Gargoum, Suliman, Li, Yang, El-Basyouny, Karim, & Kim, Amy

Factors Affecting Classification of Road Segments into High-and Low-Speed Collision Regimes.

AUTHOR POST PRINT VERSION

Gargoum, S., Li, Y., El-Basyouny, K., & Kim, A. (2017). Factors Affecting Classification of Road Segments into High-and Low-Speed Collision Regimes. Transportation Research Record, 2659(1), 98-105. https://doi.org/10.3141/2659-11

Factors Impacting the Classification of Road Segments into High/Low Speed-Collision Regimes

Suliman Gargoum, M.Sc.*
Department of Civil and Environmental Engineering
University of Alberta, Edmonton, AB, Canada
+1-780-200-0161
Email: gargoum@ualberta.ca

Yang Li, M.Sc.
Department of Civil and Environmental Engineering
University of Alberta, Edmonton, AB, Canada
+1-780-710-4616
Email: li18@ualberta.ca

Karim El-Basyouny, Ph.D., P.Eng.
Department of Civil and Environmental Engineering
University of Alberta, Edmonton, AB, Canada
+1-780-492-9564
Email: basyouny@ualberta.ca

Amy Kim, Ph.D., P.Eng.
Department of Civil and Environmental Engineering
University of Alberta, Edmonton, AB, Canada
+1-780-492-9203
Email: amy.kim@ualberta.ca

* Corresponding Author

Submitted for presentation at the 96th Annual Meeting of Transportation Research Board and publication in *Transportation Research Record: Journal of the Transportation Research Board*

TRB 96th Annual Meeting, Washington, D.C.
January 8-12, 2017

Word count: 6211 words + 4 (1 Figure & 3 Tables)*250 = 7211 equivalent words

ABSTRACT
The safety of locations operating under high-speed conditions could significantly differ from that of locations operating under low-speed conditions. Therefore, different approaches must be adopted when analyzing and managing speed and safety at locations operating under different regimes. However, first it is necessary to understand the factors affecting which speed-collision classification a site falls under. Locations operating under high speeds are typically expected to have more collisions compared to locations where speeds are low. Some locations, however, might experience a high collision rate even when speeds are low, or vice versa. This paper aims to identify the factors that affect a sites classification into any of those categories, using data collected at roads in Edmonton, Canada. Locations are divided into four speed-collision bins (high collision, high speed; high collision, low speed; low collision, high speed; low collision, low speed) and GIS maps of locations are produced to explore the spatial distribution of those locations. Moreover, logistic regression is used to understand the role different factors play in identifying the speed-collision bin a certain location belongs to. The results reveal that locations with a high collision rates but low speeds, have a relatively high population of heavy vehicles and trucks as well as high speed variability. As for locations with a low collision rates at high speeds, these sites were found to have a high level of protection through the presence of medians and shoulders with relatively low access density.

KEYWORDS
Low-Speed Severe Collisions; Speed-Collision Regime; Multinomial Logistic Regression; Road Features; Road Design; Speed Limit Credibility.

BACKGROUND

Speed has been shown to be a major contributing factor in all severe collisions (*1*). Statistics show that, in 2012, 30% of all fatal collisions in the US were speed-related crashes (*2*). Similarly, Transport Canada reported that, in 2011, almost a third of fatalities and a fifth of serious injuries were speed-related (*3*). While speed may not be the main factor in every severe collision, it is still one of the contributing factors in almost all types of severe collisions. Considering the vulnerability of the human body and the law of conservation of momentum, it is difficult to imagine a fatal or serious collision where speed was not a factor at all.

The significant contribution of speeding in both fatal and injury crashes leads to the belief that locations where a high severe-collision rate is observed would be expected to have higher speeds compared to locations where a severe-collision rate is low. Similarly, it is reasonable for a severe-collision rates to be low at locations with low speeds. What is less likely, however, is for a location to have a relatively high severe-collision rate and low speeds or low severe-collision rate with high speeds. While the latter may be acceptable, given that the location is still safe despite the high speeds observed at the location, the former situation where a location experiences high severe-collision rate despite low speeds is not so common. Since speeds are low, a high severe-collision rate could be attributed to the road design and geometry at those particular locations; another reason could be a high population of vulnerable road users at those locations, such as pedestrians or cyclists. Similarly, locations with a high population of heavy vehicles could also experience a high severe-collision rate even if they are operating at relatively low speeds. The high momentum involved in heavy vehicle crashes due to the size of those vehicles can have lethal impacts even at low speeds. Regardless of the reason, safety management and analysis at locations operating under high speeds must differ from those operating under low speeds.

