Investigating the Relationship between Speed Variability Measures and Road Safety: A Case Study of Residential Urban Streets in the City of Saskatoon

Hassan Hamad, M.Sc. student,
University of Saskatchewan

Emanuele Sacchi, Ph.D., P.Eng.,
University of Saskatchewan

Paper submitted for the Lowering Vehicle Speeds to Improve Road Safety session

2021 TAC Conference & Exhibition

Introduction

Traditionally, road safety has been measured by means of motor-vehicle collision events. Motor-vehicle collisions are, in fact, those events resulting in property damage, injury or fatality and can be related directly to safety (Garber and Hoel 2014). Collision events are commonly used to determine collision frequency defined as the count of collisions observed at a geographical space over a specified period of time (e.g., year) which is the most common indicator to analyze road safety (Lord and Mannering 2010). However, motor-vehicle collisions are rare events, and a prolonged observational period (e.g., several years) is usually required to collected them. Moreover, evaluating safety using collision events is a reactive approach (a countermeasure is implemented only after collisions have occurred). Therefore, the use of surrogate safety measures which are able to measure road safety in place of crash frequency, has become a key area of research (Boonsiripant 2009; Li et al. 2013; Tarko 2018).

Surrogate road safety measures have been gaining an increasing popularity in recent years in highway engineering. Surrogate safety measures provide an alternative approach to traditional collision-based measures. In addition, surrogate measures can solve the issue related to the reactive nature of collision data use in statistical analysis. Regarding speed-related variable use as potential surrogate safety measure, it is important to recall from physics that the kinetic energy of a body grows with the square of its speed. Therefore, the higher the speed during a collision, the higher the kinetic energy transferred from the bullet vehicle to the subjective vehicle, and higher the severity of a collision. Moreover, traveling at higher speeds results in longer perception-reaction distances and longer stopping distances. Finally, it is important to mention that also speed variability among vehicles is responsible for interactions among them (virtually interactions among vehicles would be impossible to occur without speed variability) and, therefore, the more the interactions the more the possibility of “unsafe” events, conflicts and ultimately collisions. Because of all these reasons, it is reasonable to assume that motor-vehicle travel speeds might be associated to road safety and a practical method for converting speed-related measurements into a corresponding collision frequency and/or severity can be developed.

Quantifying and establishing a comprehensive understanding of the relationship between speed and safety and the use of speed as a surrogate for safety evaluation is an important subject of research. Several research studies over the years investigated speed-related variables as surrogate measures of safety (Tarko 2018; Elvik, Christensen, and Amundsen 2004; Gargoum and El-Basyouny 2016; Taylor, Lynam, and Baruya 2002; X. Wang et al. 2018). However, more research is needed to understand the relationship between speed variability measures and safety. It is, in fact, important to understand whether vehicles travelling slower or faster than the mean speed of traffic are more often involved in collisions than vehicles travelling at a speed close to the mean speed. Despite several studies explored the relationship between the speed levels and safety, some studies have argued that speed variability variables (e.g., standard deviation, variance, and coefficient of variation) can be better predictors of safety (Taylor, Lynam, and Baruya 2000b; Lee, Saccomanno, and Hellinga 2002). Hence, the investigation of variables related to speed variability could potentially yield to new results in the field of surrogate safety measures (especially in low-speed urban environment).

Research Statement

In recent years, the City of Saskatoon has conducted several speed surveys within its neighborhood traffic reviews (NTRs) to gain an understanding of the traffic patterns at the neighborhood level opposed to the case-by-case analysis prior to that (“Neighbourhood Traffic Reviews” 2018). Around eight neighborhoods are selected each year for the analysis based on prioritization criteria that considers the residents' concerns, collision history, the stage of development and the age of the neighborhoods. Within this context, the use of speed-related variables as surrogate measures of safety, and in particular speed variability measures, could overcome the limitation of collision data availability or complement the use of collision data in road safety evaluations. However, a clear relationship between collision frequency and speed variability has to be demonstrated.

Therefore, the objective of this research work was to investigate the relationship between speed variability variables and predicted crash frequency with the aim of supporting the use of speed-related variables in road safety analysis. Three speed variability measures based on field free-flow speeds collected for 150 residential urban segments in the City of Saskatoon (Canada) were estimated and their association to collision frequency was explored using path analysis. Path analysis is a type of statistical technique used to model simultaneously different regression models where the association between speed variability and safety (collision frequency) can be mediated, moderated, or related directly to several different road and traffic factors. The analysis was conducted using Bayesian statistics.

Literature Review

It is part of the state-of-the-practice in road safety to use a safety performance function (SPF) to predict collision frequency at a road site (AASHTO 2010). In its canonical form, exposure factors like traffic volume in the form of Average Annual Daily Traffic (AADT), and the length of the road segment are the fundamental variables in SPFs. Collision counts are modeled using a negative binomial distribution to account for overdispersion property of collision data (Poch and Mannering 1996). Other site-related (roadway and roadside) characteristics that are expected to affect collision frequency are usually added to the canonical form of an SPF (e.g., number of lanes, access point and driveway density). Therefore, SPFs are developed to understand what roadway and traffic factors are related to collision frequency with the ultimate goal to make rationale decisions related to the planning, design, operation, and maintenance of roadway networks.

