International Journal of Modern Research in Engineering and Technology (IJMRET)
www.ijmret.org Volume 1 Issue 1 ǁ June 2016.

An Exploratory Computational Piecewise Approach to Characterizing and Analyzing Traffic Accident Data.

Farhan Saleem, Eric Asa[1] and Joseph Membah.
Associate Professor, North Dakota State University, Department of Construction Management and Engineering, Dept. 2475, Box 6050, Fargo, ND, 58108-6050

Abstract : *In spite of the many advancementsinsafety technology, roadway design and engineering as well as several policy initiativesaimed at addressingtraffic crashes (and it concomitant injuries and fatalities); it continuesto saddle humanity and present significant health hazards and threats to the socio-economic wellbeing of the inhabitants of this earth. Even though federal and state transportation engineers, policy makers, planners and researchers have spent large sums of money and effort on this complex and ubiquitous problem, traffic crashes is one of the top causes of fatal and non-fatal injuries in the United States. In this work a piecewise approach was used to perform an exploratory,systemic characterization and analysis of the six-year (2008-2013) traffic crash data from North Dakota to discover the nature, peculiarities and trends in the data.Several important features and trends in the data were discovered and the outcome of this paper can be further used for engineering design, planning and policy analysis. The research approach could be duplicated in any other state to enhance its societal benefits. Heterogeneity and uncertainty will be fully addressed in future work.*

Key words: *Traffic accidents; Safety; Exploratory analysis; Piecewise.*

Introduction

From an engineering standpoint the most challenging problem in highway safety is the persistence of traffic crashes in spite of several improvements in engineering design and planning, technological advances and policy initiatives to halt this unfortunate phenomenon (Mannering et al. 2016). Traffic accidents are a major source of property damage, injuries and deaths in the United States and across the globe.Traffic fatalities is considered the leading cause of deaths in the first thirty years of life and the second highest leading cause of deaths before age 75 in the U.S (Conner and Smith 2014; CDC, 2010). Although several policies and preventive measures have been implemented in the United States over the years, an average of about 35,000 people lose their lives to traffic accidents every year. According to the National Safety Council38,300 people lost their lives to traffic accidents and another 4.4 million (1.37% of the entire US population)were injured on this country'sroads in 2015. It was the largest one-year percentage increase in traffic fatalities in the United States in half a century.In North Dakota, traffic accidents accounted for 135 fatalities and 5,289 personal injuries in 2014 (NDDOT, 2014). In North Dakota, a traffic accident occurred every 33 minutes; one person was injured from a traffic-related accident every 1.7 hours; and one person died every 2.7 days from traffic accidents in 2014 (NDDOT, 2014). It is quite evident that traffic accidents and the concomitant fatalities and injuries remain a major health problem and present a significant challenge to the socio-economic well-being of the inhabitants of the United States.Generally, ttraffic safety analyses are employed to evaluate potential safety issues and identify opportunities for improving safety. In order to identify the challenges and address the traffic accident problem, several methods have been employed to model and analyze traffic accident data over the years.The predominant approach in traffic crash research is the use of generalized linear models (GLMs) to characterize crash data, develop crash modification factors,analyze the relationship between traffic accidents and different covariates, predict values, identify hot spots and screen variables (Geedipally et al. 2012; Wu, Chen et al. 2014; Shirazi et al. 2016).Other benefits of traffic safety analyses are the identification of low cost/high-impact improvement options;promotion of safety conscious planning, design, and implementation culture;provision of data and information to facilitate good decision making in future; economic benefits from reduced accidents; liability cost reduction due to safer roads and others (Mahmud et al, 2015). The two commonly used methods are negative binomial and Poisson regression models. Flavors of the Poisson-based modelsare Poisson lognormal conditional-autoregressive (Wang and Kockelman 2013), full Bayesian multivariate Poisson lognormal (El-Basyouny et al. 2014), multivariate Poisson-lognormal regression (Ma et al. 2008), Poisson-Weibull (Cheng et al. 2013), Poisson-gamma mixture (Mothafar et al. 2016), geographically-weighted Poisson (Hadaayeghi et al. 2010), diagonal inflated bivariate Poisson regression (Lao et al. 2011), Poisson regression(Powers at al.2010), non-canonical Poisson regression (Polus and Cohen 2012) and other Poisson-based models. Examples of the recent applications of the negative binomial models to traffic accident data are finite mixture of negative binomial (Zou et al. 2014), negative binomial-Lindley (Lord and Geedipally 2011),semiparametric negative binomial generalized linear (Sherazi et al. 2016), finite mixture of negative binomial with varying weight parameter (Zou et al. 2013), negative binomial-generalized exponential (Vangala et al. 2015), random parameters negative binomial panel (Coruhet al. 2015), negative binomial and a generalized estimating equation (Mohammadi et al. 2014), random effect negative binomial (Chin and Quddus2003) and other models. In view of the limitations of the negative binomial and Poisson models, machine learning, Bayesian and other classes of models have recently attracted considerable attention. Examples are quantileregression (Wu, Gao et al. 2014), time series (Quddus 2008), empirical Bayes approach (Gkritza et al 2014), Bayesian networks (de Ona et al. 2013), latent class cluster (de Ona et al. 2013; Cerwick et al. 2014),spatial analysis (Pirdavani et al. 2014), spatial regression (Rhee et al. 2016), mixed logit (Cerwick et al. 2014; Wu et al. 2014), multinomial logit (Ye and Lord 2014), ordered probit (Ye and Lord 2014),Tobit regression (Anastasopoulos 2008), discrete generalized Pareto distribution (Prieto et al. 2014), multinomial logistic regression (Bham et al. 2012), fuzzy logic (Effati et al. 2015), data mining – classification trees and rules discovery (Montella et al. 2011; Pakgohar et al. 2010), clustering (Anderson 2009), support vector machines (Li et al 2012), artificial neural networks (Abdelwahab and Abdel-Aty, 2001) and other models.

Wu et al. (2014)and Effati et al. (2015) argued that fully parametric (as well as mixed and random) models, lack robustness, cannot adequately handle heterogeneity and could result in biased estimation of crash incidence and severity as homogeneity and independence are assumed during modeling. Researchers have noted that traffic crash data could (1) contain many zero values and thus be highly dispersed (Lord and Geedipally 2011; Lord and Geedipally 2012; Shirazi et al. 2016); (2) exhibit high variance (low amount of information) or low variance (high amount of information); (3) be heterogeneous – process and statistical (Rhee et al 2016);(4) complex (Mannering et al. 2016 and Delorme and Lassarre 2014); and (5) uncertain (Rhee et al. 2016).Even though methodological advances like random parameters count models (Barua et al. 2016), random parameters Tobit models (Anastaopoulos et al. 2012), random parameters generalized count model (Bhat et al. 2014), latent-class (finite mixture) model (Yasmin et al. 2014), Markov switching count model (Malyshkina and Mannering 2010), and bivariate/multivariate models with random parameters (Barua et al 2015, 2016) have been developed to address the unobserved heterogeneity; addressing heterogeneity is still a significant challenge as the disadvantages of these approaches outweigh their advantages (Zou et al. 2013; Vangala et al. 2015; Mannering et al. 2016). In general; traffic crash data were collected from different scales (cities, counties, states), roads (urban and rural); road characteristics/geometries (intersections, curves, type of surface and others) and others and are aggregated when used in modeling. The quantitative models were applied to the traffic accident data without prior exploratory analysis to discover the trends; peculiarities; and the effects of the various factors on traffic crashes (Rhee et al 2016). In view of the inherent complexity and uncertainty in traffic crash data, an understanding of the relationships among the various factors and traffic crashes could be a useful source of information for transportation engineers (Rhee et al. 2016; Mannering et al 2016).Traffic accidents are driven by several discrete variables and research to understand the peculiarities of the variables and how they relate to crashes is needed (Quddus 2008; de Ona 2013; Zou et al. 2014). Kim et al. (2006) indicated in their work that there are differences in the relationship between different crash types and the explanatory variables and using crash type model could enhance understanding of traffic crashes and the effectiveness of countermeasures. According to Zou et al. (2014) and Mannering and Bhat (2014), sub-groups of data are homogeneous as compared to aggregates of data from different sites, which are heterogeneous. Segregating the data into different subgroups could improve modeling and analysis. In thisresearch an exploratory piecewise approach was used to characterize the traffic accident data types to discover their peculiarities, nature and trends.

Research Approach

This research is aimed atunderstanding the nature, peculiarities and trends in the traffic accidents data/factors to aid prediction, engineering design/planning and policy analysis. These issues have not been addressed together in the literature in the way it is presented in this work. The computational approach employed in the research is summarized in Figure1.