Realizing the differences between collisions occurring at different levels of speeding, Abdel-Aty, Uddin and Pande (*4)* separated data into high-speed and low-speed regimes when developing collision prediction models. The authors argue that characteristics of crashes at high speed are significantly different from those at low speeds. In fact, one of the main findings of the study was that, although model development under the low-speed and high-speed regimes was similar, the variables with significant effects on collisions under the two regimes were different. This led the authors to conclude that it was of utmost importance to model crashes separately under different operating speed regimes. In work analyzing the indirect relationships between road geometry and collisions through speeds, Gargoum and El-Basyouny (*5)* found that the impacts of some geometric features on collisions were indeed mediated through speeds, indicating the importance of treating safety for locations operating at different speed regimes differently.

Taylor, Baruya and Kennedy (*6)*, developed a Poisson regression model to assess the effects of several variables, including speeds, on crashes observed at rural single carriage roads across the UK. Data used in the study included injury accident data, traffic flows, individual vehicle speeds, and geometric and other features of the roads. The data was collected from 174 different road sections where the speed limits were all 60mph. Because of the difference in road conditions among the different road segments, the authors created homogenous "road quality" groups into which segments were classified, analyzing the relationship between collisions and speed. The authors found that average speed was positively correlated with collisions, but these findings were only reached once the locations were classified into homogenous groups based on road quality. The reason why such a classification is necessary, according to Elvik, Christensen and Amundsen (*7)*, is that the "best roads tend to have the highest speed limits" — a better road here is a road with higher quality attributes than other roads. In fact, even in Taylor et al.'s study

(*6*), the variables on which the grouping was based were all attributes of road quality. In an earlier study, Taylor, Lynam and Baruya (*8)* studied the effects of speed on urban classified roads; data was collected from 300 road sections and linked with 1590 injury crashes. In this case, a statistical cluster analysis was used to classify the sites based on their speed characteristics. The speed attributes used in the classification included average speed, variability in speeds and the proportion of slow vehicles. This classification yielded four different groups. The findings of the study revealed that, among other variables, average speeds were strongly related to accident frequencies. As in Taylor et al.'s other study (*6*), the classification of sites was necessary to find these strong relationships. Although pervious research does acknowledge the importance of classifying roads into homogenous groups in order to properly analyze the safety of a particular location *(4)*, to the best of the authors' knowledge, no studies have assessed the factors that affect a location belonging to a certain speed-collision classification.

This paper aims to explore the factors affecting the likelihood of road segments belonging to one of the four speed-collision regimes, using empirical data from the city of Edmonton, Canada. Data from over 600 urban roads is considered and locations were classified as operating under one of four regimes (high collisions, high speeds; high collisions, low speeds; low collisions, high speeds; low collisions, low speeds). A GIS map of the locations was generated to understand the spatial distribution of those sites around the city. In addition to spatial analysis, multinomial logistic regression models were developed to assess the influence of factors such as the road geometry, traffic composition had on the likelihood of a location being classified into one of the speed-collision bins. Multinomial logistic regression is a form of regression which analyses the effects of several independent variables on dependent variable of nominal levels.

DATA DESCRIPTION

Speed and Collision Data

Data used in this study was collected from 643 urban roads in Edmonton, Canada during the period from 2009 to 2014. The locations included arterial and collector roads which were spread around different parts of the city. The speed data was collected through speed surveys conducted by the city. Speed survey data was then used to calculate speed summary data at each of the locations (e.g. average speed, 85th percentile speeds, etc.). Speed data was linked to the collision frequency observed at each of the locations during the same period of time. Only severe collisions (fatal and injury collisions) observed at the locations of interest were included in the dataset, since these are the collisions where speeds are expected to have the highest impacts (*1*). Information about traffic flow at each location was also measured during speed survey visits and included in the analysis.

Features of the Road Environment

In addition to speed and collision information, the dataset was augmented to include features of the road environment. These features were manually collected at each site using Google Maps images; views from multiple years were compared to ensure that the feature of interest did not change. Most of the features were represented using binary variables, where "1" represents the presence and "0" represents the absence of that attribute. These features included the horizontal curves on the road segment (0 for tangent segment and 1 for curved), the vertical grade on the road, and the presence or absence of shoulders, bus stops and service roads on the road segments. A service road is a road directly adjacent to a higher volume road (typically an arterial), which is used for local access. It is worth noting that the shoulder variable represents the lateral clearance on the edge of the road, used in case of an emergency; this is not as an extra travel lane.

