Speed (Km/hr)	0	26	11.6	5.4
Intensity of Major Culverts per Kilometre	0	76.2	0.4	3.6
Intensity of Minor Culverts per Kilometre	0	571.4	3.6	21.1
Intensity of Bridges per 10 Kilometres	0	9.5	0.1	0.4
Intensity of Educational Centres per 10 Kilometres	0	16	0	0.35
Proximity to population centres (Km)	0	1456.7	54	146.5
Categorical Variables	**Observation Frequency**		**Sample Share**	
High Speed Limit (>100 Km/hr)	2442		50%	
Medium Speed Limit (>50 and <100 Km/hr)	2386		48%	
Low Speed Limit (<50 Km/hr)	85		2%	
Terrain[1]	866		18%	
Pavement Seal Conditions[2]	4670		95%	
LOS[3]	3370		68%	

[1]0 (if Level), 1(if rolling and/or mountainous)
[2]0 (if un-sealed), 1 (if sealed)
[3]0 (if A, B, C or D), 1(if E or F)

Results and Discussion

Negative Binomial (NB) regression model is the widely accepted safety performance function to establish the relationship between traffic crashes and contributing factors (Poch & Mannering, 1996). Thus, estimating a traditional NB model with a single risk process was the first task in this study. Table 2 presents the results of NB model estimated in Bayesian framework. According to Table 2, the 90% credible intervals for the dispersion parameter (Φ) of NB model does not include zero. This indicates the presence of significant over-dispersion in crash data and thus it is necessary to utilise the NB model to account for such an over-dispersion. Thirteen variables out of all factors used in the study were significant in the NB model with 90% certainty. Some of these variables had positive effects, while others had negative effects on the total crash count. The AADT, length and terrain configuration of road segments had positive coefficients indicating that greater volume of traffic, longer road segments and rolling and/or mountainous terrain results in higher number of crashes. Furthermore, positive coefficients for low and medium speed limits along road segments intuitively indicated that compared with motorways, arterial roads are more associated with traffic crashes. The percentage of heavy vehicles, rain conditions, solar conditions, average number of thunder days per year, wind speed, intensity of bridges and schools had negative coefficients, indicating that these variables have decreasing effects on the total crash count. In adverse weather conditions, drivers may adapt and drive more cautiously, which might have resulted in negative association with total crashes. Pavement seal conditions and LOS also had negative coefficients, indicating that changing from unsealed to sealed and from congested to free flow conditions result in less crashes.

Table 2. Regression results of the Traditional NB model with a single risk process

Variables	Mean	Std. Deviation	Bayesian Credible Interval (BCI) 10% Value	90% Value
Constant	-9.664	0.509	-10.290	-9.122
AADT	0.755	0.026	0.722	0.786
Length	0.665	0.025	0.634	0.694
Percent of HV	-0.030	0.002	-0.032	-0.027
Terrain	0.083	0.039	0.033	0.133
Pavement Seal	-0.289	0.098	-0.430	-0.178
Low Speed Limit	0.845	0.144	0.661	1.031
Medium Speed Limit	0.732	0.035	0.687	0.777
LOS	-0.290	0.043	-0.346	-0.235
Rain Conditions	-0.123	0.095	-0.247	-0.002
Solar Conditions	-0.068	0.034	-0.115	-0.025
Thunder Days	-0.371	0.193	-0.609	-0.107
Wind Speed	-0.234	0.071	-0.326	-0.140
Intensity of Bridges	-2.424	0.667	-3.298	-1.566
Intensity of Schools	-1.027	0.448	-1.603	-0.454
Φ	1.961	0.079	1.860	2.064
Number of Observations (Sample Size)	4913			
Number of Parameters	16			
Log Likelihood	-9145			
Bayesian Information Criteria (BIC)	18426			