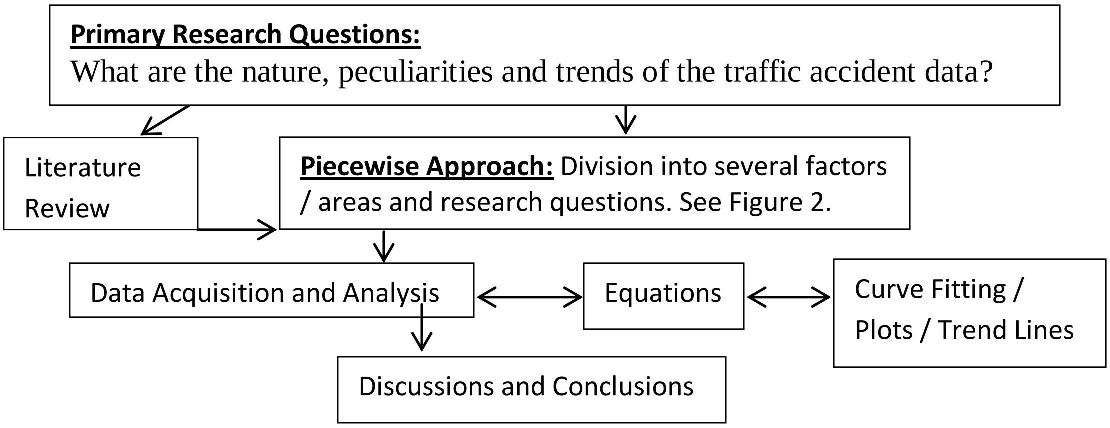


Figure 1: Research Approach

A piecewise approach was used to breakdown the main research questionsinto several research questions thataddresses the various aspects of the data (Figure 2).In view of the complexity, uncertainty, and heterogeneity, fitting a complex response equation to represent the large number of unknown model coefficients is difficult. The relationships among the traffic crashes and the various factors could be linear and/or nonlinear. Thus the high dimensional,multivariate surface (equation 1) could be reduced to a one dimensional space representing each of the factors (equation 2).

$$Y(X) = f_1(x_1) + f_2(x_2) + \cdots + f_{n-1}(x_{n-1}) + f_n(x_n) + e_n \qquad (1)$$

$$y(x_1) = f_1(x_1) + e_1 \ldots\ldots\ldots.for\ subfactor\ 1$$

$$y(x_2) = f_1(x_1) + e_2 \ldots\ldots\ldots.for\ subfactor\ 2$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$y(x_n) = f_n(x_n) + e_n \ldots\ldots\ldots.for\ subfactor\ n \qquad (2)$$

Major research question: What are the nature, peculiarities and trends of the traffic accident data?

- Area of jurisdiction
  - Agencies – what agencies are responsible for reporting accidents?
  - Counties - which counties had the most accidents?
  - Cities - which cities experienced the most accidents?
  - Rural or urban - which areas experienced more accidents – rural or urban?
- Visibility + weather
  - Visibility –which light condition is more critical to traffic accidents?
  - Weather – which weather-related factors drive accidents?
- Road infrastructure
  - Road condition – what are the effects of road conditions on accidents?
  - Type of surface – what type of road surface is more prone to accidents?
  - Surface condition - does the surface condition affect the accident rate?
  - Geometrics – is there any impact of road geometrics on accidents?
  - Intersection type – do more accidents occur at intersections?
  - Functional class – what class of road is more dangerous?
- Crash related factors
  - Crash severity - what are the various levels of severity of traffic crashes?
  - Classes of objects - what classes of objects are associated with accidents?
  - Vehicle type - which vehicles are more likely to be involved in accidents?
  - Access control - how does access control relate to accidents?
  - Manner of collision - are there any similarities in accidents related to the manner of collision?
  - Travel direction – does the direction of travel affect the accident rate?
  - Type of road - what type of road is more critical to traffic accidents?
  - Road structure - which one is more dangerous- roadway, shoulder, or median?
  - Type of junction - which type of junction is associated with more accidents?
- Human related factors
  - Alcohol/drugs – does alcohol and/or drugs play in a role in accidents?
  - Visual obstruction - does visual obstruction contribute to traffic accidents?
  - Evasive action - do drivers take evasive actions when confronted with potential crash situations?
  - Hit and run - are hit and run incidents a major component of accidents?
- Time related factors
  - Year - which year had the most accidents during the study period?
  - Month - what months are more critical for accidents?
  - Day – which days of the month did more accidents occur?
  - Time of day – in which time slots did more traffic accidents occur?
- Other factors
  - Type -what is the relationship between VMT and the types of accidents?
  - Licensed drivers – how does the licensed drivers relate to accidents?
  - Registered vehicles – how does registered vehicles relate to accidents?
  - Population - what is the relationshipbetween population and

Figure 2: Piece-wise Research Approach

Where $x_1, x_2, \ldots, x_n$ are the subfactors and $e_1, e_2, \ldots, e_n$ are the errors. Thus the piecewise approach involves the use of several, one-dimensional linear/nonlinear equations to accurately and efficiently approximate the complex nonlinear surface (Ngueveu et al. 2016; Li and Cao 2008).

This approach fully captures the nature, peculiarities and trends underlying the various factors driving traffic crashes.The research employed six-years (long term in relation to the growth North Dakota's economy) traffic accident data from the state of North Dakota.

Analysis and Discussions

North Dakota'sroad traffic accident data from 2008 to 2013was obtained from the Safety Division of North Dakota Department of Transportation. This large data contain about 107,000 crash reports from 2008 to 2013. The data comprises of different parameters including their description and coding as per NDDOT standards.Figure 3 represents a graphical distribution of the total number of crashes within the study period (2008-2013).

Figure 3: Map of North Dakota Showing the Total Number of Crashes from 2008 to 2013.

Table 1 is a summary of the population, licensed drivers, registered vehicles and the total number of crashesthat occurred in the state of North Dakota from 2008 to 2013.

Table 1: Summary of data on population, drivers, vehicles and thetotal number of crashes

Year	Population	Change %	Licensed Drivers	Change %	Registered Vehicles	Change %	Crashes	Change %
2008	641,421		475,129		774,346		16387	
2009	646,844	0.845	479,921	1.009	728,376	-5.937	17673	7.848
2010	672,591	3.980	487,489	1.577	726,937	-0.198	17076	-3.378
2011	683,932	1.686	496,543	1.857	889,213	22.323	18823	10.231
2012	699,628	2.295	509,195	2.548	844,617	-5.015	18356	-2.481
2013	723,393	3.397	520,083	2.138	804,332	-4.770	18977	3.383

A positive trend was observed in the state's population from 2008 to 2013. The population of North Dakota increased each year during the study period. From 2008-09, the population increased by about 0.85% but from 2009-10 the increment was 3.98%;which is the highest for the study period. In 2010-11 and 2011-2012 the percent increase in population was 1.68% and 2.29%;respectively (Table 2). The increase in population from 2012 to 2013 was 3.39%.

There was a steady positive increase in licensed drivers during the study period. From 2008-09, the percent change in the number of licensed drivers in North Dakota was about 1%, then the numbers increased to 1.57%, 1.85%, 2.54% for the periods of 2009-2010, 2011-2012 and 2012-2013, respectively. The highest percent change (2.54%) in the number of licensed drivers was from 2011 to 2012. In 2012-2013 the percent change slightly dropped from 2.54% to 2.13%.

The number of registered vehicles in North Dakota varied for the entire study period (2008-13). The percent change in registered vehicles was - 5.93% from 2008 to 2009 as the number of registered vehicles dropped by about 45,970 vehicles from 2008 to 2009. For 2009-10 the percent change in registered vehicles was only -0.19% as the numbers of vehicles were more or less the same for the 2009-10 and the difference was of only 1,439 registered vehicles. There was a big jump in the number of registered vehicles between 2010 and 2011;as a total of 162,276 vehicles were registered in that period. The percent change in the number of registered vehicles was (22.32%) from 2010 to 2011. The number of registered vehicles declined afterwards. The change in the number of registered vehicles was -5.01% and -4.76% for year 2011-2012 and 2012-2013, respectively.The number of accidents in North Dakota has generally increased from year to year (Table 2). There was about 7.84% increase in the number of crashes from 2008-2009; as 1,286 more crashes were reported in 2009 compared to 2008. The percent change in crashes from 2009-2010 was - 3.37% due to a decrease in the number of crashes (about 597 reported in that period). The numbers of crashes increased quite rapidly from 2010 to 2011, as 1,747 more crashes occurred in the one year period (representing an increase of 10.23%). This rapid change could be the result of a large increase in the number of vehicles (162,276) registered in that period. After that the numbers of crashes decreasedin 2012 as 467 less crashes were reported as compared to 2011 and the percent change in crashes was - 2.48% from the previous year. The number of crashes increased in 2013 by 621 from the previous year, despite a decrease of 40,285 in the number of registered vehicles. The percentage changes in population, the number of licensed drivers, the number of registered vehicles and the number of crashes during the study period (2008-2013) is also shown in Table 2.

There is a positive trend in the number of crashes from year to year (2008-2013), even though there were slight increases in 2010 and 2012. A curvilinear equation (3) was fitted to the data.

$$y = 10.935x^3 - 66006x^2 + 1E{+}08x - 9E{+}10 \quad (3)$$

The R^2 is 0.7843, and so the model could be used to predict the approximate number of crashes with respect to year.Linear equations(4) and (5) were fitted to the data of population and licensed drivers of North Dakota.The results show a strong positive trend in the case of population and licensed drivers with R^2values of 0.979 and 0.978 respectively. So,these equations could be used to infer meaningful results from the data.

$$y = 16559x + 620013 \quad (4)$$

$$y = 9189.9x + 462562 \quad (5)$$

A curvilinear equation (6) was fitted to the data of registered vehicles. It showed that the number of registered vehicles went up and down during different years of the study period (2008-2013). R^2 is 0.7686 which could be used to predict the approximate number of registeredvehicles.

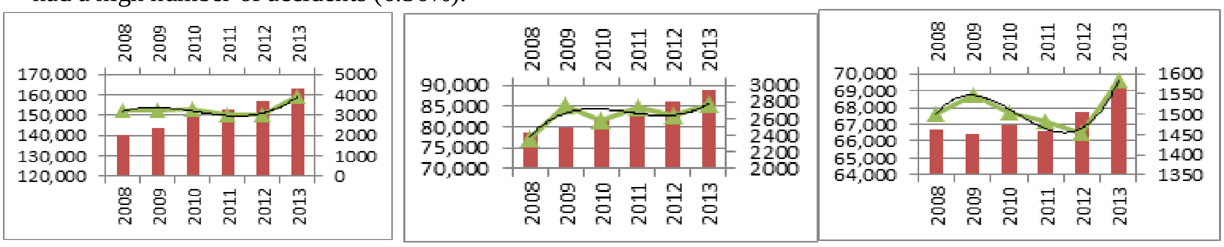
$$y = -12156x^3 + 125064x^2 - 348440x + 1E{+}06 \quad (6)$$

The following sections addresses the questions raised in Table 1 and discusses the detailed analysis of road traffic accident data grouped into different categories to ascertain the intricacies of the traffic accident data.