For other design features, the variables were non-binary. For on-street parking, three levels were defined (0 for no parking, 1 for one-sided parking and 2 for two-sided parking). Similarly, for the separation between the two approaches on the road segment (i.e., the median), three levels were defined (0 for no median, 1 for a physical median and 2 for a Two-Way-Left-Turn-Lane (TWLTL). For each site, the number of lanes per direction was also recorded. Sidewalk was another categorical variable which was initially coded as 0 if there was no sidewalk, 1 for a road with a sidewalk on one side and 2 for a road where a sidewalk was present on either side of the road segment. Continuous variables considered were the total road width in meters and the number of access points per kilometer (i.e. the number of access points per segment divided by the segment length). Due to the fact that some categories were underrepresented, the median variable and the sidewalk variable were both recoded as binary variables. It is worth mentioning here that the number of bus stops on a segment was not included as a variable since the majority of segments that had bus stops (around 25% of the segments) only had one bus stop (i.e. the variable would have been underrepresented). Descriptive statistics of the dataset are provided in Table 1.

SITE CLASSIFICATION & HYPOTHESIS

Road segments considered in the study were divided into four different groups. Segments were assigned to different groups based on the collision rates and the speeds recorded at the location. Group I included locations where a high severe-collision rate and high speeds were observed (HCHS); Group II included locations with a high severe-collision rate but low speeds (HCLS); Group III included segments where severe collision rate was low but speeds were high (LCHS); Group IV included locations where both severe collision rate and speeds were low (LCLS).

High/Low Collision

In order to determine whether a site belonged to the high or low collision group, an acceptable threshold for the number of severe and fatal collisions, for a site to be considered a high collision site, had to be defined. Since different types of roads have different road features that affect the collision risk, locations were split into homogenous groups, and a different threshold was defined for each group of sites. These road groups are categorized based on road classification (arterial roads/collector roads), number of lanes and whether the average daily traffic is above the mean or less than or equal to the mean. Since arterial roads are more likely to have a greater number of lanes than collectors, arterial roads and collector roads were sub-classified into three categories and two categories, respectively. The sub-classification was based on the number of lanes. Specifically, for any of the arterial groups, these were categorized by the number of lanes: roads with one lane in each direction, with two to three lanes per direction or with more than three lanes per direction. Whereas, collector road groups are categorized into one type that has only one lane per direction, and another type that has more than one lane per direction.

Collision rate at every road within a group was compared to the average collision rate at locations of similar characteristics and similar road features. The 643 road segments were classified into 10 groups, and high collision sites in each of the 10 groups were identified.

High/Low Speed

Regarding speeds, the threshold was defined based on the average speed and the posted speed limit at a certain location. Locations where the average speed was higher than the posted speed limit were considered to be operating in a high-speed regime, while locations where the average speed was less than the speed limit were considered low-speed locations. Average speed has previously been used to separate high-speed and low-speed regimes in (*4*).

Hypothesized Relationships

Before exploring the effects of different factors on a road segment belonging to a certain speed-collision bin, it was first necessary to establish the hypothesized effects of those variables. First of all, locations with a higher road classification (i.e., arterials compared to collectors and collectors compared to local roads) and locations with more lanes are typically associated with higher travel speeds, and consequently, a higher severe-collision rate is expected at those locations. Therefore, it was hypothesized that the higher the road class or the number of lanes, the more likely it is for a road segment to belong to the high collision, high speed class.

Given that speeding is a leading factor in road trauma (*9*), it is likely for locations with high speeds to have high collisions. However, if the sites are designed in such a way that makes the driving environment safer, for example, equipping sites with lateral clearance on shoulders, installing median barriers separating the two carriageways or reducing access point density, these sites may experience low collision rates irrespective of high travel speeds. In contrast, although low operating speeds are expected at locations with horizontal curves, it is still expected that those locations would experience a high number of collisions compared to straight segments. Hence, it is more likely for a location with a horizontal curve to belong to the HCLS category.

Similarly, locations that have sidewalks or pedestrian crossing facilities are typically designed to accommodate pedestrian and cycling activities. Hence it is reasonable to assume that at these locations speeds ought to be relatively low. Moreover, since pedestrians and cyclists are vulnerable road users who are at high risk of severe injury and death even at low speeds, it is expected that sites equipped with pedestrian facilities may experience high severe-collision rate even when operating at low speeds, assuming high vehicular and pedestrian activity.

Operating speed on a road segment is often considered to be 85th percentile speed, since 85 percent of drivers will drive at a speed equal to or higher than that speed. Since 85 percent of the driving population chooses this speed, it is reasonable to assume that the majority of drivers feel that this speed is safe and that it could be the most credible speed for this location. If the 85th percentile speed exceeds the posted speed limit at a site by a high margin, it is likely that the speed limit is set too low for the site, and therefore, assuming that driver assessment of a safe speed is valid, we expect the number of severe collision at such a site to be low despite high speeds.

