EFFECTS OF GEOMETRIC AND TRAFFIC FACTORS ON FREQUENCIES OF TRUCK-INVOLVED CRASHES ON ONTARIO HIGHWAYS

Ran Ran and Chris Lee
Department of Civil and Environmental Engineering
University of Windsor

Introduction

A vehicle is the most important transportation mode in the modern world. However, the safety of road traffic becomes a major concern. There were 2,077 fatalities and 165,172 injuries caused by vehicle crashes in Canada in 2012 (Transport Canada, 2015). In particular, as the demand of freight in surface transportation system increases, the number of truck-involved crashes will also increase. In general, a large number of truck-involved crashes lead to fatality and injury. According to the U.S. National Highway Traffic Safety Administration (NHTSA) (2014), large trucks (gross vehicle weight rating greater than 10,000 pounds) were involved in the traffic crashes which led to 3,921 fatalities and 104,000 injuries in 2012 – i.e., 18% and 4% increase compared to 2011, respectively (NHTSA, 2014). Thus, it is important to analyze truck-involved crashes and identify their unique characteristics.

For this task, we need to understand the influence of road geometry, traffic, environment, and driver behavior on crashes. Generally, higher traffic volume, higher traffic density, higher post speed limit, more number of lanes and bad weather conditions will lead to higher crash frequency. For instance, driver's exceeding speed limits or driving too fast is more likely to cause crashes. Nearly 55 percent of speed-related crashes were attributed to exceeding speed limits

(NHTSA, 2014). Kotikalapudi and Dissanayake (2013) also observed that a driver of large trucks is 1.56 times more likely to have a higher injury severity than the drivers of the other vehicle types if the driver is speeding. Choi et al. (2014) reported that the effects of speed-related variables on injury severity of truck-involved crashes are more statistically significant than the effects of volume-related variables.

Road geometric characteristics are also closely associated with crash frequency. For instance, more number of lanes on the road generally increases the chance of crashes. This is because more number of lanes increases opportunity of changing lanes and the number of conflicts among vehicles (Caliendo et al., 2013). Lane width is also a significant factor affecting crash frequency. Dong et al. (2004) observed that the number of car–truck crashes was higher at intersections with wider lanes of both minor and major roads. Width of shoulders also affected crash frequency. Haleem et al. (2013) demonstrated that wider outside shoulder can reduce the number of total crashes. They also found that segments with 9 feet or more outside shoulder had lower probability of fatal and injury crashes.

Traffic volume was also found to have significant effects on crash frequency in previous studies. Caliendo et al. (2013) observed that the relationship between average annual daily traffic volume (AADT) and crash is not linear. In free-flow conditions, the number of crashes increased with AADT. However, in congested conditions, the number of crashes decreased with AADT. Furthermore, Dong et al. (2004) found that as truck percentage in total traffic volume increases, the opportunity of a collision involving with at least one truck would also increase. On the other hand, Kotikalapudi and Dissanayake (2013) observed that angle crashes on the major roads tend to increase as truck percentage increases.

However, there is a lack of studies on the comparison between truck-involved crashes and total crashes, and nonlinear effects of truck percentage on truck-involved crashes. The objectives of this study are 1) to identify the factors affecting frequency of total crashes and truck-involved crashes on road segments and 2) to analyze their effects on crash frequency based on their relationships.

Data

The crash, road geometry and traffic data obtained from Ontario Ministry of Transportation were used in this study. Seven-year (2004-2010) data were collected from 6,475 roadway segments in all Ontario provincial highways. In this study, only the crashes that occurred within road segments without being influenced by intersections were analyzed.

The database consists of three different data sets: road geometry data, crash frequency data and traffic volume data. These data were combined by matching LHRS (Linear Highway Referencing System) numbers which are the identification number of each road segment. Road geometry is unique characteristics of each segment whereas crash frequency and traffic volume change every year. Each segment may have different length.