The next step was to apply the multiple generating process SPFs on the data. Crash contributing factors were categorised into two sources: engineering factors and spatial factors. Engineering factors included segments length, percentage of heavy vehicles, general terrain, pavement surface conditions, speed limit, LOS and rutting conditions of road segments. Spatial factors included rain conditions, solar conditions, average number of thunder days per year, average annual wind speed, number of major and minor culverts as well as number of bridges along the road segments, intensity of educational centres, and the proximity of road segments to population centres. Since exposure factors play a vital role in crash occurrence models, the exposure variable and its coefficients were set to be the same in both risk processes following the formulation in eq. 6. This implies that there exists a base crash count associated with exposure, irrespective of any other crash contributing factors. As a result, for the multiple risk process model eq. 6 is simplified to the following:

$$\mu = F^{\alpha 1} \times \left(\frac{1}{2\ \omega_1}\ \mu_1 + \frac{1}{2\ \omega_2}\ \mu_2\right) \qquad \text{Equation [6]}$$

where:

$$\mu_1 = e^{(\alpha_3 X_3 + \alpha_4 X_4 + \alpha_5 X_5 + \cdots)} e^{(\varepsilon_{i1})}$$

$$\mu_2 = e^{(\beta_3 Z_3 + \beta_4 Z_4 + \beta_5 Z_5 + \cdots)}\ e^{(\varepsilon_{i2})}$$

F is the exposure variable and the remaining notations are the same as previously stated. According to the literature (Washington & Haque, 2013), unobserved spatial factors account for approximately 5 to 10 percent of all crashes and thus a uniform distribution ranging from 0.05 to 0.15 was used in this study as a prior distribution for the proportion of spatial risk process. Although the MCMC simulation resulted in Markov chains which were stabilised and converged for most of the regression parameters, the ultimate convergence of the model needs to be improved in future efforts. The regression results for the multiple risk process model are presented in Table 3.

A comparison of the two tables shows that the traditional NB regression analysis with a single risk process results in a model with Log likelihood of -9145 while the Log likelihood of the multiple risk process model is -7335, demonstrating 20% improvement compared to the traditional NB model. It should be noted, however, that separating two generating processes leads to an increase in the number of parameters to be estimated. While the number of observations in the three models is 4913, the NB and the multiple risk process models have 16 and 20 parameters to be estimated respectively. According to Tables 2 and 3, the BIC value of the multiple risk process model (14840) is smaller than the BIC value of the traditional NB model (18426) which clearly shows the dominance of the former model in goodness of fit.

The prominent result of this study, however, is that according to Table 3, crash contributing factors originate from two distinct sources associated with two latent risk processes including engineering and spatial factors. The mean proportions of these two sources across observations (w_i) are 90% and 10% respectively. The variance and confidence intervals of such weights show that the contribution of such sources is significant with 90% certainty.

The coefficients of all significant variables excluding the average number of thunder days per year had the same sign in both models. However, the NB model resulted in a negative coefficient for the average number of thunder days while separating the two risk processes caused the coefficient sign to become positive. This result could be considered more intuitive in which increasing the number of thunder days per year results in an increased crash counts at road segments.

Table 3. Regression results of the multiple risk process model

	Mean	Std. Deviation	Bayesian Credible Interval (BCI) 10% Value	90% Value
Exposure Factor				
AADT	0.685	0.021	0.663	0.713
Engineering Factors				
Constant	-9.099	0.270	-9.393	-8.810
Length	0.664	0.010	0.652	0.677
Percent of HV	-0.037	0.002	-0.040	-0.034
Terrain	0.054	0.038	0.005	0.103
Pavement Seal	-0.332	0.043	-0.391	-0.278
Low Speed Limit	0.675	0.148	0.486	0.862
Medium Speed Limit	0.696	0.036	0.651	0.744
LOS	-0.204	0.045	-0.261	-0.148
Spatial Factors				
Constant	-11.590	1.512	-13.780	-9.970
Solar Conditions	-1.540	2.107	-1.809	-0.817
Thunder Days	5.648	3.341	2.261	10.380
Wind Speed	-9.324	4.210	-14.850	-5.127
Random Terms				
ε_1	0.059	0.053	0.007	0.137
ε_2	0.761	0.526	0.142	1.521
σ_{11}	4.182	1.713	2.551	7.015
σ_{22}	0.326	0.127	0.189	0.490
$\sigma_{12} = \sigma_{21}$	-0.767	0.410	-1.383	-0.321
Average Risk Process Weighs				
w_1	0.900	0.000	0.900	0.901
w_2	0.100	0.000	0.099	0.101
Number of Observations (Sample Size)	4913			
Number of Parameters	20			
Log Likelihood	-7335			
Bayesian Information Criteria (BIC)	14840			