SITE VISUALIZATION

Once the sites were classified into the four speed-collision bins, GIS maps of the locations were plotted in order to visualize the distribution of the sites around the city. Figure 1 depicts the distribution of locations in the city of Edmonton based on four speed-collision groups into which locations were classified in this paper. Since locations with high collisions are considered critical, road segments were color coded as follows: group I (HCHS) in red, group II (HCLS) in yellow, group III (LCHS) in green and group IV (LCLC) in grey.

The distribution of the segments among the four speed-collision bins is quite promising since the majority of sites (over 60%) belonged to the low collision categories; this includes 246 sites in the HSLC group and 192 sites in the LSLC group. The HCHS category had the lowest number of locations, with only 74 sites. The remaining segments (131) were locations with a high collision rate despite low speeds.

As evident from Figure 1, site distribution seems highly random; however, a few important observations can be drawn from the Figure 1. First of all, it can be inferred that in the downtown area (central area of the figure), the majority of the sites fall under the high-collision regime. In other words, most of the locations in that region are either high collision, high speed (HCHS) segments (red) or high collision, low speed (HCLS) segments (yellow). HCLS locations also seem to spread into areas north of downtown up to Yellowhead Trail. In the region between Yellowhead Trail and up north towards the Anthony Henday Drive, safety seems to improve again with most of the sites there belonging to the low-collision bins. In fact, it is worth noting that despite high speeds on the Anthony Henday Drive and the Yellowhead Trail segments, these roads seem to be relatively safe, with the majority of segments belonging to the LCHS (green) bin.

In contrast to the downtown region and the region slightly north of downtown, LCLS and LCHS locations were predominant in the southwestern parts of the city, with the majority of the segments in that region colored in black or green. A low number of collisions was found in the west side of the city, with only a few HCHS and HCLS locations in that region.

In general, the figure shows that central locations around downtown and the University of Alberta area seem more likely to be high-collision segments regardless of the speeds (i.e., they could be HCLS or HCHS). On the contrary, LCHS and LCLS segments are more common in residential areas located on the southern, western and northern outskirts of the city.

REGRESSION ANALYSIS

In order to develop a better understanding of the factors that influence a location belonging to one of the four groups, a multinomial logistic regression model was developed. Multinomial logistic regression was used since it does not make any assumptions about the distribution of the independent variables, which in the case of this study were either binary or categorical.

Multinomial Logistic Regression

Multinomial logistic regression models the probability of event occurrence based on a set of independent variables (*10*). It can be seen as a form of binary logistic regression where the dependent variable can include more than two levels. In binary logistic regression, the response variable is binary (i.e., it is either a 0 or a 1). Multinomial logistic regression can be ordinal or nominal. In ordinal logistic regression choices on the scale of the dependent variable are ordered (i.e. despite the dichotomous nature of the dependent variable, a higher option means an increase/decrease in the dependent variable). In nominal logistic regression, there is no order to the response scale of the dependent variable (i.e. moving from one category to another does not have any quantitative meaning). The four speed-collision bins considered in this paper were nominal, hence, multinomial logistic regression with a nominal response was used to model the data. The general notation for logistic regression is described in the following paragraph.

Logistic regression is used to predict the odds of a certain event occurring, based on the values of several independent variables. Let Y_i be the dependent variable that denotes the group to which the location *i* belongs. In this case, $p_i = P(Y_i = k)$ represents the probability of location *i* belonging to group *k*. Let $p_j = P(Y_i = j)$ represent the probability of location *i* belonging to reference group *j*. The log-odds of location *i* belonging to category *k* can then be expressed using a logit model of the following functional form:

$$\ln\left(\frac{p_i}{p_k}\right) = \beta_o + \beta_1 X_{i1} + \dots\dots\dots + \beta_n X_{in} \quad (1)$$

which can also be expressed as

$$\frac{p_i}{p_k} = e^{\beta_o + \beta_1 X_{i1} + \dots\dots + \beta_n X_{in}} \quad (2)$$

where, $\beta_0, \beta_1 \dots\dots \beta_m$ are regression parameters of the model, and X_{ij} are the covariates representing several factors that have alleged effects on a location belonging to a certain category. The regression parameters are evaluated using an iterative maximum likelihood (ML) procedure. All analysis was done using IBM SPSS Statistics version 23. Before proceeding with the goodness of fit statistics for the models developed, it is worth noting that in logistic regression the odds of an event occurring are evaluated relative to a reference category. Therefore, all results are estimated relative to one of the levels in the dependent variable known as the reference category. In other words, each category of the predictor variable (except the reference category) is compared to the reference category.