Since crash frequency was generally higher for longer segment, crash rate (i.e., crash frequency divided by length of segment) was computed for each segment. Injury severity of crashes was classified into five levels: fatal, major, minor, minimal, and property damage only (PDO). The variables in the data were listed in Table 1.

Table 1. Description of Variables

Numeric Variables	Mean	Standard Deviation
AADT (veh/day)	34891.68	70015.94
Truck Percentage (%)	16.25	11.21
Truck AADT (veh/day)	4453.44	72046.01
Length (km)	6.32	6.55
Posted speed limit (km/h)	87.81	10.63
Number of lanes	3.39	2.30
Lane width (m)	3.52	0.38
Surface width (m)	11.25	8.68
Streams	1.43	0.66
Median shoulder width (m)	0.58	1.02
Median width (m)	4.36	8.50
Shoulder width (m)	2.26	0.87

Table 1. Description of Variables (Continued)

Categorical variables	Categories
Road classification	Freeway, Arterial, Collector, Local
Shoulder type	Gravel, Paved, Partially paved
Median barrier	Divided, Undivided
Terrain	Flat, Mountainous, Rolling
Road surface type	Asphalt, Not asphalt

Figure 1 shows the comparison of crash rates on different road classifications for total crashes and heavy truck-involved crashes. Heavy truck-involved crashes are defined as the crashes involving at least one heavy truck. The figure shows that total crash rate was highest on freeways among the four road classifications. Heavy truck-involved crash rate was also highest on freeways. Similar patterns were observed for fatal and injury crash rates.

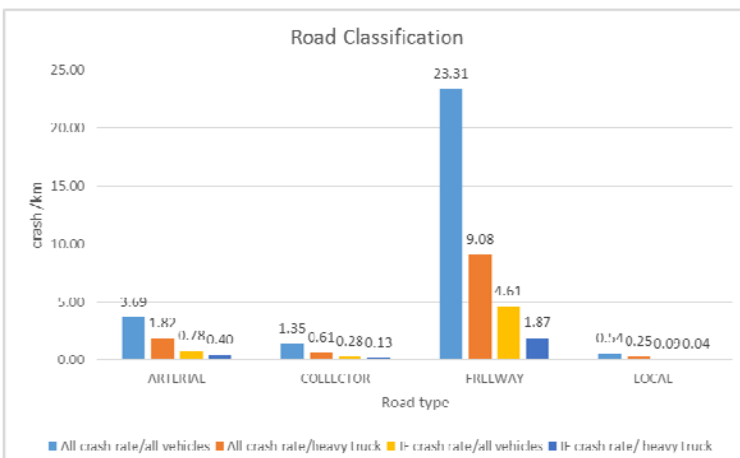


Figure 1. Crash rates by road classification

Figure 2 shows that crash rates for all crash types were higher for divided roads than undivided roads. This is because divided road segments generally have higher AADT and posted speed limit, which will increase the crash rate.

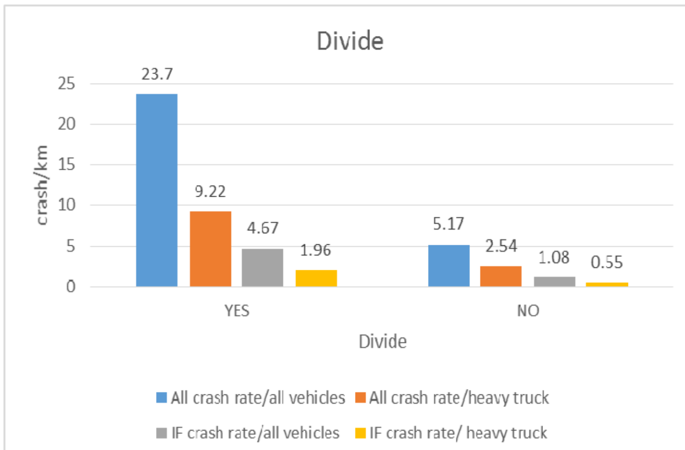


Figure 2. Crash rates by presence of median barrier

Figure 3 demonstrates that crash rates of total and fatal/injury crashes were highest for the segments with 100 km/h of the posted speed limit. However, although it is expected that crash rate generally increases as speed limit increases, the figure shows that the relationship between crash rate and speed limit is not linear.