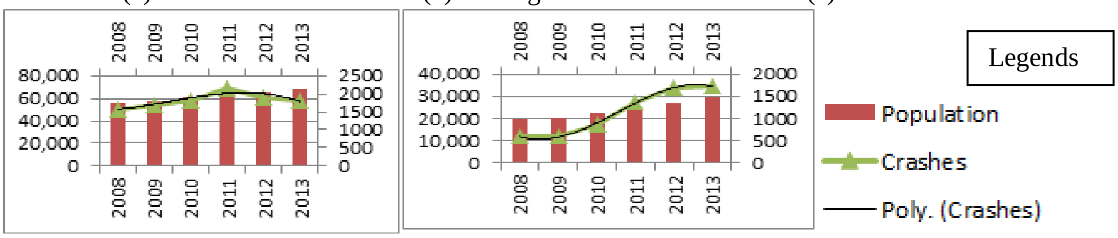
Area and Jurisdiction Factors

Agencies: The statistical analysis shows that city police departments reported the most accidents (54.45%), followed by county police (25.35%) and highway patrol (19.43%). The rest of the traffic accidents were reported by campus police and BIA. Military and park rangers reported the lowest number of accidents. Thus a greater proportion of accidents occurred within the limits of cities.

Counties: Figure 4 is the comparison of crashes versus population in the five major counties of North Dakota. It is evident that graphs of crashes against population do not followthe same trend in the five major counties (Figure 4). A total of 58.28% of traffic accidents occurred in the top five counties with respect to the number of accidents are Cass, Burleigh, Ward, Grand Forks and Williams. Cass County is obviously on top due to its high population (the highest in the state), as more people may translate into more traffic accidents. Williams County on the other hand, which has less population (29,563); almost half than the Grand Forks population (69,311); had a high number of accidents (6.36%).



(a) Cass (b) Burleigh (c) Grand Forks



(d) Ward (e) Williams

Figure 4: North Dakota Counties (Crashes Vs Population 2008-13)

Cass County:Figure 4(a) shows that the number of crashes in Cass County increased each year from 2008 to 2010 but not in considerable amount. In 2011 the number of crashes decreased substantially by 284. There was a steep rise (922) in number of crashes from 2012 to 2013. A curvilinear equation (7) was fitted to the crash data.

$y = 59.667x^3 - 549.64x^2 + 1427.3x + 2250$ (7)

R^2 is 0.9415 which is strong enough to infer reliable results for the number of crashes in Cass County.

Burleigh County:The results from Figure 4(b) depicted that a mixed trend as the number crashes went up and down throughout the study period (2008-2013). A curvilinear equation (8) was fitted to the data of the number of crashes.

$y = 21.537x^3 - 243.32x^2 + 855.43x + 1756.7$ (8)

R^2 is 0.7311, which is quite good and could be used to predict the number of crashes in Burleigh County.

Grand Forks County:The results from Figure 4(c) show aslow and steady decline in the number of crashes from 2008 to 2012; except 2009. In 2013 the number of crashes jumped considerably by 133.A curvilinear equation (9) was fitted to the number of crashes.

$y = 10.843x^3 - 105.08x^2 + 285.79x + 1308.7$ (9)

R^2 is 0.949 and it could be used to infer reliable results for number of crashes in Grand Forks County.

Ward County:In Ward County the number of crashes rose gradually from 2008 to 2011 (Figure 4d). The number of crashes declined noticeably in the last two consecutive years(2012 and 2013). A curvilinear equation (10) was fitted to data.

$y = -13.75x^3 + 98.464x^2 - 54.643x + 1540$ (10)

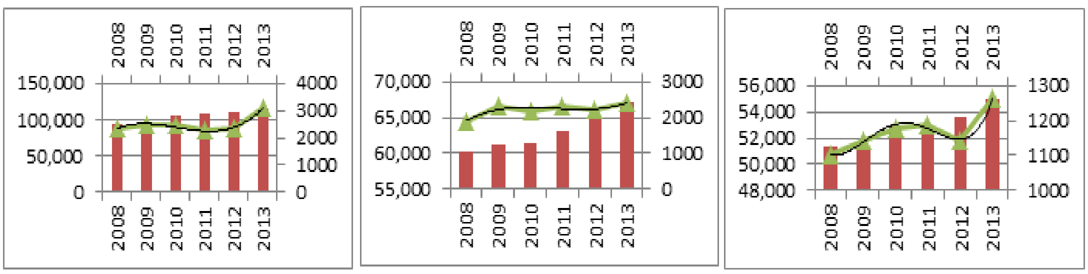
R^2 is 0.8106, which is quite reasonable and the model could be used to predict the number of crashes in Ward County.

Williams County:Figure 4(e) shows that there is a positive trend and the number of crashes increased continuously over the study period (2008-2013). A curvilinear equation (11) was fitted to the data of the number of crashes.

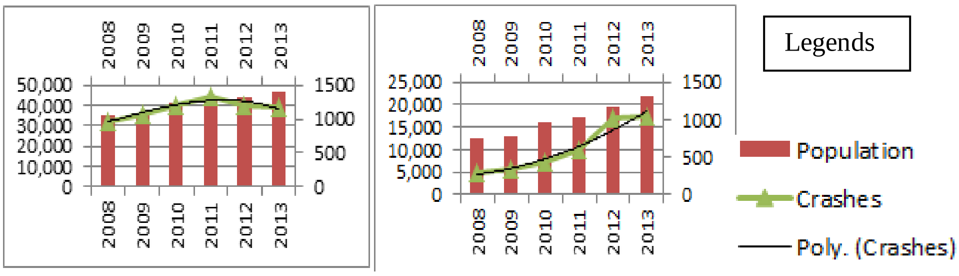
$y = -36.574x^3 + 388.35x^2 - 922.36x + 1166.7$ (11)

R^2 is 0.9991, so the model could be used confidently to forecast the number of crashes in Williams County. The Williams County experienced the highest increase in the number of crashes over the study period.

Cities:The following (Figure 5) is the comparison of crashes against population in the five major cities in North Dakota.Fargo and Bismarck are the two top cities with respect to the number of crashes reported; almost 44% of the accidents occurred in these two big cities in the state. Bismarck is the worst city with respect to the number of accidents compared to the population. Its accident rate (total count with respect to population) is 0.198 as compared to Fargo which has an accident rate of 0.13. After these two cities, about 21.75% accidents occurred in Grand Forks and Minot. West Fargo (2.94%) is the safest among the big cities in North Dakota. In order to investigate the trends in the number of crashes from year to year (2008-2013) for all the big cities of North Dakota, curvilinear equations (12), (13), (14), (15), and (16) were fitted to their crash data.



(a) Fargo (b) Bismarck (c) Grand Forks



(d) Minot (e) Williston

Figure 5: North Dakota Cities (Crashes Vs Population 2008-13)

Fargo: Figure 5(a) shows that the number of crashes in Fargo went up and down from 2008 to 2012. There was aconsiderable jump in number of crashes (about 759)from 2012 to 2013. A curvilinear equation (12) was fitted to data of number of crashes.

$y = 47.843x^3 - 436.29x^2 + 1145.1x + 1581.7$ (12)

R^2 is 0.977 which is strong enough to infer reliable results for the number of crashes in Fargo.

Bismarck:The results from Figure 5(b) showa mixed trend as the number crashes varied throughout the study period (2008-2013). A curvilinear equation (13) was fitted to the crash data.

$y = 22.954x^3 - 257.96x^2 + 916.94x + 1230.7$ (13)

R^2 is 0.8149, which is quite good and the model could be used to predict the number of crashes in Bismarck.

Grand Forks: Figure 5(c) depicts the steady increase in the number of crashes from 2008 to 2011. In 2012 the number of crashes reduced to 1140 and then it increased in 2013; reaching 1262.A curvilinear equation (14) was fitted to data.

$y = 5.1042x^4 - 64.282x^3 + 266.55x^2 - 392.35x + 1289.2$ (14)

R^2 is 0.9762, which is good enough to infer reliable results for number of crashes in Grand Forks.

Minot: Like Grand Forks, there was a gradual rise in the number of crashes from 2008 to 2011 (Figure 5d). The number of crashes declined considerably in 2012 and 2013. A curvilinear equation (15) was fitted to the crash data.

$y = -4.787x^3 + 18.675x^2 + 111.32x + 835.33$ (15)

R^2 is 0.8514, which is quite good and the model could be used to predict the number of crashes in Minot.

Williston: The results from Figure 5(e) show that there is a positive trend and the number of crashes increased throughout the study period (2008-2013). A curvilinear equation (16) was fitted to the data of the number of crashes.

$y = 21.946x^2 + 15.861x + 229.3$ (16)

R^2 is 0.9375, so the model could be used to forecast the number of crashes in Williston.

Table 5 summarizes the nature of accidents in percentages of all the major cities of North Dakota. Minot is on the top with respect to fatal accidents (0.216%), whilst Fargo and Williston were the top cities with respect to accidents involving injuries (26.276%) and property damage only (86.481%). Detailed analysis of each sub category (fatal, injury, PDO) can be viewed in Table 2 for all the major cities of North Dakota.