The effect of speed-related variables on collision frequency using an SPF has been investigated in several studies. A large body of research literature suggests that higher speed levels are associated to lower level of safety expressed as increased collision frequency (Elvik, Christensen, and Amundsen 2004; Taylor, Lynam, and Baruya 2002; C. Wang, Quddus, and Ison 2009). On the other hand, this positive relationship between speed-collision levels is challenged by other empirical evidence. For example, Taylor et al. (2000a) developed simple SPF regression models. The analysis of the data revealed that the average speed was negatively associated with crash frequency. The authors attributed this finding to the difference in road quality at the surveyed road segments; therefore, they created homogenous groups through which the effects of road quality on the relationship between collisions and speed could be captured. Further analysis of the data revealed that average speed was positively correlated with collisions while including a group variable. Therefore, the authors concluded that the speed-crash relationship is expected to be positive unless there is one or more unobserved factors influencing this relationship. Another study by Gargoum and El-Basyouny (2016) investigated the relationship between average speed, crash frequency and their predictors in the urban area. The modeling was conducted using path analysis where partial and full mediation effects of speed on crash frequency predictors were analyzed. The most important outcome from this study was the mediation effects of speed and its relationship with crash frequency which was found to be positive and statistically significant.

Regarding speed variability measures, in a study by Lee et al. (2002) traffic and environmental variables were modeled as factors affecting collision frequency. The authors used collision and traffic flow data extracted from loop detectors along a 10-km stretch of the Gardiner Expressway in the city of Toronto. Detectors were located at various locations upstream of on-ramps/off-ramps or lane transition zones (lane increase/lane reduction) and upstream of straight sections. The authors concluded that the variability in speeds within the lane and across lanes was positively associated to crash rates. In a subsequent study by Elvik et al. (2004), researchers used a theoretical model to demonstrate that speed variance would generate more conflict situations among vehicles. Higher conflict rates can translate to higher probability of collision events (i.e., lower safety). Two scenarios were compared: a scenario A where all vehicles travelled at the same speed level, leading to 32 conflict points and a scenario B where speed variation among the same number of vehicles was allowed. In this latter theoretical scenario, the number of conflicts grew to 42, which was about 30% more than scenario A. It was, therefore, theoretically demonstrated that high levels of speed variability contributed to higher conflict rate which can potentially relate to higher crash frequency.

Overall, these studies offered several insights to establish the effect of speed-related variables to safety. While higher speed levels seemed to be more associated to higher collision levels, more research is needed to demonstrate the theoretical results that speed variability variables (e.g., standard deviation, variance, and coefficient of variation) can be associated to higher collision frequency levels.

Data Collection and Analysis

In this work, speed data was collected at 140 locations in 7 neighborhoods - suburban development areas (SDAs) of the City of Saskatoon - as part of the city's NTRs. Speeds were measured between the months of April and October for the years 2017 and 2018 using pneumatic tubes placed at mid-segments. Twenty-four hours of data were recorded at all sites for an average of 6 days per location. In order to develop operating speed prediction models, speed measurements needed to reflect free-flow conditions. Therefore, all speed measurements for vehicles with headway less than 6 seconds were removed which is equivalent to volumes of less than 600 veh/h (Hassan and Sarhan 2011; Gargoum and El-Basyouny 2016). Nighttime and weekend measurements were also removed from speed data to eliminate biases due to visual impairment in low-light environments or unusual travel patterns/behavior in non-weekdays (Moses, Mtoi, and Ozguven 2014).

Raw speed data were aggregated in 15 speed bins (i.e., <10 km/h, 10-15 km/h, 15-20 km/h, 20-25 km/h, ..., >90 km/h), in addition to vehicle counts aggregated in 15-min intervals; speed data were extracted by plotting the speed curves given aggregated speeds and bin values. Speed measurements such as the average speed, standard deviation of speed, variance of speed along with the coefficient of variation of speeds were determined mathematically from the speed curves. The first variable measured was the average spot speed (time-mean speed, or TMS) for aggregated speed data. Eq. 1 shows the formula for calculating TMS where $\bar{v}_{Ti}$ is the average speed (TMS) per site 'i', 'g' is the number of speed groups (bins) where 'j' is the order of the speed group, f_j is the number of observations per speed group, v_j is the midpoint speed per speed group and 'N' is the total number of speed observations. Moreover, speed variance (σ_T^2) of TMS was calculated using Eq. 2. The standard deviation of speed (SD) corresponding to the calculated TMS can be obtained by calculating the square-root of the speed variance.

$$\bar{v}_{Ti} = \frac{1}{N}\sum_{j=1}^{g}\left(f_j\, v_j\right) \qquad \text{(Eq. 1)}$$

$$\sigma_{Ti}^{2} = \frac{\sum_{j=1}^{g} f_j (v_j)^2 - \frac{1}{N}\left(\sum_{j=1}^{g} (f_j\, v_j)\right)^2}{N-1} \qquad \text{(Eq. 2)}$$