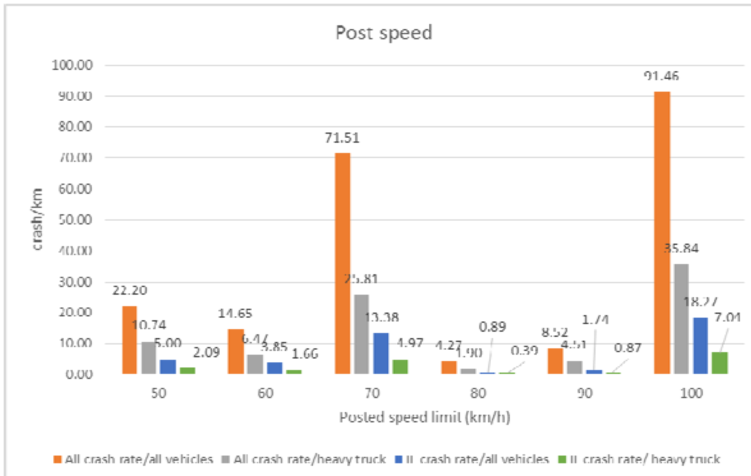
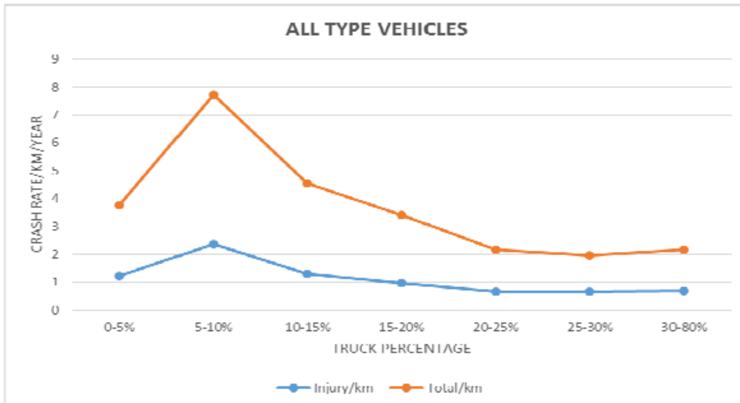
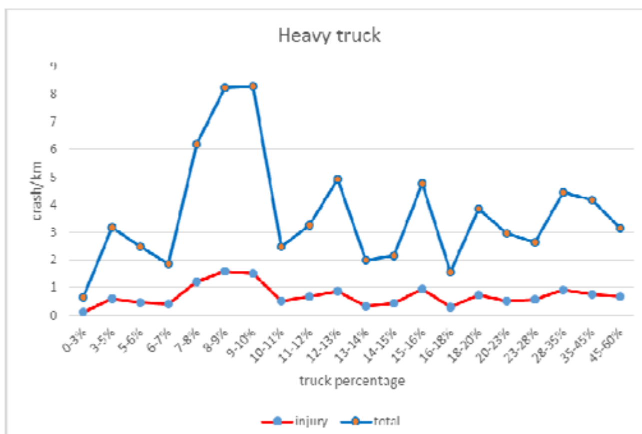


Figure 3. Crash rates by posted speed limit

Figure 4(a) shows the trend of total crash rates for different truck percentages. The crash rate was highest for 5-10% for both total crashes and fatal/injury crashes. Similar pattern was observed for truck-involved crashes as shown in Figure 4(b). In truck-involved crashes, truck percentage categories have smaller intervals. The figure shows that the crash rate was highest for 8-10% for both total crashes and fatal/injury crashes.