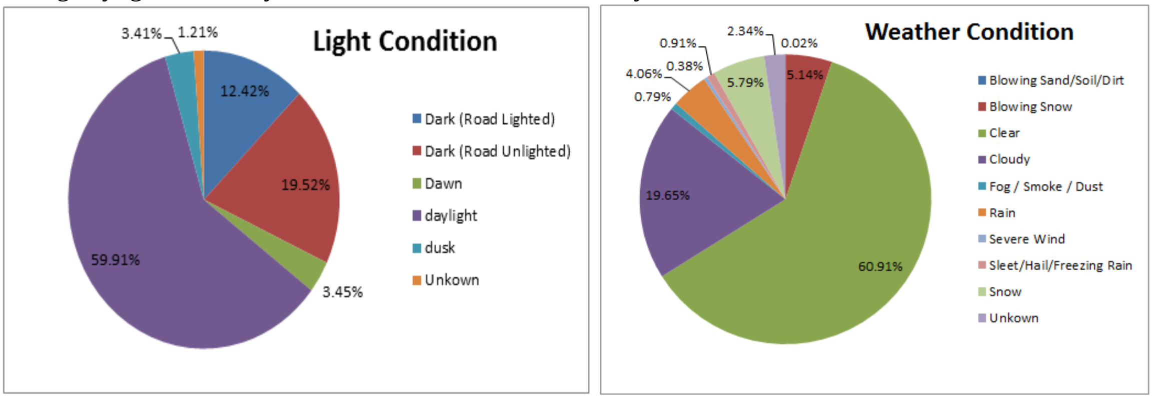
Table 2: Nature of accidents in big cities of North Dakota (%)

City	Nature of Accidents (%)		
	Fatal	**Injury**	**PDO**
Bismarck	0.075	20.906	79.019
Fargo	0.100	26.276	73.624
Grand Forks	0.100	24.704	75.196
Minot	0.216	16.157	83.626
Williston	0.135	13.383	86.481

Rural and Urban: An analysis of the data show that most of the crashes occurred on rural (44%) and on urban minor arterial (41%) roads. Urban principal arterial (15%) is the safest road.

Visibility and Weather Related Factors

Visibility and Weather Conditions:The results from Figure 6 (a) show that, almost 60% accidents happened during daylight. This may be due to lack of concentration by drivers.



(a) (b)

Figure 6: Effects of Light and Weather Conditions on Crashes

About 32% of accidents can be attributed to darkness (road is lighted or unlighted). About 7% of accidents happened at dawn and dusk.The pie chart Figure 6(b) shows that most accidents occurred on clear days (when there is no weather event). Clouds, snow, blowing snow and rainare other weather conditions that contributed to accidents. However all the other factors (aside from clear day) account for less than 40% of traffic incidents in North Dakota.

Road Infrastructure related Factors:

Table 3 summarizes the road infrastructure related factors involved which had effects on the number of crashes with respect to top five contributing sub factors and their percentages.

Table 3: Road infrastructure related factors and sub factors

Road Infrastructure related Factors											
Road Condition		***Surface Type***		***Surface Condition***		***Road Geometrics***		***Intersection Type***		***Functional class***	
Sub Factors	**(%)**	**Sub factors**	**(%)**	**Sub factors**	**(%)**	**Sub factors**	**(%)**	**Sub factors**	**(%)**	**Sub factors**	**(%)**
Normal	95.5	Asphalt	68.2	Dry	55.1	Straight (Level)	79.0	Non-intersection	73.6	Urban Minor Arterial	24.5
Under Construction	1.47	Concrete	24.5	Ice / Compacted Snow	27.3	Straight (Grade)	11.7	Four Way Intersection	18.0	Urban Principal Arterial (Other)	19.7
Loose Material Surface	1.23	Gravel/ Scoria	6.27	Wet	8.27	Curve (Level)	4.61	T Intersection	7.45	Rural Principal arterial (Other)	12.2
Holes/Ruts / Bumps/ Washout	0.61	Dirt	0.66	Snow	5.91	Curve (Grade)	3.49	Five + Roads	0.51	Rural Local Road or Street	8.26
Debris on Road	0.36	Concrete Bridge Deck	0.20	Slush	1.36	Hillcrest	0.89	Y Intersection	0.26	Urban Collector	7.10

Road Condition:Table3shows that 95.5% of accidents that occurred in North Dakota in the study period happened under normal road conditions; only a few other conditions like loose material surfaces (1.233%)and construction areas (1.472%)contributed to more accidentsthan the other factors.

Surface Type:Table 3 depicts that almost 98.21% of accidents occurred on good surfaces like asphalt and concrete that again shows the driver - related behavior may be the main contributor to the high traffic accident rates.Table 4 shows the crash severity with respect to major surface types. It depicts that 71.6% fatal accidents occurred on asphalt surfaces. Whereas, 16.9% and 10.5% fatal accidents occurred on gravel/scoria and concrete surfaces respectively.In terms of accidents involving injuries and PDOs, asphalt surfaces are on top, while concrete and gravel/scoria surfaces are in second and third positions respectively.

Table 4: Crash severity with surface type

Surface Type	**Crash Severity (%)**		
	Fatal	**Injury**	**PDO**
Asphalt	71.608	65.0748	69.2416
Concrete	10.490	25.4219	24.3977
Gravel/Scoria	16.923	8.4148	5.5008
Other	0.979	1.0885	0.8601

Surface Condition:Aside from types of road surface, surface conditions are also important. The results from Table 3 show that almost 55% of crashes were reported when the surface was dry and good for driving, while ice/ compacted snow, snow and wet surface type contributed to
about 41.75 % of the accidents.

Road Geometrics:In terms of road geometrics,Table 3 depicts that79% accidents occurred on straight and level sections of the road, which is quite unusual as compared to straight on grade (11.7%), curve on grade (4.6%) and curve on level (3.5%).

Intersection Type:Intersections are an important aspect ofthe road structure and they could present challenges to drivers. Table 3reveals that about 18% of accidents occurred at four way intersections and 7.4% happened at T intersections. The majority of accidents (73.6%)occurred at sections of the road where there were no intersections.

Functional Class:The results from Table 3 show that the top three dangerous functional classes of road with respect to accidents were urban minor arterial (24.5%), urban principal arterial (19.8%) and rural principal arterial (12.3%).

Crash Related Factors

Table 5 summarizes the crash related factors with respect to the top five contributing sub factors and their percentages during the study period (2008-2013).

Table 5: Crash related factors and sub factors

Crash related Factors											
Crash Severity		***Classes of objects***		***Vehicle Type***		***Access Control***		***Manner of Collision***		***Original Direction of Travel***	
Sub Factors	**(%)**	**Sub factors**	**(%)**	**Sub factors**	**(%)**	**Sub factors**	**(%)**	**Sub factors**	**(%)**	**Sub fact ors**	**(%)**
PDO	79.9	MV in Transport	57.2	Passeng er Car	44.0	No Control	91.1	Non-Coll. w/Motor Veh.	31.2	W	24.0
Non-incapacitat ing injury	9.70	Deer	16.0	Pickup - Van - Utility	43.9	Full Control	8.28	Rear End	23.9	E	23.6
Possible Injury	8.09	Parked Motor Vehicle	5.67	Truck Tractor	3.28	Partial Access Control	0.61	Angle (Not Specific)	21.1	N	23.2

Incapacitating Injury	1.61	Overturn / Rollover	4.92	Motorcycle	1.12	--	--	Sideswipe (same dir.)	7.69	S	23.2
Fatal	0.66	Ran Off Roadway	3.03	3+ Axle	1.05	--	--	Right Angle	7.47	SE	1.81

(N = North; S = South; E = East; W = West; SE = Southeast)

Crash Severity:From the data provided in Table 5property damage only accidents were about 80% of the total crashes in the study period. The most critical accidents are fatal and non-fatal injuries. Fatal accidents were 715 about (0.66%), whilst incapacitating injuries and non- incapacitating injuries were 1.6% and 9.7% respectively.
Classes of objects:In all the crashes reported the top five classes of objects were motor vehicle in transport (57.2%), deer (16%), parked motor vehicle (5.6%), overturn/ rollover (4.9%) and ran off roadway (3%) accidents (Table 5).
Vehicle Type (Unit Configuration):Table 5shows that passenger cars (44%) and pickup-van–utility vehicle (43.9%) were the top two types of vehiclesthat were involved in most crashes in North Dakota. Truck tractors and motorcycleswere also involved in accidents but in much lesser numbersthan the abovementioned, about 3.3% and 1.1% respectively.
Access Control:Table 5 shows that only 8.3% accidents are related to full control (only ramp entry and exit) while the rest 91.1% are related to no control (unlimited access).
Manner of Collision:Non collision with motor vehicle in transport was the top manner of collision in most of the cases (Table 5). The second highest is rear end collision and the third is angular collisions (but there were no specifics).
Original Direction of Travel:The results (Table 5)show that accidents occurred in all the directions and the distributionis almost symmetrical in North Dakota.
Type of road:Most accidents occurred on undivided roads (72%) and about 18% happened on divided highway (Figure 7).