Lastly, the coefficient of variation of speed was also measured. The coefficient of variation (CV) is the ratio of the standard deviation to the average (Eq. 3). It is a unitless value that is considered a standardized measure of variation. A large CV indicates a high speed variability in the sample, normalized by its mean speed, and a small CV indicates a low variability of speeds (AASHTO 2010).

$$CV_i = \sqrt{\sigma_{Ti}^{2}}/v_{Ti} \qquad \text{(Eq. 3)}$$

It is worth to mention that the average spot speed (TMS) is not representative of the average speed of vehicles along the length of the road segment. The average speed over the length of a road segment (i.e., space-mean speed, or SMS) is always lower than the TMS which yields higher traffic densities, thus more relevant to design-related applications. Speed measurements can be converted from TMS to SMS using Eq. 4 and the obtained measurements can be employed to derive the speed variation parameters (SD and CV); where, v_S is the space-mean speed (SMS), v_T is the average spot speed (TMS) and σ_T^2 is the variance of the TMS. The variance of the space-mean speed can then be calculated knowing the values of the TMS and the SMS as shown in Eq. 5.

$$v_S = v_T - \left(\sigma_T^{2}/v_T\right) \qquad \text{(Eq. 4)}$$

$$\sigma_S^2 = (v_T - v_S)\, v_S \qquad \text{(Eq. 5)}$$

Along with speed data, collision counts, traffic volumes and roadway/traffic features for the 140 locations were collected. The total number of crashes in 5 years (2014 to 2018) was selected as the temporal frame for this evaluation. This is because crashes are rare events that require prolonged observational periods to be employed for statistical inference. Moreover, traffic volumes were obtained from the speed data sheets (i.e., NTR data). Daily traffic can be estimated by summing all hourly traffic volumes in a day and, by considering all days in the data set, the average daily traffic (ADT) for each road site can be estimated. After, a monthly modification factor was applied to ADTs to determine the AADT for each site.

Furthermore, data describing road and traffic features of selected road segments were obtained from Google Street View and tools integrated into Google Earth. The survey made use of 2018 observations. Detailed description of these variables along with speed, collision and traffic volume variables used in the modeling is provided in Table 1 and the corresponding summary statistics in Table 2.

Table 1 Description of variables

Variable	Description
T5Y	The total number of crashes in 5 years (from 2014 to 2018)
SMS	Space-mean speed; the average (mean) speed of vehicles measured by taking the average of vehicles' speed measurements along the length of the traffic segment
SD	Standard deviation of speed; the measure of deviation of speeds from the mean

Variable	Description
CV	Coefficient of variation of speed; calculated based on the space-mean speed and its standard deviation
AADT	Average annual daily traffic volume (vehicles/day) on a road segment
L	Length of the road segment, measured between two traffic control devices interrupting the flow (signal, yield or stop sign)
TWW	Travelled-way width: the cross-sectional width of the paved surface (curb to curb)
Crv	Presence of horizontal curvature at the road segment; 0 = no curvature; 1 = curved road segment
Lane	Lane configuration; 0 = baseline condition (2-lane roadway); 1 = lane configuration otherwise
Med	Raised median presence; 0 = not present; 1 = present
CLP	Centerline presence: centerline marking dividing two opposing way of traffic. 0 = not present; 1 = present
OSP	On-street parking presence: 0 = not present on both sides; 1 = all other cases
BL	Bike lane presence: 0 = not present on both sides; 1 = all other cases
SW	Sidewalk presence: 1 = present on both sides; 0 = all other cases
AccPD	Density of access points: the count of stop or yield-controlled accesses along the road segment per unit of length; an access point has a minimum flow rate of 10 vehicles/hour
DW	Density of driveways: the count of accesses to properties along the road segment per unit length; a driveway has a maximum flow rate of 10 vehicles/hour
PedX	Pedestrian crossings presence: 0 = pedestrian crossings not present; 1 = pedestrian crossings present; pedestrian crossing points are only considered present when a sign or pavement marking is present.
BS	Bus stop presence: 0 = not present; 1 = at least 1 bus stop present
SchZn	School zone presence: accounts for the presence of a school zone sign indicating a lower speed limit (30 km/h). 0 = school zone not present; 1 = school zone is present

Table 2 Summary statistics of variables

Variable	Unit of Measure	Maximum	Minimum	Average	Std. Dev.
T5Y	Collisions/5 years	116.00	0.00	15.39	18.71
SMS	km/h	77.60	18.47	39.72	8.75
SD	km/h	10.78	4.75	7.83	1.17
CV	(Unitless)	0.41	0.12	0.21	0.05
AADT	Vehicles/day	18,499	153	3,594.25	3,915.72
L	m	1,863	100	702.03	412.33
TWW	m	25.00	6.75	12.91	3.04
Crv	0/1	1.00	0.00	0.43	0.50
Lane	0/1	1.00	0.00	0.10	0.30
Med	0/1	1.00	0.00	0.09	0.29
CLP	0/1	1.00	0.00	0.38	0.49
OSP	0/1	1.00	0.00	0.92	0.27
BL	0/1	1.00	0.00	0.06	0.25
S	0/1	1.00	0.00	0.89	0.32
AccPD	Access/km	18.13	0.00	7.43	4.46
DWD	Driveway/km	132.40	0.00	47.02	27.24
PedX	0/1	1.00	0.00	0.64	0.48
BS	0/1	1.00	0.00	0.51	0.50
SchZn	0/1	1.00	0.00	0.20	0.40