(a) Total crashes



(b) Truck-involved crashes

Figure 4. Crash rates by truck percentage

Method

Over past decades, the methodologies have been developed to identify the relationship between crash frequency and contributing factors. A majority of the previous studies used the generalized linear models (GLM). As an extension of traditional linear models, GLM can fit data with distributions in exponential family and allow independent variables linearly related to dependent variables through a nonlinear link function. GLM describes a dependent variable in a function of explanatory variables as follows:

$$Y = \exp(\alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k) \tag{1}$$

where Y is the expected crash frequency during a certain time period; X_k is the explanatory variable related to crash frequency; α is a constant, and β_k is the coefficients for the explanatory variables X_k. A positive coefficient β_k indicates that as the value of X_k increases, crash frequency also increases.

GLM can be developed by choosing different distributions of crash frequency such as Poisson and negative binomial (NB) distributions. Poisson regression models have been widely used in predicting crash frequency. Kumara and Chin (2005) used this methodology with modification of parent Poisson regression model. They found that left-turn volume, number of signal phases during one cycle and shorter sight distance were significant factors affecting crash frequency at three-legged signalized intersections. Other researchers (Ye et al., 2013; Li et al., 2013) also applied the Poisson regression model to analyze crash frequency data.

However, since Poisson distribution assumes that the mean and the standard deviation are equal, the distribution is not valid if the variation in crash frequency is larger than the mean crash frequency (i.e., over-dispersion). To account for over-dispersion, the negative binomial distribution with the error term following Gamma distribution has been applied to crash frequency models (Hauer, 2001). Unlike Poisson distribution, negative binomial distribution allows the standard deviation of crash frequency to vary with the mean crash frequency. More specifically, the standard deviation of the crashes equals the square root of the mean + mean^2 / k where k is the over-dispersion parameter which is determined from the data. For instance, Hu (2010) used the NB regression model to determine the factors that are statistically significant to traffic collisions at highway-railroad grade crossings.

Results and Discussion

GLMs were developed to estimate the influence of variables on crash frequency using SAS 9.2 (SAS Institute, 2012). To investigate potential nonlinear relationship between crash rate and truck percentage, dummy variables for different ranges of truck percentage were created as shown in Table 2:

Table 2. Dummy Variables of Truck Percentage

Variable name	Value
Truck%(< 5)	= 1 if 0 ≤ percentage < 5%, = 0 otherwise
Truck%(5-9)	= 1 if 5% ≤ percentage < 10%, = 0 otherwise
Truck%(10-14)	= 1 if 10% ≤ percentage < 15%, = 0 otherwise
Truck%(15-19)	= 1 if 15% ≤ percentage < 20%, = 0 otherwise
Truck%(≥ 20)	= 1 if percentage ≥ 20%, = 0 otherwise

There were sufficient samples in each category of truck percentage shown in Table 2. The result of GLM for total crashes is summarized in Table 3. Only variables statistically significant at a 95% confidence level were included in the model. For truck percentage, truck percentage lower than 5% was taken as the base case. The result shows that crash frequency was higher for truck percentage of 5-9% compared to the base case. However, crash frequency gradually decreases as truck percentage increases from 10%. This result verifies that the relationship between truck percentage and crash frequency is nonlinear. This non-linear relationship is potentially because vehicle conflicts are less when truck percentage is very low. On the other hand, as truck percentage increases, car drivers tend to be more cautious to avoid conflicts with trucks.

Table 3. Estimated Parameters of GLM (Total crashes)