A further assessment of the mean values for the regression coefficients indicated that while significant engineering variables had the same increasing/decreasing effect in both models, their coefficients changed very slightly in magnitude from one model to another. However, separating the two risk processes caused a dramatic change in the coefficient magnitudes of solar conditions (from -0.068 to -1.540), average number of thunder days per year (from -0.371 to 5.648) and wind speed (from -0.234 to -9.324). Bearing in mind that this dramatic change occurred in the proposed model where six other spatial variables became insignificant, it can be inferred that separating the two risk processes caused the three previously mentioned variables to absorb the majority of spatial effects. The BCI values for ε_1, ε_2, σ_{11} and σ_{22} indicated that the random terms and their variance were significant with 90% certainty for each distinct risk process. It is noteworthy that the BCI values for σ_{12} and σ_{21} did not include zero, indicating that there is a correlation between the two risk processes.

This finding could be indicative of the fact that the two risk processes were distinct and yet interrelated.

Conclusions

This study aimed to demonstrate the principle of a multiple risk process mechanism for crash causation and its corresponding methodological approach. The objective was achieved by differentiating between two distinct crash generating processes including engineering and unobserved spatial factors. A traditional NB count model was initially estimated for the means of comparison with the new proposed multiple risk process model.

It was concluded that the crash data is over-dispersed and thus the traditional NB model is appropriate to capture the over-dispersion. A comparison of BIC values for the NB and the multiple risk process models clearly showed the dominance of the latter in terms of goodness of fit. Further, a comparison of significant variables indicated that while many spatial factors were not significant when separately modelled, solar conditions, average number of thunder days and wind speed were significant in the spatial risk process. Moreover, significant changes occurred in coefficient magnitudes of such variables when spatial factors are separately modelled. This clearly shows that these three spatial factors play influential roles among other spatial variables in the spatial risk process. Furthermore, the decreasing effect of average number of thunder days per year on total crash counts changed to an increasing effect after separating the two risk processes, consistent with expectations.

In summary, it can be seen that the performance of SPFs in goodness of fit is significantly improved by separating the two distinct processes. Further, the multiple risk process SPF methodology illuminates the true significance and influence of crash contributing factors on crash occurrence. Future research should include all three crash generating processes, including engineering factors, spatial factors, and human behavioural influence, into the SPFs and demonstrate their implications for black spot identifications.

Limitations

The scope of this research was limited to investigating the influence of postulating two risk processes including engineering and spatial factors on the crash occurrence over the state controlled road network in Queensland. Human behavioural data such as gender, age, and possession of driving licence is directly associated with the third risk generating process (behavioural factors) which is out of the current scope of this research. Although a comprehensive assessment of crash causation process should include all three distinct sources of crash contributing factors, i.e. roadway geometric, spatial, and human behavioural factors, this study aimed to demonstrate the principle of postulating multiple risk processes for crash causation and corresponding methodological approach. Future research efforts should expand the proposed model and include all three underlying processes of crash occurrence. Moreover, vehicle characteristics may be considered as another source of crash contributing factors. However, mitigating such factors is beyond the scope of road agencies and thus it was not dealt with in this context.

The focus of the study was on the development of a theoretical model and thus a representative dataset was collected for the state controlled roads in Queensland to validate the proposed model with real-world data. Although the network only consisted of roadway segments (excluding intersections), the distinction between rural and urban roads was not available in collected data. However, geometric characteristics of the segments (e.g. length and AADT) partially accounted for the principal differences between rural and non-rural segments. Future research should apply the proposed model on a more extensive dataset that consists of road segments in urban and rural road environment separately and includes a wide range of roadway geometric and traffic control characteristics.