Methodology

Since collisions are discrete, random and non-negative events (i.e., count data), it is not possible to model collision frequency as response variable (Y) of a simple linear regression model. Following the literature, collisions can be modelled using a Poisson-Gamma (negative binomial) distribution, which accounts for the fact that collisions are count data and for their overdispersion (Lord and Mannering 2010). A Poisson-Gamma model for a set of crash data Y_{ik} at site i of type k can be written as:

$$Y_{ik} \sim \text{Poisson}(\lambda_{ik}), \quad i = 1, 2, 3, \dots, n \quad k = 1, 2, 3, \dots, j \qquad \text{(Eq. 6)}$$

where λ_{ik} is the expected mean of crashes of category k for the i-th site and can be modeled for road segments as:

$$\ln(\lambda_{ik}) = \beta_{0k} + \beta_{1k}\ln(AADT_i) + \beta_{2k}\ln(L_i) + \sum_{3}^{m}\beta_{mk}X_{im} + \varepsilon_{ik} \qquad \text{(Eq. 7)}$$

where β_{0k} is the intercept for type k, β_{1k} and β_{2k} are the respective coefficients of the average annual daily traffic (AADT) and the segment length (L), β_{mk} the coefficient of the m-th explanatory variable and category k, X_{im} the value of the m-th explanatory variable for the i-th site and category k, and ε_{ik} is a gamma distributed term with overdispersion parameter equal to κ.

An alternative to the Poisson-gamma model is the so-called Poisson Log-Normal (PLN) model where ε_{ik} is a normally distributed term with zero mean and σ as standard deviation. This latter modeling was the one employed in this study as recent research has showed that crash prediction models using the Poisson Log-Normal distribution result in better goodness of fit then Poisson-Gamma (Barua, El-Basyouny, and Islam 2016; El-Basyouny and Sayed 2009; X. Wang et al. 2015).

Equation 7 describes a standard SPF, which can be included and modeled simultaneously in a path analysis framework (see next sub-subsection). In this way, speed-related variables can be employed as a mediator that can predict collision frequency (Y) and can be predicted by site-related characteristics (road and traffic features). Bayesian statistics can be employed for SPFs to estimate model parameters (β). An alternative is to employ the generalized linear modeling approach and the maximum likelihood estimation for model parameters.

Path Analysis using Speed Variability as Mediator

The proposed methodology to model speed variability and collision frequency relationship is path analysis. Path analysis is a form of Structural Equation Modeling (SEM) where all variables are measured (i.e., observed variables). Path analysis is a technique that allows to test multiple relationships simultaneously. The use of path analysis modeling allows to mediate or moderate the relationship between speed and crash frequency by different roadway and traffic factors. Mediation analysis is, in fact, a type of path analysis which is used to understand how a variable x is related to another variable y. In other words, mediation is used to test whether the effect of x on y is (i) direct only, (ii) indirect only (through a mediator variable, m) or (iii) both direct and indirect. Case (ii) is known as full mediation, whereas the effect in case (iii) is considered to be partially mediated. The focus of this work is the mediation effect of speed-related variables and their effect on crash frequency. This can be summarized as illustrated in Figure 1 where traffic and roadway characteristics can predict crash frequency directly, indirectly through a speed-related variable or directly and indirectly simultaneously.

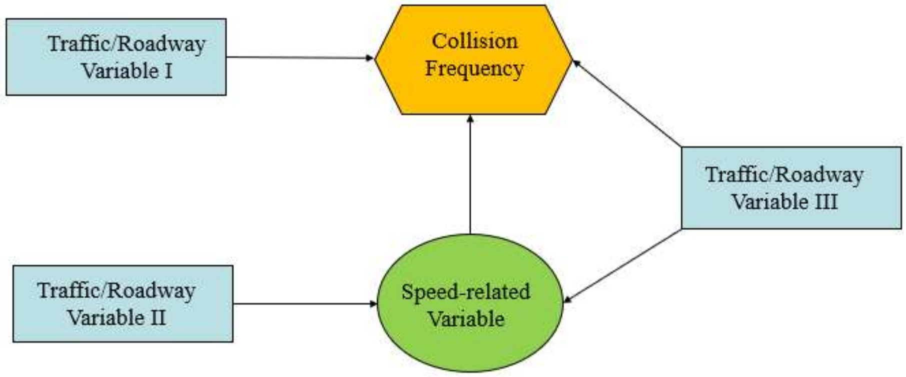

Figure 1 Path Analysis using a Speed-related variable as Mediator

In symbols, a simple mediation model depicts a model where x has an effect on y through a single mediator m. That is, in addition to the direct relationship between x and y, x is assumed to have an effect on m and this effect then propagates to y. Eqs. 8 and 9 are regression equations that represent a simple mediation model. Specifically, Eq. 8 represents the combination of the paths from m to y and x to y, and Eq. 9 represents the path from x to m (Gargoum and El-Basyouny 2016)