Parameter	Estimate	Standard Error	p-value
Intercept	-0.7394	0.147	<.0001
Truck%(< 5)*	-	-	-
Truck%(5-9)	0.6307	0.0421	<.0001
Truck%(10-14)	0.5580	0.0434	<.0001
Truck%(15-19)	0.4851	0.0476	<.0001
Truck%(≥ 20)	0.3987	0.0459	<.0001
AADT	0.0113	0.0004	<.0001
Local street*	-	-	-
Arterial	1.0844	0.0493	<.0001
Collector	0.8687	0.0455	<.0001
Freeway	1.0434	0.0867	<.0001
Length	0.0179	0.0021	<.0001
Streams	0.6048	0.0679	<.0001
Lane width (m)	0.1263	0.0325	<.0001
Non gravel shoulder*	-	-	-
Gravel shoulder	-0.1938	0.0322	<.0001
Asphalt	0.3137	0.0712	<.0001
Flat terrain*	-		
Rolling or mountainous terrain	0.0698	0.0267	0.0089
No. of observations: 4800			
Deviance = 4891.66			
Log likelihood ratio = -16234.25			

*Base case.

The result also shows that wider lanes, higher traffic volume and longer length of segments were associated with higher crash frequency. It was found that crash frequency was higher for collector, arterial and freeway than local street. This indicates that crashes are more likely to occur on high-speed roadways with higher traffic volume. Also, crash frequency was lower for gravel shoulder than other types of shoulder where it was higher for rolling or mountainous terrain than flat terrain.

To analyze the truck-involved crash only, the crashes which involve at least one truck were extracted from the original data base. Dummy variables were also created to capture the nonlinear relationship between crash frequency and truck percentage. However, when truck percentage was categorized as shown in Table 2, the effects of dummy variables in the truck-involved crash model were not significant. Thus, more refined categories of truck percentage with narrower ranges were used for truck-involved crashes.

The result of GLM for truck-involved crashes is shown in Table 4. Again, only variables statistically significant at a 95% confidence level were included in the model. The result shows that crash frequency was significantly higher for truck percentage of 8-10% compared to truck percentage less than 5%. However, crash frequency was lower for truck percentage of 16-18%. This result indicates that frequency of truck-involved crashes was much lower at very high truck percentage than very low truck percentage unlike total crashes. This difference reflects that at high truck percentage, drivers are more likely to avoid following trucks and the chances of collision with trucks decrease. However, this results in an increase in non-truck-involved crashes due to more frequent lane changes to avoid following trucks.

Some additional variables are only significant in the truck-involved crash model such as speed limit, number of lanes, divided/undivided road and shoulder width. It was found that as speed limit, number of lanes and shoulder width increase, the frequency of truck-involved crashes also increases. This is because higher speed limit, more number of lanes and wider shoulder are correlated with higher AADT. On the other hand, frequency of truck-involved crashes is lower on divided road segments than undivided road segments. This is potentially because truck drivers generally choose the outer lanes and they are less likely to hit a median barrier on divided roads.

Table 4. Estimated Parameters of GLM (Truck-involved crashes)

Parameter	Estimate	Standard Error	Pr > ChiSq
Intercept	0.4482	0.3078	0.1453
Speed limit (km/h)	0.0057	0.0031	0.0718
Truck%(< 5)*	-	-	-
Truck%(≥ 8 and < 9)	0.2412	0.0813	0.003
Truck%(≥ 9 and < 10)	0.2043	0.081	0.0117
Truck%(≥ 16 and < 18)	-0.2865	0.0934	0.0021
Local street*	-	-	-
Arterial	1.0053	0.0906	<.0001
Collector	0.827	0.083	<.0001
Freeway	1.4162	0.15	<.0001
Length	0.0141	0.004	0.0005
Number of lanes	0.229	0.0191	<.0001
Non gravel shoulder	-	-	-
Gravel shoulder	-0.1756	0.0536	0.0011
Divided (1=divided, 0= undivided)	-0.318	0.1224	0.0094
Shoulder width (m)	0.2844	0.0361	<.0001
Lane width (m)	0.0858	0.0443	0.0529
No. of observations: 5728			
Deviance = 6297.58			
Log likelihood ratio = -17214.00			

Conclusions and Recommendations

This study identifies the factors affecting frequency of truck-involved crashes and total crashes on road segments and analyzes their effects on crash frequency using GLMs. It was found that crash frequency generally increases as AADT and length of segment increase. Also, road classification, shoulder type and lane width are associated with crash frequency. However, there was a difference in the effect of truck percentage on crash frequency between total crashes and truck-involved crashes. Unlike total crashes, frequency of truck-involved crashes was lower at higher truck percentage (16-18%) compared to truck percentage less than 5%. Also, some variables are only significant in truck-involved crashes but not in total crashes. This difference in the results reflects that truck-involved crashes have unique characteristics and they should be analyzed separately.