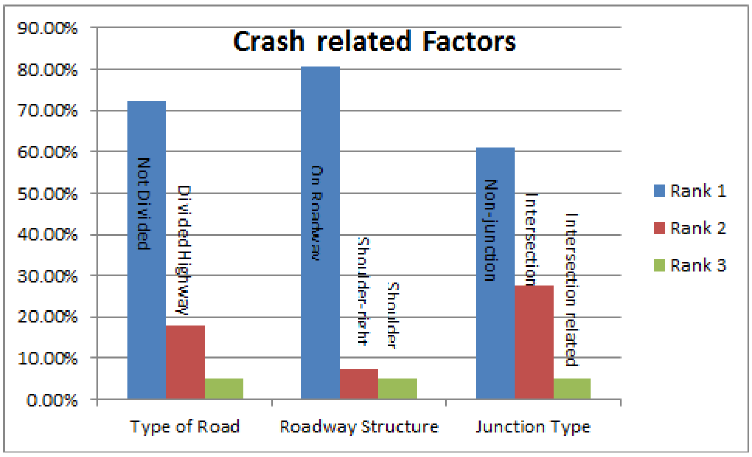
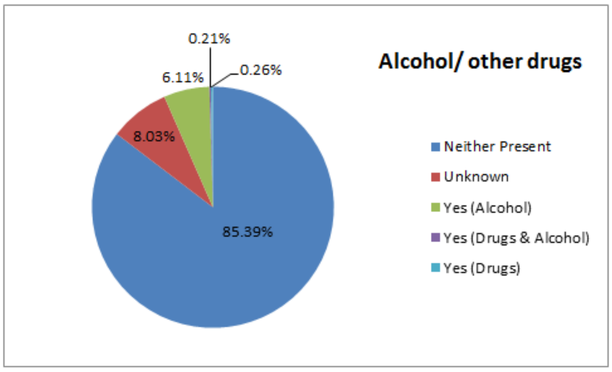


Figure 7: Crash related sub factors with ranking

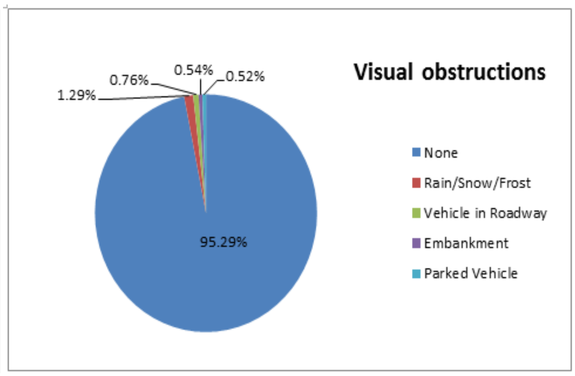
Roadway structure: The results from Figure 7 show that more accidents happened on the main roadway (80.4%) as compared to shoulders (15.95%) and median (2.5%).
Type of junction: About 27.6% of crashes occurred at intersections whereas 60.8% of accidents occurred where there were no junctions or intersections and so the vehicle was travelling in a curve or straight section of road (Figure 7).

Human related Factors

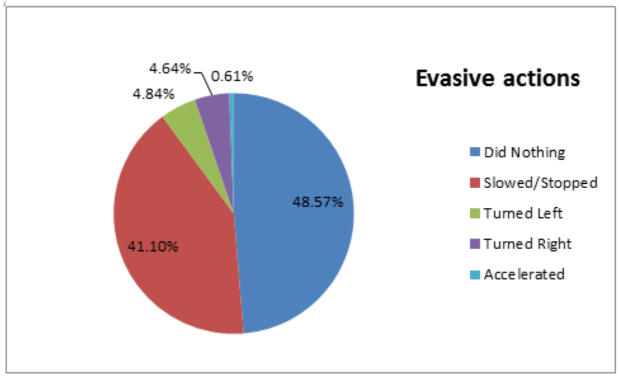
Alcohol/ Other Drug Involvement:Figure 8(a) shows that only 6.57% accidents were reported as drivers under the influence of alcohol/ drugs or both. In 85.4% of the cases neither was present.



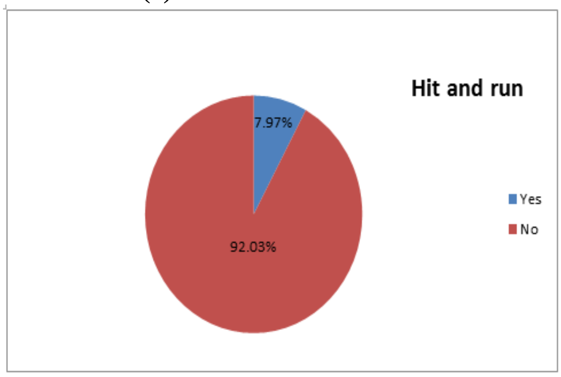
(a)



(b)



(c)



(d)

Figure 8: Effects of Human related Factors on Crashes

Visual Obstruction: In 95.3 % of crashes there were no visual obstructions when the accident happened whereas,rain/snow/frost and fog/smoke/dustcontributed to 1.28% and 0.38% respectively, Figure 8(b).

Evasive Action: A total of 48.5% of drivers did nothing while entering into a potential crash and 41.1% slowed or stopped the vehicle, while about 9.4% turned left or right, Figure 8(c). Thus the majority of drivers did not take any evasive actions when confronted with a potential crash.The accidents might have taken them by surprise.

Hit and Run: Figure 8(d) shows that in almost 92% accidents the parties involved did not leave the scene of the accident whilstin 8% of the cases, drivers left the accident scene.

Factors Related to Year, Month, Day and Time

Crashes by Year:The numbers of accidents increased in each year except 2010 and 2012 when there were slight decreases in percentages from previous years.

Crashes by Month:November, December, January and February are the most critical months during the entire period of 2008-2013. December is the deadliest among all the months.

Crashes by Day of Month:The 2nd and 11th day of the month were the most critical, while the rest of the days of a typical month had close to the same probability of accidents occurring.

Crashes by Time of Day:The most risky time slots were 3:00 pm to 4:00 pm and 5:00 pm to 6:00 pm in the evening and 7:00 am to 8:00 am in the morning.

Other Factors

Crash Severity - Rate per 100 Million Vehicle Miles Traveled (2008-2013):The following section discusses the detailed analysis of total crashes and different types of crashes namely fatalities, injuries and property damage only (PDO), with respect to 100 million vehicle miles traveled in North Dakota. Equations were fitted to see the positive or negative trends for the different types of crashes with respect to 100 million vehicle miles traveled in North Dakota.

Fatal Crashes:Figure 9 (a) shows the rate per 100 million VMT for fatal crashes in North Dakota during the study period (2008-2013). The rate per 100 million VMT for fatal accidents increased from 2008 to 2013. The rate per 100 million VMT for fatal accidents was 1.46 for year 2009 which was the highest within the study period (2008-2013). The rate decreased from 1.46 to 1.1 in 2010. Due to an increase in the number of fatalities in years 2011 and 2012, the rate per 100 million VMT for fatal accidents increased to 1.41 in 2011 and 1.45 in 2012. It dropped from 1.45 to 1.3 in 2013. A curvilinear equation was fitted to the rate per 100 million VMT for fatal crashes (Eqn. 17).

$y = -0.0231x^4 + 0.3135x^3 - 1.4415x^2 + 2.5864x - 0.1476$ (17)

The R^2 is 0.572, which is relatively low to deduce meaningful results for the rate per 100 million VMTfor fatal crashes.

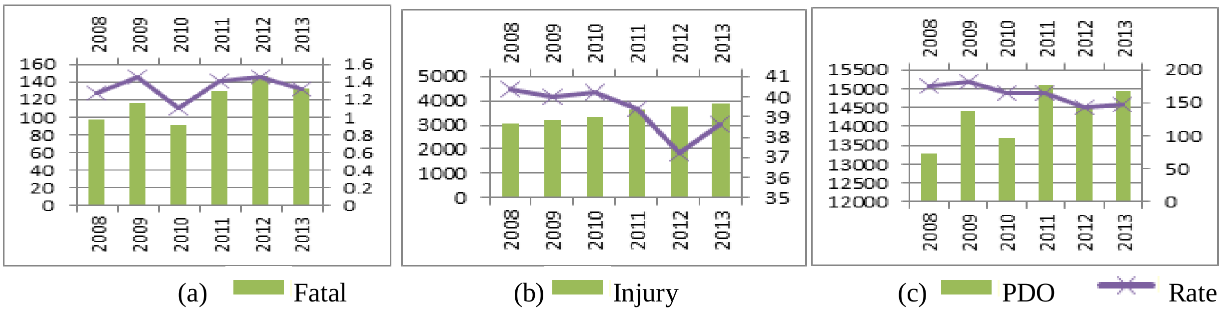


Figure 9: Crash Severity - Rate per 100 Million Vehicle Miles Traveled (2008-2013)

***Injury Crashes**:*Figure 9 (b) shows the rate per 100 million VMTfor crashes involving injuries in North Dakota during the study period (2008-2013). The rate per 100 million VMTfor injury crashes ranges from 37.2 to 40.3 and generally decreased from 2008 to 2013. The highest rate (40.3) was recorded in 2008, whereas the lowest rate was recorded in 2012. A curvilinear equation was fitted to the rate per 100 million VMT for injury crashes (Eqn. 18).

$y = 0.1288x^3 - 1.3663x^2 + 3.6677x + 37.724$ (18).

The results show a negative trend in the case of rate of injury crashes per 100 million VMT. R^2 is 0.768 that could be used to estimate injury crashes.

***Property Damage Only (PDO)**:*Figure 9(c) shows the rate per 100 million VMTfor crashes involving property damage in North Dakota during the study period (2008-2013). The rate per 100 million VMTfor crashes involving property damage decreased significantly between 2008 and 2013. The rate consistently dropped each year from 174.5 to 143.4 from 2008 to 2012. In 2013 the rate again increased from 143.4 to 147.9. A curvilinear equation was fitted to the rate per 100 million VMT for property damage only crashes (Eqn. 19).

$y = 1.2331x^3 - 13.5x^2 + 35.909x + 151.26$ (19)

The R^2 is 0.9041, so it could be used to predict the rate per 100 million VMT for property damage only crashes. The results also show a negative trend in the case of PDO crashes.

Total Crashes: Despite the fact that the VMT in North Dakota increased each year in the study period, the rate of total crashes per 100 million VMT decreased from 2008 to 2013. The only exception is 2009, where the rate per 100 million VMT for total crashes increased from the previous year (from 215.4 to 222.5). The rate per 100 million VMT for years 2010 and 2011 were almost the same. In 2012 the rate per 100 million VMT was dropped to 181.8 and it increased in 2013 and reached 187.9. An equation was fitted to the rate per 100 million VMT for total crashes (Eqn. 20).