$$y_i = \beta_0 + \beta_1 m_i + \beta_2 x_i + \varepsilon_{1i} \quad \text{(Eq. 8)}$$

$$m_i = \gamma_0 + \gamma_1 x_i + \varepsilon_{2i} \quad \text{(Eq. 9)}$$

where, y_i denotes the outcome variable; m_i is the mediator variable; x_i represents the set of independent variables (exogenous); ε_{1i} and ε_{2i} are the errors associated with each of the components of the model structure; β_0 and γ_0 denote the intercepts of the models; and β_1, β_2 and γ_1 are all regression coefficients. It is worth to mention that in this study Eq. 8 was modeled as an SPF (see Eq. 7) and Eq. 9 was modeled as a multiple linear regression predicting the speed-related variable of interest. Figure 2 shows a diagram of the path analysis variables and parameters.

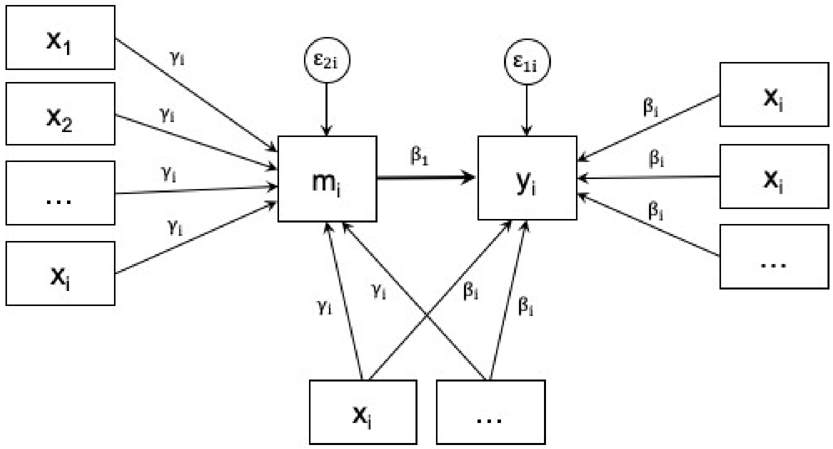

Figure 2 Path analysis variables and parameters

The coefficient γ_1 represents the magnitude of change in m associated with a unit change in x; similarly, the coefficient β_2 represents the magnitude of change in y associated with a unit change in x, which also denotes the direct effect of x on y. Moreover, the coefficient β_1 represents the magnitude of change in y associated with a unit change in m. To that end, the indirect effect of x on y can then be estimated using the product-of-coefficient estimator $\gamma_1 \beta_1$ (Hayes 2013).

Results

This section describes the results of the path analysis modeling using a speed variability variable as mediator to predict collision frequency for the 140 road segments described before. First, the selection of the mediating (speed-related) variable was investigated in the following way: four baseline SPFs with the same mathematical form of Eq. 7 were developed, where each potential speed-related variable was introduced as independent predictor (X) of crash frequency (λ). The speed-related variables selected for the preliminary modeling were three speed variability variables, i.e., the coefficient of variation of speed (CV), the standard deviation of speed (SD), the variation of speed (Var). SPFs were developed with SAS software, using the traditional negative binomial regression (GLM approach). All models showed an acceptable overall goodness of fit but the results showed that among speed-variability variables, CV was the most significant variable associated to collision frequency (significant at the 90% confidence level). In contrast SD and Var were found to be insignificant in their association to collision frequency. Therefore, CV was selected as the speed-related variable for this research.

Development of Path Analysis

Developing a path model (as presented in Figure 1 and 2) involved the selection of the independent variables that predicted CV, collision frequency or both, where the effect of some of these variables on collision frequency was mediated by CV. PLN distribution modeling technique was employed to develop the SPF component of the path model with T5Y as the 5-year collision frequency. Multiple linear regression was employed for the CV prediction model component. The best-fitting path model was obtained by eliminating variables from the initial set of variables (Table 2) that were not statistically significant at the 95% confidence level. However, multiple modeling trials suggested the inclusion of variables like AccPD (predicting CV) and BL (predicting T5Y) significant at the 90% confidence level. This allowed to retain in the model other variables significant at the 95% confidence level. WinBUGS 1.4.3 was employed for path analysis so that Bayesian estimates of model parameters could be obtained. Posterior estimations from the two path models were obtained via two chains with 20,000 iterations each, out of which 10,000 were excluded as burn-in samples. The trace plots for all model parameters were checked to monitor convergence. The deviance information criterion (DIC) was used for model selection.

Figure 3 provides a diagram of the resulting path model using the variables collected in this study. The resulting coefficients along with their standard deviation, confidence intervals are presented in **Table** 3. The standard deviation of the random error terms ε_{2i} and ε_{1i} in Eqs. 8 and 9 were denoted as $sigma_{CV}$ and $sigma_{SPF}$ indicating the standard deviations of error terms of the CV prediction model and the T5Y crash prediction model, respectively.