The goodness of fit of the developed models was assessed using the chi-squared of the -2 log likelihood ratio and the Pseudo R-squared estimates. Both measures indicated that the estimated models were good fit of the data with a significant chi-squared measure ($p<0.0005$) and a Nagelkerke *R*-squared value of 0.821. The significant chi-squared value indicates that the final model provides a significant improvement in prediction when compared to the baseline model (i.e., with intercept only) and the *R*-squared value of 0.821 indicates that the model explains more than 80% of the variance.

MODELING RESULTS

Table 2 shows the results of the regression analysis relative to the HCHS category. Variables were added to the models if it was hypothesized that they had effects on the classification of a site into one of the four categories (HCHS, HCLS, LCHS or LCLS). All variables with any statistically significant effects on the dependent variable at a 10% significance level were kept in the model. The models were run while changing the reference category, however, the parameter estimates are not presented in table 2 for brevity. It is worth noting that although the parameter estimates are not shown, the results are discussed. As seen in Table 2, eight variables of those considered had shown statistically significant effects on whether a road segment belonged to a certain class. These eight variables are the difference between operating speeds and the speed limit, access density, the presence of a shoulder lane, the presence of a median, road class, the standard deviation of speeds, the number of lanes and the percentage of trucks.

In general, examining the outcomes of the regression, it could be inferred that HCHS segments are more likely to be collector segments where operating speeds exceed speed limits by high margins with a low speed variability among road users. The segments also had a high number of lanes, no median and a high access density. HCLS segments, on the other hand, were more likely to be arterial segments with high speed variability but with operating speeds closer to the PSL. The segments were also more likely to have no shoulder, no median, a high percentage of nighttime crashes and a high percentage of trucks.

LCHS segments were more likely to have low access density and a low number of lanes. These segments were also more likely to have a lateral shoulder clearance and a median barrier. In contrast, LCLS segments were more likely to experience high speed variability but operating speeds closer to the speed limit. Segments with medians and a low proportion of nighttime crashes were also more likely to belong to this class.

The results show that high collision, low speed (HCLS) locations were more likely to be arterial road segments with low operating speeds but high speed dispersion when compared to high collision, high speed (HCHS) segments. As expected, locations with a high percentage of trucks were more likely to belong to the HCLS class. As already noted, the size of heavy vehicles means that, even at low speeds, high momentum is expected in crashes involving these type of vehicles, and consequently, a high rate severe-collisions was expected at locations where these vehicles are present. What is interesting in the case of HCLS locations is that geometric features did not seem to have significant effects in distinguishing between HCHS and HCLS, even though the initial hypothesis was that segments with high collisions, despite having low speeds, could have design deficiencies or high pedestrian activity. With that being said, geometric attributes did play a role in distinguishing between LCLS and HCLS sites. In that context, it was found that a location was more likely to be a HCLS segment if it had no median barrier and no shoulder. Although these attributes are typically found on segments with high speeds, it seems they are also effective in improving safety at low-speed locations too. In addition to geometric features, increases in the difference between operating speeds and the PSL also increased the likelihood that a location belonged to the HCLS class. Locations with a higher percentage of nighttime crashes were also more likely to belong to the HCLS category; this is reasonable considering that impaired driving, which increases at night, is an important cause of collisions even at low-speed locations.

When comparing LCHS segments to HCHS segments, it was noticed that the increase in the number of lanes and in access density both increased the likelihood of a segment being a HCHS segment; similarly, segments with no physical median barrier and no lateral shoulder were also more likely to belong to the HCHS category. This implies that for a segment to have low collisions despite high speeds, it must have fewer lanes, lower access density, a median barrier and a shoulder.

These findings seem highly intuitive. Lane-changing activity at high speeds causes conflicts between vehicles in different lanes, and these conflicts at high speeds are likely to lead to collisions. Therefore, a high speed segment is more likely to have low collisions if it has fewer lanes. Similarly, a segment with high access points is also expected to experience more conflicts between road users and potentially a higher number of collisions as shown in previous research (*11*).

The findings with respect to presence of a median were also in line with expectation. The presence of a median increases the likelihood of a segment being a LCHS segment as opposed to HCHS; this indicates that the median barrier does provide the anticipated protection. Although medians might be associated with an increase in fixed-object collisions, they help limit head-on collisions, which are typically severe collisions. In fact, even when comparing HCHS locations to LCLS locations, the model shows that the presence of a median significantly increases the likelihood of a location belonging to the LCLS category. In addition to medians, presence of a shoulder lane also increases the likelihood of a segment being a LCLS location as opposed to a HCHS location, which is another indicator that these design features seem to be effective in limiting the dangers associated with high speeds. The findings discussed in the last few paragraphs are summarized in Table 3. The table shows a selection of factors which influence a road segment belonging to bin 1 over bin 2 and vice versa. For instance, if the road is a collector it is more likely to be a HCHS site than a HCLS.