$y = 1.3726x^3 - 14.982x^2 + 40.079x + 189.2$ (20)

The results also show the negative trend in case of total rashes. The R^2 is 0. 0.8991and the equation could be used to infer meaningful results.

Crash Severity - Rate per 1,000 Licensed Drivers (2008-2013):The following section discusses the detailed analysis of the total crashes as well as the different types of crashes - fatalities, injuries and property damage only (PDO) with respect to 1000 licensed drivers in North Dakota. Equations were fitted to the data.

Fatal Crashes:Figure 10 (a) shows the rate per 1000 licensed drivers involved in fatal crashes in North Dakota during the study period (2008-2013). The rate per 1000 licensed drivers for fatal accidents increased from 2008 to 2013. The maximum rate per 1000 licensed drivers was 0.28 in 2012. The maximum number of fatal accidents in 2012 was 144. The lowest rate per 1000 licensed drivers was 0.18 in year 2010 which was the safest year of the study period. A curvilinear equation was fitted to the rate per 1000 licensed drivers for fatal crashes (Eqn. 21).

$$y = -0.0074x^3 + 0.0783x^2 - 0.2196x + 0.3554 \quad (21)$$

The results show a positive trend in the case of fatal crashes. The R² is 0.9715, which is strong enough to deduce meaningful results for the rate of fatal crashes per 1000 licensed drivers.***Injury Crashes:***Figure10 (b) shows that the rate per 1000 licensed drivers for injury crashes in North Dakota during the study period (2008-2013). The rate per 1000 licensed drivers for injury crashes steadily increased from 2008 to 2013 and the range varied from 6.4 to 7.5. It shows that about 16% increase in the rate of injuries per 1000 licensed drivers from 2008 to 2013.

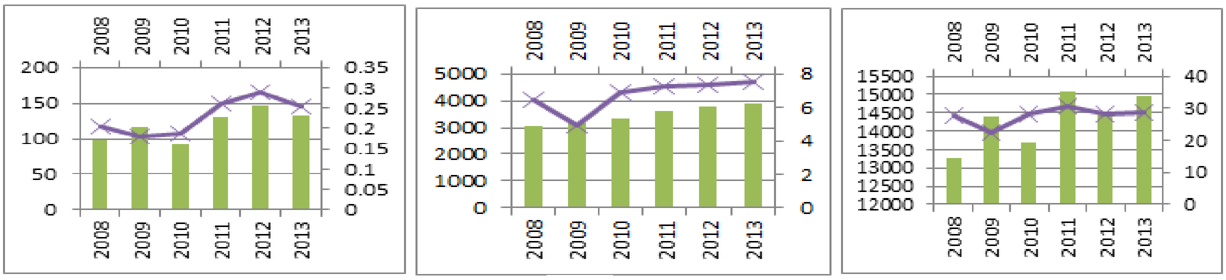


Figure 10: Crash Severity - Rate per 1,000 Licensed Drivers (2008-2013)

A curvilinear equation (22) was fitted to the rate per 1000 licensed drivers for injury crashes.

$$y = -0.1275x^3 + 1.3573x^2 - 3.7998x + 8.8152 \quad (22)$$

The results also show the positive trend in the case of rate per 1000 licensed drivers for injury crashes.R² is 0.738 and the equation could be reasonably used to estimate the rate of injury crashes per 1,000 licensed drivers. ***Property Damage Only:***Figure 10 (c) shows property damage crashes per 1000 licensed drivers for North Dakota during the study period (2008-2013). The rate per 1000 licensed drivers involved in property damage accidents went up and down between 2008 and 2013. The rate was 27.9 in 2008, and then it went down to 22.2 in 2009 and went up to 28 in 2010. The maximum rate was 30.4 which was recorded in 2011.A curvilinear equation was fitted to the rate per 1000 licensed drivers for property damage only crashes (Eqn. 23).

$$y = 0.4512x^4 - 6.7656x^3 + 34.79x^2 - 69.305x + 68.747 \quad (23)$$

The R² is 0.995, so it could be used to predict the rate per 1000 licensed drivers for property damage only crashes. The results show that there is a positive trend in the case of PDO crashes.

Total Crashes:As the number of licensed drivers in North Dakota increased during the study period, the rate per 1000 licensed drivers for the total crashes also increased. In 2011 the rate per 1000 licensed drivers for the total crashes was 37.9 which was the highest recorded for the study period and the lowest rate was 34.9 (2008). After that the rate per 1000 licensed drivers for total crashes decreased from 37.9 to 36 in 2012 and then it went up to 36.4 in 2013. A curvilinear equation was fitted to the rate per 1000 licensed drivers for total crashes (Eqn. 24).

$$y = -0.5798x^3 + 6.0835x^2 - 17.234x + 45.214 \quad (24)$$

The results show the positive trend in case of total rashes.R² is 0.6242, not good enough to infer meaningful results.

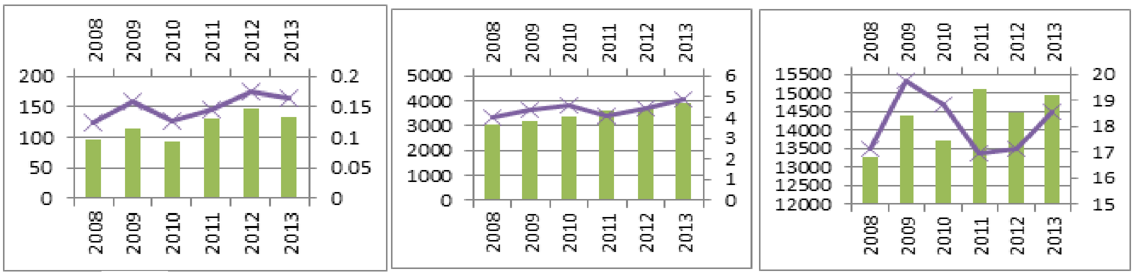
<u>Crash Severity - Rate per 1,000 Registered Vehicles (2008-2013)</u>: The following section discusses the detailed analysis of total crashes and the different types of crashes - fatalities, injuries and property damage only (PDO) with respect to 1000 registered vehicles in North Dakota. Equations were fitted to the data to study the trends in the different types of crashes with respect to 1000 registered vehicles in North Dakota.

Fatal Crashes:Figure 11 (a) shows the rate per 1000 registered vehicles for fatal crashes in North Dakota during the study period (2008-2013). The rate per 1000 registered vehicles for fatal accidents varied from 2008 to 2013 and ranges from 0.12 to 0.17. The lowest value was recorded in 2008 and the highest value was recorded in 2012. The rate per 1000 registered vehicles dropped from 0.17 in 2012 to 0.16 in 2013 as the numbers of registered vehicles decreased by 40,285. A curvilinear equation was fitted to the rate per 1000 registered vehicles for fatal crashes (Eqn. 25).The results show the positive trend in the case of fatal crashes

$$y = -0.0034x^4 + 0.0479x^3 - 0.2289x^2 + 0.4324x - 0.1222 \quad (25).$$

The R² is 0.9494, and it can be used to infer meaningful results for the rate per 1000 registered vehicles for fatal crashes.

Injury Crashes:Figure 11 (b) shows the rate per 1000 registered vehicles for crashes involving injuries in North Dakota during the study period (2008-2013). The rate per 1000 registered vehicles for injury crashes ranges from 3.9 to 4.8 and generally increased during the study period (2008-2013). The rate per 1000 registered vehicles increased from 3.9 to 4.6 from 2008 to 2010 and then it dropped to 4.0 in 2011. The rate went up to 4.4 in 2012 and then again increased to



(a) Fatal (b) Injury (c) PDO Rate

Figure 11: Crash Severity - Rate per 1,000 Registered Vehicles (2008-2013)

4.8 in 2013. A curvilinear equation was fitted to the rate per 1000 registered vehicles for injury crashes (Eqn. 26).

$$y = 0.055x^3 - 0.5658x^2 + 1.7807x + 2.6877 \quad (26)$$

R^2 is 0.8192, which is good enough to infer results from the model. The results also show a positive trend in the case of rate per 1000 registered vehicles for injury crashes.

Property Damage Only:Figure 11 (c) shows the rate per 1000 registered vehicles for property damage only crashes in North Dakota during the study period (2008-2013). The rate per 1000 registered vehicles for property damage crashes was not consistent in the study period (2008-2013). The maximum rate per 1000 registered vehicles of 19.7 was recorded in 2009 and the minimum of 17.1 was recorded in 2012. A curvilinear equation was fitted to the rate per 1000 registered vehicles for property damage only crashes (Eqn. 27).

$$y = 0.3057x^3 - 3.2377x^2 + 9.8118x + 10.37 \quad (27)$$

The R^2 is 0.9307, so it could be used to predict the rate per 1000 registered vehicles for property damage only crashes. The results show that there is mixed trend in the case of PDO crashes as the rate varied during different years of the study period (2008-2013).