Table 3 Coefficient estimates of path analysis

Variable		Posterior Estimate	SD	Percentiles			
				2.50%	5.00%	95.00%	97.50%
$Intercept_{CV}$		0.344	0.028	0.290	0.299	0.390	0.398
SchZn	a1	0.042	0.010	0.023	0.026	0.059	0.063
Med	a2	-0.063	0.021	-0.103	-0.097	-0.029	-0.022
Lane	a3	0.080	0.026	0.029	0.037	0.122	0.130
AccPD	a4	-1.93E-03*	1.03E-03	-3.93E-03	-3.61E-03	-2.33E-04	1.06E-04
DWD	a5	3.00E-04	1.46E-04	1.36E-05	6.19E-05	5.40E-04	5.86E-04
TWW	a6	-0.009	0.002	-0.013	-0.012	-0.005	-0.004
L	a7	-3.18E-05	1.20E-05	-5.53E-05	-5.16E-05	-1.21E-05	-8.40E-06
BL	a8	-0.038	0.018	-0.073	-0.067	-0.009	-0.004
PedX	a9	-0.020	0.010	-0.041	-0.037	-0.004	-0.001
σ_{CV}		0.043	0.003	0.038	0.039	0.048	0.049
$Intercept_{SPF}$		-8.531	1.316	-11.130	-10.700	-6.403	-5.991
Ln(AADT)	b1	0.688	0.086	0.526	0.550	0.836	0.862
Ln(L)	b2	0.667	0.138	0.394	0.436	0.898	0.929
CV	b3	4.054	1.920	0.310	0.932	7.171	7.928
AccPD	b4	0.046	0.019	0.008	0.015	0.076	0.082
BL	b5	-0.566*	0.319	-1.209	-1.095	-0.037	0.058
σ_{SPF}		0.786	0.061	0.675	0.690	0.892	0.917
Total DIC		312.548					

* Parameter statistically significant at the 90% CL,
All other variables were statistically significant at 95% CL.

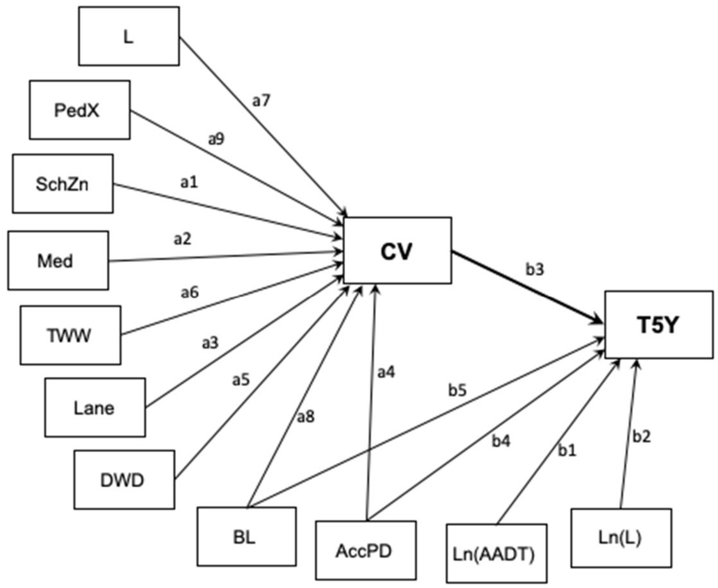

Figure 3 Resulting Path Model with CV as Mediator

With regard to factors predicting crash frequency (T5Y), the presence of bike lanes (BL) on roadway segments was found to be associated to safer conditions by contributing directly to the reduction of crash frequency; in fact, BL parameter estimate was found negative and equal to -0.5664 and significantly different from zero at the 90% CI. Conversely, higher access point density (AccPD; b4 = 0.0460) and higher speed variations (CV; b3 = 4.054) were associated to higher crash frequency (significant estimates at the 95% CI). Parameter estimates of exposure variables such as AADT (b1 = 0.6879) and L (b2 = 0.6671) were found positively associated to collision frequency which was expected as well. With regard to factors affecting CV, roadway segments in the proximity of school zones (SchZn), segments with lane configurations (Lane) different from the baseline case (two-way two-lane residential streets), and segments with higher density of driveways (DWD) were found to be associated to higher variability in speeds (CV) being the parameter estimates positive and equal to 0.0425, 0.0796 and 0.0003, respectively. On the other hand, longer roadway segments (L; a7 = -0.000032), segments with pedestrian crossing locations (PedX; a9 = -0.0205), median-divided segments (Med; a2 = -0.0626), wider cross-sections (TWW; -0.0086), segments with bike lanes (BL; a8 = -0.0379), and segments with more stop-controlled access points (AccPD; a4 = -0.00193) were associated to lower speed variability (CV). Finally, the presence of more access points was expected to be associate to higher speed variation, but the parameter estimate was significant only at the 90% confidence level. On the other hand, all other CV predictors were significant at the 95% confidence level.