It is worth noting that, some of the variables that were initially hypothesized to have effects on the classification were not found to be statistically significant; these include measures of pedestrian activity, such as the existence of a pedestrian crossing facility and the existence of sidewalks along the side of the road. The initial hypothesis was that these variables could be associated with higher pedestrian activity, and hence would increase the likelihood of a site being a high-collision location despite low speeds. However, the fact that these variables were not statistically significant indicates that this was not the case. Since the usage levels of those facilities are unknown, this could be a matter of these variables not truly reflecting pedestrian activity and hence not having an impact on the classification. Another variable where the hypothesized effects did not hold was the effect of horizontal curves on the classification. The initial hypothesis was that segments with horizontal curves would have lower speeds due to the design of the curves but higher collisions due to the difficulty of driving along those segments. Nevertheless, the variable was again found to be statistically insignificant.

CONCLUSIONS & FUTURE RESEARCH

Whether a particular location is operating under high-speed conditions or low-speed conditions has significant influence on the safety of that location. Therefore, different approaches must be used when analyzing and managing speed and safety at locations operating under those different regimes. However, first it is necessary to establish an understanding of the factors affecting which speed-collision classification a site falls under. In this paper, data collected from 2009 to 2014 in the city of Edmonton, Canada was used to perform a detailed analysis. The study analyzed factors that affect the likelihood of a particular location belonging to one of four speed collision bins (high collision, high speed (HCHS); high collision, low speed (HCLS); low collision, high speed (LCHS); and low collision, low speed (LCLS)). The results revealed that, among other factors, the difference between operating speeds and the speed limit, access density, the presence of a shoulder lane, the presence of a median, road class, the standard deviation of speeds, the number of lanes and the percentage of trucks were all significant variables in identifying whether a site belonged to a particular class.

Although the study provides valuable insight into the effects of road features and operating conditions on site classification, future research could expand on the analysis performed in this paper in many ways. One focus would be to explore the impacts of dynamic factors, such as pedestrian activity, on locations being classified into a certain category. Although this study does use a proxy for pedestrian activity, such as the presence of a pedestrian crossing facility and the presence of sidewalks, these factors had hardly any significant impacts. This finding could be due to information about the occupancy rates of those facilities being unknown. Another opportunity for future work is to reclassify sites that belong to each of the speed-collision bins. For instance, a different definition could be used to classify high/low speeds and high/low collisions; these changes may provide more insight into the impacts of different factors on classification. Moreover, future research could also consider integrating the effects of collision injury type into the analysis. Although outside the scope of this paper, the type of injury experienced during a low-impact collision could be critical in a severe collision. Research shows that even at low impact severe injuries to soft neck tissues can be experienced, potentially resulting in disability (*12*). Jakobsson, Isaksson-Hellman and Lindman (*13)* state that a large number of occupants reporting neck injuries is found for collisions where only repair of the bumper system is required. Although safety systems, such as airbags, are effective in reducing those injuries, these systems are often not activated during low-impact collisions, indicating the importance of considering injury type in future research.

ACKNOWLEDGEMENTS
This study was sponsored by NSERC Discovery Grant. The authors would like to thank the City of Edmonton and Office of Traffic Safety for providing the data to conduct this study.

REFERENCES

[1] Nilsson, G. Effects of speed limits on traffic accidents in Sweden. 1982.
[2] Administration, N. H. T. S. Traffic Safety Facts, 2012 Data:. Vol. 65, No. 4, 2015, p. 452.
[3] Road Safety Canada Consulting. *Road Safety in Canada.* Publication 978-1-100-18621-4, Transport Canada, 2011.
[4] Abdel-Aty, M., N. Uddin, and A. Pande. Split models for predicting multivehicle crashes during high-speed and low-speed operating conditions on freeways. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1908, 2005, pp. 51-58.
[5] Gargoum, S. A., and K. El-Basyouny. Exploring the association between speed and safety: a path analysis approach. *Accident Analysis & Prevention,* Vol. 93, 2016, pp. 32-40.
[6] Taylor, M. C., A. Baruya, and J. V. Kennedy. *The relationship between speed and accidents on rural single-carriageway roads*. TRL, 2002.
[7] Elvik, R., P. Christensen, and A. Amundsen. Speed and road accidents. *An evaluation of the Power Model. TØI report,* Vol. 740, 2004, p. 2004.
[8] Taylor, M. C., D. Lynam, and A. Baruya. *The effects of drivers' speed on the frequency of road accidents*. Transport Research Laboratory Crowthorne, 2000.
[9] OECD. Speed Management.In, Joint Transport Research Committee, Paris, 2006.
[10] Hosmer Jr, D. W., S. Lemeshow, and R. X. Sturdivant. *Applied logistic regression*. John Wiley & Sons, 2013.
[11] Leur, P. d., and T. Sayed. Development of a road safety risk index. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1784, 2002, pp. 33-42.
[12] Distner, M., M. Bengtsson, T. Broberg, and L. Jakobsson. City safety—a system addressing rear-end collisions at low speeds.In *Proc. 21st International Technical Conference on the Enhanced Safety of Vehicles*, 2009.
[13] Jakobsson, L., I. Isaksson-Hellman, and M. Lindman. WHIPS (Volvo cars' Whiplash Protection System)—the development and real-world performance. *Traffic injury prevention,* Vol. 9, No. 6, 2008, pp. 600-605.