Total Crashes: Even though there was an increase in the number of vehicles registered in North Dakota each year, the rate of total crashes per 1000 registered vehicles did not increase consistently from 2008 to 2013. A maximum rate of 24.2 was recorded in 2009, which was surprisingly quite high as the highest numbers of vehicles (889,213) were registered in 2011. The rate per 1000 registered vehicles for years 2008, 2011 and 2012 were almost same. In 2013 the rate per 1000 registered vehicles increased to 23.6. A curvilinear equation was fitted to the rate per 1000 registered vehicles for total crashes (Eqn. 28).

$$y = 0.3626x^3 - 3.8223x^2 + 11.664x + 13.065 \quad (28)$$

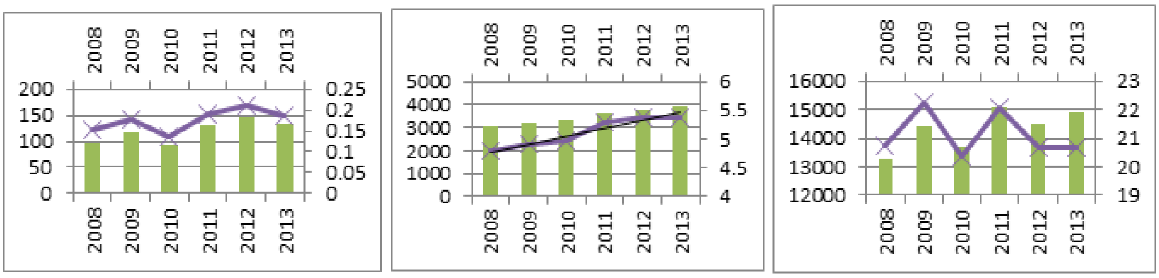
The results show that there is a mixed trend in the case of total crashes.R^2 is 0.9139, so it could be used to predict the rate per 1000 registered vehicles for total crashes.

<u>Crash Severity - Rate per 1,000 Population (2008-2013):</u>This section discusses the detailed analysis of the total crashes and the different types of crashes - fatalities, injuries and property damage only (PDO) with respect to 1000 population in North Dakota. Equations were fitted to the data to study the trend.

Fatal Crashes:Figure12 (a) shows the rate per 1000 population for fatal crashes in North Dakota during the study period (2008-2013). The rate per 1000 population for fatal accidents increased from 2008 to 2013; but not consistently. The maximum rate per 1000 population was 0.21 in 2012, where the maximum number of fatal accidents was 144. The lowest rate per 1000 population was 0.13 in year 2010. A curvilinear equation was fitted to the rate per 1000 population for fatal crashes (Eqn. 29).

$$y = -0.0037x^4 + 0.0499x^3 - 0.2241x^2 + 0.3974x - 0.0666 \quad (29)$$

The results show the positive trend in fatal crashes. The R^2 is 0.8123, which could be used to infer meaningful results for the rate per 1000 population for fatal crashes.



(a) Fatal (b) Injury (c) PDO Rate

Figure 12: Crash Severity - Rate per 1,000 Population (2008-2013)

Injury Crashes:Figure 12 (b) shows that the rate per 1000 population for crashes involving injuries in North Dakota during the study period (2008-2013). The rate per 1000 population for injury crashes increased from 4.7 to 5.4 in 2008 to 2013. It shows that about 12.6% rate per 1000 population for injury crashes increased from 2008 to 2013. Linear equation was fitted to the rate per 1000 population for injury crashes (Eqn. 30).

$$y = 0.1346x + 4.6467 \quad (30)$$

The R^2 is 0.9348, which could be used confidently to infer meaningful results. The results show asteady positive trend in case of rate per 1000 population for injury crashes.

Property Damage Only:Figure 12 (c) shows property damage crashes per 1000 population for North Dakota during the study period (2008-2013). The rate per 1000 population involved in property damage accidents varied from 2008 to 2013. It did not exhibit consistency and went up and down each year. The maximum rate was 22.2 which was recorded in 2009. The rate was 20.3 in 2010, and then it went up to 22.0 in 2011 and went down to 20.6 for the years 2012 and 2013. A curvilinear equation was fitted to the rate per 1000 population for property damage only crashes (Eqn. 31).

$$y = -0.0538x^4 + 0.7897x^3 - 4.0771x^2 + 8.4726x + 15.668 \quad (31)$$

The R^2 is 0.2651, so it could not be used to predict the rate per 1000 registered vehicles for property damage only crashes. The results show that there is a mixedtrend in the case of PDO crashes.

Total Crashes:The population in North Dakota increased each year during the study period. However, the number of total crashes per 1000 population did not increase substantially from 2008 to 2013. In 2011 the total number of crashes per 1000 population was 27.5; which was the highest recorded for the study period and the lowest rate was 25.3 (2010). After 2011, the rate per 1000 population for total crashes decreased from 27.5 to 26.2 in 2012 and remained the same in 2013. A curvilinear equation was fitted to the rate per 1000 population for total crashes (Eqn. 32).

$$y = 0.023x^3 - 0.3544x^2 + 1.583x + 24.515 \quad (32)$$

The results show that there is a varied trend in case of rate per 1000 population for total rashes.
R^2 is also quite low (0.1488) to infer meaningful results.

Conclusions

Traffic crashes and its associated fatalities, injuries and property damages remain a long-standing problem in road transportation; despite several advances in technology, highway design and engineering and safety policy initiatives.Various flavors of generalized linear models - negative binomial and Poisson are the predominant approaches used to characterize traffic crash data. Of late machine learning and other models have been used. Generally these models were used without prior exploratory analysis of the traffic data; even though traffic accidents are driven by several discrete variables and understanding them could throw more light on research results. In view of its pure vastness, diversity, complexity, heterogeneity, uncertainty, and other characteristics of traffic crash data,aggregating the data for use during modeling could lead to wrong conclusions. Research is needed to ascertain the nature, peculiarities and trends in the various factors underlying traffic accidents. In this work a piecewise approach was used to break down the research question into several sub-questions (organized around seven major factors) to enhance the understanding of the nature, peculiarities and trends in the data.

It was evident that each county, city, and other sub-areas have their own trends and peculiarities. There was a positive trend statewide in the number of crashes from year to year even though the number of registered vehicles went up and down during the period of the study (2008-2013). City Police departments documented most of the crashes (54.45% compared with 25.35% by the counties), and William's County (in the oil producing Bakken region of North Dakota) experienced the highest increase in the number of crashes. Even though the most accidents occurred in the two largest cities – Fargo and Bismarck- Minot experienced the highest number of traffic fatalities; whilst Fargo and Williston saw the most accidents involving injuries and property damage only, respectively. The irony is thata high proportion of the accidents (60%) occurred during the day, when the weather was clear. Almost all the accidents (95.5%) occurred under normal road conditions and asphalt paved roads saw the highest number of accidents (68.2%). Asphalt roads accounted for the most accidents - fatalities (72%), injuries (65%) and property damages only (69%) and is the most dangerous road surface to travel in the state. The highest number of accidents occurred on straight and level sections of the roadway (79%); where there were no intersections (73.6%). Property damages only accidents account for the highest percentage of accidents in the state (80%) and almost all the accidents occurred on sections of the main roadway (80%) where there is unlimited (no control) access (91%). Undivided roads are the most dangerous in the state and accounted for 72% of accidents. A high percentage of the accidents did not involve drunk driving (85.4%) and even though there were no visual obstructions (95 %); a greater proportion of the drivers did not take any evasive actions (48.6%). The months of November, December, January and February saw the most accidents and the most risky days were 2nd and 11th days of the month. It is evident that the approach used in this research brought a lot of salient information about the accident data to light. Even though this research was limited to North Dakota, it could be extended to other states to enhance it societal benefits. Further work is needed to fully address the uncertainty and heterogeneity associated with the traffic accident data.

Acknowledgements

This research was fully performed by the Computational and Sustainable Infrastructure Research Laboratory and was not funded by any public or private entity. The data used in the research was provided by the Safety Office of North Department of Transportation; for which we are very grateful.

References

[1.] Abdel-Aty, M. A., and Abdelwahab, H. T. (2004). "Predicting Injury Severity Levels in Traffic Crashes: A Modeling Comparison." Journal of Transportation Engineering, 130(2), 204-201.

[2.] Abdelwahab, H.T., Abdel-Aty, M.A. (2001). "Development of Artificial Neural Network Models to Predict Driver Injury Severity in Traffic Accidents at Signalized Intersections." Transportation Research Record, 1746, 6–13.

[3.] Anastasopoulos, P., Mannering, F., Shankar, V., and Haddock, J. (2012). "A Study of Factors Affecting Highway Accident Rates Using the Random-parameters Tobit Model." Accident Analysis and Prevention, 45 (1), 628–633.

[4.] Anderson, T. K. (2009). "Kernel Density Estimation and K-Means Clustering to Profile Road Accident Hotspots." Accident Analysis and Prevention, 41, 359-364.

[5.] Barua,S., El-Basyouny, K., and Islam, M. (2015). "Effects of Spatial Correlation in Random Parameters Collision Count-data Models." Analytic Methods in Accident Research, 5–6, 28–42.

[6.] Barua,S., El-Basyouny, K., and Islam, M. (2016). "Multivariate Random Parameters Collision Count-data Models with Spatial Heterogeneity." Analytic Methods in Accident Research, 9, 1-15.

[7.] Bham, G. H., Bhanu, S. J., and Manepalli, U. R. R. (2012). " Multinomial Logistic Regression of Single Vehicle and Multi Vehicle Collisions on Urban U. S. Highways in Arkansas." Journal of Transportation Engineering, 138(6), 786-797.

[8.] Bhat, C. R., Born, K., Sidharthan, R., and Bhat, P. C. (2014). "A Count Data Model with Endogenous Covariates: Formulation and Application to Roadway Crash Frequency at Intersections." Analytic Methods in Accident Research, 1, 53–71.

[9.] CDC, Centers for Disease Control and Prevention. (2010),WISQARS (Web-based Injury Statistics Query and Reporting System), US Department of Health and Human Services, CDC, Atlanta, GA (http://www.cdc.gov/injury/wisqars. Accessed on August 23rd, 2013)

[10.] Cerwick, D. M., Gkritkza, Shaheed, M. S. and Hans, Z. (2014). "A Comparison of the Mixed Logit and Latent Class Methods for Crash Severity Analysis." Analytic Methods in Accident Research, 3-4, 11-27.