While variables such as AADT, L, CV, AccPD and BL estimated the association to collision frequency directly, the effects of the predictors of CV were indirectly mediated either fully or partially through CV. Indirect effects of the predictors of CV on the likelihood of crashes were measured as the product of the coefficient associated to these predictors (a_i) and the coefficient estimate of CV in the crash prediction model component (b_3). In symbols, the indirect effect (i_i) can be estimated by the following equation:

$$\mathrm{i_i} = \mathrm{a_i} * \mathrm{b_3} \quad \text{(Eq. 10)}$$

Since the coefficient estimate of CV in the crash prediction model component was positively associated with the likelihood of crashes, the sign of CV predictors was directly carried over in the prediction of T5Y. This meant that variables that contributed to lower levels of speed variability in the population of drivers also contributed to lower levels of crashes, and vice versa. Moreover, the total effects (t_i) of variables that predict CV and T5Y simultaneously (partial mediation) can be estimated as the sum of their direct (b_i) and indirect effects (i_i) as shown in the following equation:

$$\mathrm{t_i} = \mathrm{b_i} + \mathrm{i_i}$$
$$\mathrm{t_i} = \mathrm{b_i} + (\mathrm{a_i} * \mathrm{b_3}) \quad \text{(Eq. 11)}$$

The summary statistics of the indirect and total parameter estimates is presented in Table 4.

Table 4 Indirect effects of path analysis

Variable		Posterior Estimate	SD	Percentiles			
				2.50%	5.00%	95.00%	97.50%
SchZn	i1	0.172	0.094	0.011	0.036	0.338	0.378
Med	i2	-0.254	0.153	-0.599	-0.533	-0.042	-0.010
Lane	i3	0.322	0.190	0.013	0.054	0.670	0.752
AccPD	i4	-7.81E-03*	5.95E-03	-2.16E-02	-1.88E-02	-6.90E-03	1.01E-03
DWD	i5	1.22E-03*	8.71E-04	-8.18E-05	5.90E-05	1.09E-03	3.27E-03
TWW	i6	-0.035	0.019	-0.077	-0.069	-0.033	-0.002
L	i7	-1.29E-04	8.12E-05	-3.15E-04	-2.78E-04	-1.18E-04	-1.43E-06
BL	i8	-0.153*	0.107	-0.404	-0.352	-0.011	0.006
PedX	i9	-0.083*	0.060	-0.224	-0.196	-0.004	0.005
Total Effects							
AccPD	t4	-0.010*	0.007	-0.025	-0.022	-0.001	7.12E-04
BL	t8	-0.191	0.119	-0.464	-0.412	-0.028	-0.004

* Parameter statistically significant at the 90% CL,
All other variables were statistically significant at 95% CL.

The indirect effects of the speed variation (CV) predictors on the collision frequency were significant at the 95% CI except for AccPD, DWD, BL and PedX, which were significant at the 90% CI. Bike lane presence contributed to lower speed variation and lower crash frequency also indirectly. The total effect of the bike lane presence is sum of the partial mediation through CV (a_8*b_3) and the parameter estimate of BL in the SPF (b_5). The direction of the total effect of BL on the crash frequency remained consistent (negative). The direct effect of AccPD (b_4) on the frequency of crashes was positive, however, this effect was outweighed by the partial mediation as the product of the parameter estimate of CV in the SPF (b_3) and AccPD estimate of the CV prediction (a_4) was a larger negative value compared to b_4. Hence, the total effect of AccPD on the frequency of crashes was negative meaning that the higher the frequency of access points on a road segment, the lower the frequency of crashes. However, this result was not statistically significant at the 95% confidence level.

CONCLUSIONS

This work aimed at investigating speed variability as a potential surrogate measure of safety able to mediate the relationship between collision frequency and road and traffic characteristics. The coefficient of variation (CV) of speeds was found to be the speed variability measure more significantly associated to crash frequency and, therefore, a path analysis model was built where CV was employed as a mediator variable which is predicted by roadway and traffic characteristics of the sites and can predict crash frequency. A path analysis framework can describe the direct and indirect dependencies among a set of variables. With path analysis, a relationship between an independent and a dependent variable can be direct or mediated by a third factor. The results demonstrated that CV was positively related to crash frequency, i.e., streets with higher speed variability showed lower safety levels, and the relationship was found statistically significant. Moreover, the interrelationship among roadway and traffic factors and crash frequency was analyzed, providing a better understanding of the indirect effect of independent predictors of CV on collision frequency. Overall, the results can be particularly important in the context of using speed-related variables as surrogate measures of safety, which would allow the assessment of safety levels of urban residential streets without waiting for collisions to occur.

Finally, future research should validate these results with different data set from different municipalities to account for local factors related to roadway, environmental and driver characteristics. Furthermore, more work needs to be conducted to explore more potential variables and relationships with regard to speed and safety modeling. The outcomes of the path model appear highly promising and, therefore, more combinations of variables could be explored to gain a better understanding on the effects of different speed-related variables and other independent variables on crash frequency. Although this study was aimed for to be as comprehensive as possible in the inclusion of different independent variables, there is room to explore different modeling structures.

REFERENCES

AASHTO. 2010. *Highway Safety Manual*. Vol. 1. Washington, D.C.: American Association of State Highway and Transportation Officials.