TABLE 1: Descriptive Statistics

Variable	Minimum	Maximum	Mean	Std. Deviation
Geometric Attributes				
Number of Lanes	1.00	11.00	3.18	1.55
Road Width (m)	6.60	99.74	17.58	10.15
Median	0.00	1.00	.328	.47
Horizontal Curve	0.00	1.00	.538	.50
Pedestrian Crossing	0.00	1.00	.597	.49
Sidewalk	0.00	1.00	.869	.34
Shoulder	0.00	1.00	.103	.30
One-way	0.00	1.00	.051	.22
Parking	0.00	2.00	.942	.95
Bus Stop	0.00	1.00	.739	.44
Service Road	0.00	1.00	.093	.29
Vertical Curve	0.00	1.00	.072	.26
Access Density Per km	0.0000	4.4723	.985	.79
Length (m)	80.45	6483.46	1075.6	843.91
Operating Conditions				
Posted Speed Limit (PSL)	30.00	100.00	54.51	8.62
Average Daily Traffic (ADT)	59.3	39861.8	6101.9	7250.32
Operating Speed and Speed Limit Difference	-17.25	31.65	9.05	5.78
Percent Night Crashes	0.00	1.00	.094	.22
Standard Deviation of Speeds	6.90	15.35	10.38	1.49
School Zone	0.00	1.00	.145	.35
Percentage of Trucks	.01	.21	.039	.03