However, there is a limitation in this study. Although we found that a nonlinear relationship exists between truck percentage and crash frequency, we did not investigate other nonlinear relationships. In the future work, it is recommended to capture nonlinear effects of other variables on crash frequency using more advanced models.

Acknowledgment

The authors would like to thank the Ministry of Transportation of Ontario (MTO) for providing the data. Opinions expressed in this report are those of the authors and may not necessarily reflect the views and policies of the MTO. This research was supported by the University of Windsor and Natural Sciences and Engineering Research Council of Canada.

References

Caliendo, C., De Guglielmo, L. M., and Guida, M., 2013. A crash-prediction model for road tunnels. Accident Analysis and Prevention, 55, 107– 115.

Choi, S., Oh, C. and Kim, M., 2014. Risk factors related to fatal truck crashes on Korean Freeways. Traffic Injury Prevention, 15, 73–80.

Dong, C., Clarke, B. D., Richards, H. S. and Huang, B., 2014. Differences in passenger car and large truck involved crash frequencies at urban signalized intersections: An exploratory analysis. Accident Analysis and Prevention, 62, 87- 94.

Haleem, K., Gan, A. and Lu, J., 2013. Using multivariate adaptive regression splines (MARS) to develop crash modification factors for urban freeway interchange influence areas. Accident Analysis and Prevention, 55, 12–21.

Hauer, E., 2001. Overdispersion in modeling accidents on road sections in Empirical Bayes estimation. Accident Analysis and Prevention, Vol. 33, 799-808.

Hu, W., 2010. Factors associated with median related crash frequency and severity. Ph.D. Thesis. The Pennsylvania State University, University Park, Pennsylvania, USA.

Kotikalapudi, S. and Dissanayake, S., 2013. Study of characteristics and evaluation of severity affecting factors associated with large-truck crashes. Advances in Transportation Studies an international Journal Section A, 29, 19-34.

Kumara, S. S. P. and Chin, H. C., 2005. Application of Poisson underreporting model to examine crash frequencies at signalized three-legged intersections. Journal Transportation Research Record: Journal of the Transportation Research Board, 1908, 46-50.

Li, Y., Liu, C. and Ding, L., 2013. Impact of pavement conditions on crash severity. Accident Analysis and Prevention, 59, 399–406.

National Highway Traffic Safety Administration, 2014. Traffic Safety Facts 2012. [DOT HS 811 868]. Department of Transportation, Washington, DC, USA.

SAS Institute, 2012. Cary, NC, USA.

Transport Canada, 2015. Canadian Motor Vehicle Traffic Collision Statistics: 2012. http://www.tc.gc.ca/media/documents/roadsafety/cmvtcs2012_eng.pdf. Accessed January 2015.

Ye, X., Pendyala, R. M., Shankar, V. and Kondurid, K. C., 2013. A simultaneous equations model of crash frequency by severity level for freeway sections. Accident Analysis and Prevention, 57, 140–149.

PRESENTATION PERMISSION FORM
2015 CTRF ANNUAL CONFERENCE
May 24 – 26, MONTREAL, QUEBEC

Author Name(s): Ran Ran, Chris Lee

Title of Presentation: Effects of Geometric and Traffic Factors on Frequencies of Truck-involved Crashes on Ontario Highways

- [x] I/we **give** CTRF permission to post my/our 2015 conference presentation to the CTRF website

- [ ] I/we **do not give** CTRF permission to post my/our 2015 conference presentation to the CTRF website

Signature: Ran Ran

Date: March 6, 2015