Barua, Sudip, Karim El-Basyouny, and Md Tazul Islam. 2016. “Multivariate Random Parameters Collision Count Data Models with Spatial Heterogeneity.” *Analytic Methods in Accident Research* 9: 1–15. https://doi.org/10.1016/j.amar.2015.11.002.

Boonsiripant, Saroch. 2009. “Speed Profile Variation as a Surrogate Measure of Road Safety Based on GPS-Equipped Vehicle Data.” *ProQuest Dissertations and Theses*, no. May: 298. https://vpn.utm.my/docview/304889106?accountid=41678.

El-Basyouny, Karim, and Tarek Sayed. 2009. “Accident Prediction Models with Random Corridor Parameters.” *Accident Analysis and Prevention* 41: 1118–23. https://doi.org/10.1016/j.aap.2009.06.025.

Elvik, Rune, Peter Christensen, and Astrid Amundsen. 2004. “Speed and Road Accidents: An Evaluation of the Power Model.” *TOI Report 740*. Vol. 740. http://www.trg.dk/elvik/740-2004.pdf.

Garber, Nicholas J, and Lester A Hoel. 2014. *Traffic and Highway Engineering*. Thomson-Engineering.

Gargoum, Suliman A., and Karim El-Basyouny. 2016. “Exploring the Association between Speed and Safety: A Path Analysis Approach.” *Accident Analysis and Prevention* 93: 32–40. https://doi.org/10.1016/j.aap.2016.04.029.

Hassan, Yasser, and Mohamed Sarhan. 2011. “Modeling Operating Speed.” *Modeling Operating Speed*. https://doi.org/10.17226/22864.

Hayes, Andrew F. 2013. *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach .* https://books.google.ca/books?hl=en&lr=&id=8ZM6DwAAQBAJ&oi=fnd&pg=PP1&ots=20Ef

nMYe-A&sig=58tZ79m0IAQbsdP00Ki19b_o-Nc&redir_esc=y#v=onepage&q&f=false.

Lee, Chris, Frank Saccomanno, and Bruce Hellinga. 2002. “Analysis of Crash Precursors on Instrumented Freeways.” In *Transportation Research Record*, 1–8. https://doi.org/10.3141/1784-01.

Li, Zhibin, Wei Wang, Ruoyun Chen, Pan Liu, and Chengcheng Xu. 2013. “Evaluation of the Impacts of Speed Variation on Freeway Traffic Collisions in Various Traffic States.” *Traffic Injury Prevention* 14 (8): 861–66. https://doi.org/10.1080/15389588.2013.775433.

Lord, Dominique, and Fred Mannering. 2010. “The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives.” *Transportation Research Part A: Policy and Practice* 44 (5): 291–305. https://doi.org/10.1016/j.tra.2010.02.001.

Moses, Ren, Enock T. Mtoi, and Eren E. Ozguven. 2014. “Evaluation of the 2010 Highway Capacity Manual Urban Street Free-Flow Speed Prediction Model.” *Transportation Research Record* 2461: 1–8. https://doi.org/10.3141/2461-01.

“Neighbourhood Traffic Reviews.” 2018. City of Saskatoon. 2018. https://www.saskatoon.ca/moving-around/driving-roadways/managing-traffic/traffic-studies/neighbourhood-traffic-reviews.

Poch, Mark, and Fred Mannering. 1996. “Negative Binomial Analysis of Intersection-Accident Frequencies.” *Journal of Transportation Engineering* 122 (2): 105–13. https://doi.org/10.1061/(ASCE)0733-947X(1996)122:2(105).

Tarko, Andrew P. 2018. “Surrogate Measures of Safety.” In *Transport and Sustainability*, 11:383–405. https://doi.org/10.1108/S2044-994120180000011019.

Taylor, M C, D A Lynam, and A Baruya. 2000a. “The Effects of Drivers’ Speed on the Frequency of Road Accidents.” *Transportation Research Laboratory*.

———. 2000b. “The Effects of Drivers’ Speed on the Frequency of Road Accidents Prepared for Road Safety Division, Department of the Environment, Transport and the Regions TRANSPORT RESEARCH LABORATORY.”

———. 2002. “The Relationship between Speed and Accidents on Rural Single-Carriageway Roads.” *Transportation Research Laboratory*. http://20splentyforus.org.uk/UsefulReports/TRLREports/trl421SpeedAccidents.pdf.

Wang, Chao, Mohammed Quddus, and Stephen Ison. 2009. “The Effects of Area-Wide Road Speed and Curvature on Traffic Casualties in England.” *Journal of Transport Geography* 17 (5): 385–95. https://doi.org/10.1016/j.jtrangeo.2008.06.003.

Wang, Xuesong, Tianxiang Fan, Ming Chen, Bing Deng, Bing Wu, and Paul Tremont. 2015. “Safety Modeling of Urban Arterials in Shanghai, China.” https://doi.org/10.1016/j.aap.2015.07.004.

Wang, Xuesong, Qingya Zhou, Mohammed Quddus, Tianxiang Fan, and Shou’en Fang. 2018. “Speed, Speed Variation and Crash Relationships for Urban Arterials.” *Accident Analysis and Prevention* 113 (April): 236–43. https://doi.org/10.1016/j.aap.2018.01.032.