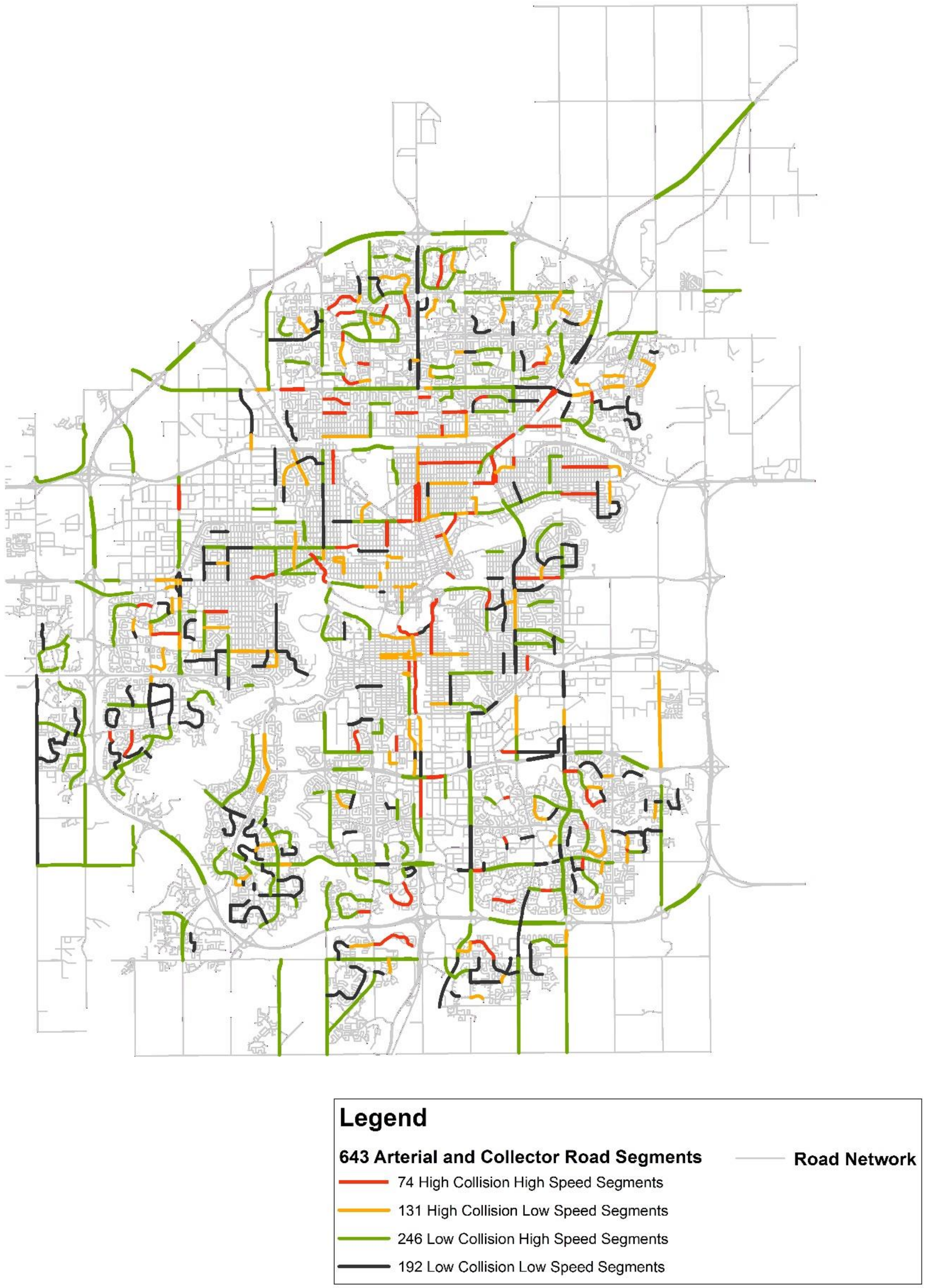

FIGURE 1: Site Distribution

TABLE 2: Regression Results

Group		B	Std. Error	Wald	Sig.	Exp(B)
HCLS	**Intercept**	2.632	3.465	0.577	0.448	
	Lanes	0.097	0.448	0.047	0.829	1.102
	Access Density	-0.762	0.501	2.314	0.128	0.467
	Operating Speed-PSL	-3.588	0.534	45.137	0	0.028
	Standard Deviation	2.555	0.46	30.881	0	12.877
	% Night Crashes	1.182	1.551	0.581	0.446	3.262
	% Trucks	38.882	17.328	5.035	0.025	7.69E+16
	Median*	-1.251	1.192	1.102	0.294	0.286
	Ped Crossing*	-0.257	0.766	0.113	0.737	0.773
	Shoulder*	2.897	1.625	3.181	0.075	18.123
	Road Class A	3.178	1.12	8.051	0.005	24
	Road Class C	0b	.	.	.	.
LCHS	**Intercept**	4.965	1.592	9.728	0.002	
	Lanes	-0.431	0.152	8.036	0.005	0.65
	Access Density	-0.732	0.217	11.327	0.001	0.481
	Operating Speed-PSL	-0.066	0.048	1.917	0.166	0.936
	Standard Deviation	0.065	0.137	0.228	0.633	1.067
	% Night Crashes	-1.099	0.667	2.714	0.099	0.333
	% Trucks	10.814	9.136	1.401	0.237	49728.12
	Median*	-1.353	0.497	7.407	0.006	0.259
	Ped Crossing*	-0.453	0.325	1.939	0.164	0.636
	Shoulder*	-1.236	0.708	3.049	0.081	0.291
	Road Class A	1.111	0.42	7.007	0.008	3.037
	Road Class C	0b				
LCLS	**Intercept**	6.321	3.357	3.544	0.06	
	Lanes	0.033	0.442	0.006	0.941	1.033
	Access Density	-0.977	0.495	3.895	0.048	0.376
	Operating Speed-PSL	-3.519	0.534	43.481	0	0.03
	Standard Deviation	2.45	0.456	28.885	0	11.592
	% Night Crashes	-1.763	1.604	1.208	0.272	0.172
	% Trucks	43.438	17.04	6.499	0.011	7.33E+18
	Median*	-2.073	1.177	3.099	0.078	0.126
	Ped Crossing*	-0.42	0.753	0.312	0.576	0.657
	Shoulder*	1.901	1.588	1.433	0.231	6.694
	Road Class A	2.079	1.096	3.6	0.058	7.998
	Road Class C	0b	.	.	.	.

a The reference category is: HCHS
* The reference category is the presence of the feature

Table 3: Factors influencing a road segment belonging to one bin over the other

Factor	Bin 1	Bin 2
	HCHS	**HCLS**
Road Class	Collectors	Arterial
Operating speeds	Higher	Lower
Speed Dispersion	Lower	Higher
%Heavy Vehicles	Lower	Higher
	HCLS	**LCLS**
Median Barrier	Absent	Present
Shoulder	Absent	Present
Operating speeds	Higher	Lower
% of Nighttime Crashes	Higher	Lower
	HCHS	**LCHS**
Number of Lanes	Higher	Lower
Access Density	Higher	Lower
Median Barrier	Absent	Present
Shoulder	Absent	Present
	HCHS	**LCLS**
Median Barrier	Absent	Present
Shoulder	Absent	Present