Furthermore, it should be noted that the analysis is still in its initial phase and requires more complex modelling techniques for better MCMC convergence, different sets of distribution assumption for weights of latent risk processes, and exploration of other possibilities like a random parameter (RP) model. RP models may potentially improve the analysis due to the fact that observations are broadly distributed over the Queensland state and thus assuming all parameters as fixed for the entire population may influence the results.

References

Aguero-Valverde, J., & Jovanis, P. P. (2006). Spatial analysis of fatal and injury crashes in Pennsylvania. *Accident Analysis & Prevention, 38*(3), pp. 618-625.

Ardekani, S., Hauer, E., & Jamei, B. (1992). Traffic impact models. *Chapter 7 in Traffic Flow Theory, Oak Bridge National Laboratory Report*.

Garber, N., & Hoel, L. (2014). *Traffic and highway engineering*: Cengage Learning.

Hauer, E. (1986). On the estimation of the expected number of accidents. *Accident Analysis & Prevention, 18*(1), pp. 1-12.

Hauer, E. (1992). Empirical Bayes approach to the estimation of “unsafety”: the multivariate regression method. *Accident Analysis & Prevention, 24*(5), pp. 457-477.

Hauer, E. (1997). *Observational Before/After Studies in Road Safety. Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety*.

Huang, H., & Abdel-Aty, M. (2010). Multilevel data and Bayesian analysis in traffic safety. *Accident Analysis & Prevention, 42*(6), pp. 1556-1565.

Joshua, S. C., & Garber, N. J. (1990). Estimating truck accident rate and involvements using linear and Poisson regression models. *Transportation planning and Technology, 15*(1), pp. 41-58.

Lyon, C., Oh, J., Persaud, B., Washington, S., & Bared, J. (2003). Empirical investigation of interactive highway safety design model accident prediction algorithm: Rural intersections. *Transportation Research Record: Journal of the Transportation Research Board, 1840*(1), pp. 78-86.

Miaou, S.-P., Hu, P. S., Wright, T., Rathi, A. K., & Davis, S. C. (1992). Relationship between truck accidents and highway geometric design: a Poisson regression approach. *Transportation research record*(1376).

Miaou, S.-P., & Lum, H. (1993a). Modeling vehicle accidents and highway geometric design relationships. *Accident Analysis & Prevention, 25*(6), pp. 689-709.

Miaou, S. P., & Lum, H. (1993). A statistical evaluation of the effects of highway geometric design on truck accident involvements *(No. CONF-930120-2).* Oak Ridge National Lab., TN (United States).

Mitra, S., & Washington, S. (2012). On the significance of omitted variables in intersection crash modeling. *Accident Analysis & Prevention, 49*, pp. 439-448.

Oh, J., Lyon, C., Washington, S., Persaud, B., & Bared, J. (2003). Validation of FHWA crash models for rural intersections: Lessons learned. *Transportation Research Record: Journal of the Transportation Research Board, 1840*(1), pp. 41-49.

Poch, M., & Mannering, F. (1996). Negative binomial analysis of intersection-accident frequencies. *Journal of Transportation Engineering, 122*(2), pp.105-113.

Qin, X., & Reyes, P. E. (2011). Conditional quantile analysis for crash count data. *Journal of transportation engineering, 137*(9), pp. 601-607.

Vogt, A., & Bared, J. (1998). Accident prediction models for two-lane rural roads: segments and intersections. *Federal Highway Administration Report FHWARD-98–133, Washington, DC*.

Washington, Simon & Haque, Md. Mazharul (2013) On the commonly accepted assumptions regarding observed motor vehicle crash counts at transport system locations. In 92nd Annual Meeting of Transportation Research Board (TRB), 13-17 January 2013, Washington DC.

Yasmin, S., & Eluru, N. Latent Segmentation Based Count Models: Analysis of Bicycle Safety in Montreal and Toronto.