[11.] Chin, H. C. and Quddus, M. A. (2003). "Applying the Random Effect Negative Binomial model to Examine Traffic Accident

Occurrence at Signalized Intersections.” Accident Analysis and Prevention, 35, 253-259.
[12.] Conner, K. A., and Smith, G. A. (2014). “The impact of aggressive driving-related injuries in Ohio, 2004–2009.” Journal of Safety Research, 51, 23-31.
[13.] Coruh, E., Bilgic, A., and Tortum, A. (2015). “Accident Analysis with Aggregated Data: The Random Parameters Negative Binomial Panel Count Data Model.” Analytic Methods in Accident Research, 7, 37-49.
[14.] Delorme, R. and Lassarre, S. (2014). “A new theory of complexity for safety research. The case of the long-lasting gap in road safety outcomes between France and Great Britain.” Safety Science, 70, 488-503.
[15.] De Ona, J., Lopez, G., Mujalli, R., and Calvo, F.J. (2013). “Analysis of Traffic Accidents on Rural Highways Using Latent Class Clustering and Bayesian Networks.” Accident Analysis and Prevention, 51, 1-10.
[16.] Effati, M., Rajabi, M. A., Hakimpour F., and Shabani, S. (2015). “Prediction of Crash Severity on Two-Lane, Two-Way Roads Based on Fuzzy Classification and Regression Tree Using Geospatial Analysis.” Journal of Computing Civil Engineering, 29(6), 04014099.
[17.] El-Basyouny, K., Burua, S., and Islam, T. M. (2014). “Investigation of Time and Weather Effects on Crash Types Using Full Bayesian Multivariate Poisson Lognormal Models.” Accident Analysis and Prevention, 73, 91-99.
[18.] Geedipally, S. R., Lord, D., and Dhavala, S. S. (2012). “The Negative Binomial-Lindley Generalized Linear Model: Characteristics and Applications Using Crash Data.” Accident Analysis and Prevention, 45, 258-265.
[19.] Gkritza, K., Souleyrette, R. R., Baird, M. J., and Danielson, B. (2014). “Empirical Bayes Approach for Estimating Urban Deer-Vehicle Crashes Using Police and Maintenance Records.” Journal of Transportation Engineering, 140(2), 04013002.
[20.] Hadayeghi, A., Shalaby, A. S., Persaud, B. N. (2010). “Development of Planning Level Transportation Safety Tools Using Geographically Weighted Poisson Regression.” Accident Analysis and Prevention, 42, 91-99.
[21.] Kim, D-G., Washington, S and Oh, J. (2006), Journal of Transportation Engineering, 132(4), 282-292.
[22.] Lao, Y., Wu Y-J, Corey, J. and Wang, Y. (2011). “Modeling Animal-Vehicle Collisions Diagonal Inflated Bivariate Poisson Regression.” Accident Analysis and Prevention, 43, 220-227.
[23.] Li, Z., Liu, P., Wang, W., and Xu, C. (2012). “Using Support Vector Machine Models for Crash Injury Severity Analysis.” Accident Analysis and Prevention, 45, 478-486.
[24.] Li, X., and Cao, Y. (2008). “Projection-Based Piecewise-Linear Response Surface Modeling for Strongly Nonlinear VLSI Performance Variations.” 9th International Symposium on Quality Electronic Design, IEEE Computer Society, 108-113, March 17th – 19th, San Jose, CA.
[25.] Lord, D. and Geedipally, S. R. (2011). “The Negative Binomial-Lindley Distribution as a Tool for Analyzing Crash Data Characterized by Large Amounts of Zeros.” Accident Analysis and Prevention, 43, 1738-1742.
[26.] Ma, J., Kockelman, K. M. and Damien, P. (2008). “A Multivariate Poisson-lognormal Regression Model for Prediction of Crash Counts by Severity, Using Bayesian Methods.” Accident Analysis and Prevention, 40, 964-975.
[27.] Mahmud, S. M. S., Ferreira, L., and Tavassoli, A. (2015). “Traditional Approaches to Traffic Safety Evaluation (TSE): Application Challenges and Future Directions.” Bridging the East and the West, 11th Asia Pacific Transportation Development Conference and 29th ICTPA Annual Conference, ASCE, May 27-29, 242-262.
[28.] Malyshkina, N., and Mannering, F. (2010). “Zero-state Markov Switching Count-data Models: An Empirical Assessment.” Accident Analysis and Prevention, 42 (1), 122–130.
[29.] Mannering, F. L. and Bhat, C. R. (2014). “Analytic Methods in Accident Research: Methodological Frontier and Future Directions.” Analytic Methods in Accident Research, 1, 1-22.
[30.] Mannering, F. L., Shankar, V., and Bhat, C. R. (2016). “Unobserved Heterogeneity and the Statistical Analysis of Highway Accident Data.” Analytic Methods in Accident Research, 11, 1-16.
[31.] Mohammadi, M. A., Samaranayake, V. A. and Bham, G. H. (2014). “Crash Frequency modeling Using Negative Binomial Models: An Application of Generalized Estimating Equation to Longitudinal Data.” Analytic Methods in Accident Research, 2, 52-69.
[32.] Motella, A., Aria, M., D’Ambrosio, A., and Mauriello, F. (2011). “Data Mining Techniques for Exploratory Analysis of Pedestrian Crashes.” Transportation Research Record. 2237, 107-116.
[33.] Mothafer, G. I. M. A., Yamamoto, T, and Shankar V. N. (2016). “Evaluating Crash Type Covariances and Roadway Geometric Marginal Effects Using the Multivariate Mixture Model.” Analytic Methods in Accident Research, 9, 16-26.
[34.] NDDOT, North Dakota Department of Transportation (2014). “North Dakota Crash Summary 2014.” Report Prepared by NDDOT.
[35.] Ngueveu, S. U., Artigues, C., and Lopez, P. (2016). “Scheduling Under a Non-reversible

Energy Source: An Application of Piecewise Linear Bounding of Non-linear Demand/Cost Functions." Discrete Applied Mathematics, 208, 98-113.

[36.] Pakgohar, A., Tabrizi, R.S., Khalilli, M., Esmaeili, A. (2010). "The role of human factor in Incidence and Severity of Road Crashes Based on the CART and LR regression: a Data Mining Approach."Procedia Computer Science 3, 764–769.

[37.] Polus, A., and Cohen, M. (2012). "A Non-Canonical Poisson Regression Model for the Prediction of Crashes on Low-Volume Rural Roads." IATSS Research, 35, 98-103.

[38.] Powers, S., Gerlach, R., and Stamey, J. (2010). "Bayesian Variable Selection for Poisson Regression with Underreported Responses." Computational Statistics and Data Analysis, 54, 3289-3299.

[39.] Pirdavani A., Bellemans T., Brijs, T. and Wets, G. 2014. "Application of Geographically Weighted Regression Technique in Spatial Analysis of Fatal and Injury Crashes." Journal of Transportation Engineering, 140(8), 04014032.

[40.] Prieto, F., Gomez-Deniz, E., and Sarabia, J. M. (2014). "Modeling Road Accident Blackspots Data with the Discrete Generalized Pareto Distribution." Accident Analysis and Prevention, 71, 38-49.

[41.] Quddus, M. (2008). "Time Series Count Data Models: An Empirical Application to Traffic Accidents, 40, 1732-7241.

[42.] Rhee, K-A., Kim, J-K., Lee, Y-i., and Ulfarsson, G. F. (2016). "Spatial Regression Analysis of Traffic Crashes in Seoul." Accident Analysis and Prevention, 91, 190-199.

[43.] Shirazi, M., Lord, D., Dhavala, S. S., and Geedipally, S. R. (2016). "A Semiparametric Negative Binomial Generalized Linear Model for Modeling Over-Dispersed Count Data with a Heavy Tail: Characteristics and Applications to Crash Data." Accident Analysis and Prevention, 91, 10-18.

[44.] Vangala, P., Lord, D., and Geedipally, S. R. (2015). "Exploring the Application of the Negative Binomial-Generalized Exponential Model for Analyzing Traffic Crash Data with Excess Zeros." Analytic Methods in Accident Research, 7, 29-36.

[45.] Wang, Y. and Kockelman, K. M. (2013). "A Poisson-lognormal Conditional-autoregressive Model for Multivariate Spatial Analysis of Pedestrian Crash Counts Across Neighborhoods." Accident Analysis and Prevention, 60, 71-84.

[46.] Wu, Q., Chen, F., Zhang, G., Liu, X. C., Wang, H. and Bogus, S. M. (2014). "Mixed Logit Model-Based Driver Injury Severity Investigations in Single- and Multi-Vehicle Crashes on Rural Two-Lane Highways." Accident Analysis and Prevention, 72, 105-115.

[47.] Wu, H., Gao, L., and Zhang, Z. (2014). "Analysis of Crash Data Using Quantile Regression for Counts." Journal of Transportation Engineering, 140(4), 04013025.

[48.] Yasmin, S., Eluru, N., Bhat, C., and Tay, R. (2014). "A Latent Segmentation Based Generalized Ordered Logit Model to Examine Factors Influencing Driver Injury Severity." Analytic Methods in Accident Research, 1, 23–38.

[49.] Ye, F. and Lord, D. (2014). "Comparing Three Commonly Used Crash Severity Models on Sample Size Requirements: Multinomial, Logit, Ordered Probit and Mixed Logit Models." Analytic Methods in Accident Research, 1, 72-85.

[50.] Zou, Y., Zhang, Y., and Lord, D. (2013). "Application of Finite Mixture of Negative Binomial Regression Models with Varying Weight Parameters for Vehicle Crash Data Analysis." Accident Analysis and Prevention, 50, 1042-1051.

[51.] Zou, Y., Zhang, Y., and Lord, D. (2014). "Analyzing Different Functional Forms of the Varying Weight Parameter for Finite Mixture of Negative Binomial Regression Models." Analytic Methods in Accident Research, 1, 39-52.