Exploring the Association between Speed and Safety: A Path Analysis Approach

Suliman A. Gargoum[*a], Karim El-Basyouny[a1]

[a] Department of Civil and Environmental Engineering, University of Alberta, Edmonton, AB, Canada T6G 1H9
[*]Corresponding author. Tel.: +1 780-200-0161, Email: gargoum@ualberta.ca
[1]Tel: +1-780-492-9564, Email: basyouny@ualberta.ca

Abstract

Road safety is influenced by many factors; these factors include characteristics of the road, climate, traffic and, most importantly, vehicle speeds. Previous research shows that increases in speed are typically associated with an increased collision risk. Moreover, previous studies have also found relationships between road and traffic characteristics and collisions. In addition, these features have also been found to affect speeds. This paper aims to model all the aforementioned relationships simultaneously using a Structural Equation Modelling approach. More specifically, the paper attempts to model the relationship between average speed and collision frequency, while taking into account the effects of factors that confound the relationship. Moreover, the analysis attempts to assess the mediated effects that some variables have on collisions through their effects on speed. The data used in this study originated from 353 two-lane urban roads in the city of Edmonton, Canada. The average speeds were obtained from 35 million speed survey observations collected over a five-year period. The speed data are linked to the crash frequency at each location during the same time frame, along with the other factors (road, traffic and climate). The results show that, among others, average speed, volume, segment length, medians and horizontal curves all have statistically significant effects on collisions. On the other hand, shoulders, speed limits and vehicle-lengths are some variables that significantly influence speeds. The results also show that the effects of some variables on safety are indeed mediated through speeds (both partial and full mediation is observed). These findings provide valuable

insight that may assist decision makers in choosing and developing alternative speed management strategies, which, in turn, could help improve safety.

Keywords

Path Analysis; Structural Equation Modelling; Speed-Safety Relationship; Indirect Effects; Causal Mediation; Mediation Analysis.

Highlights

- SEM was used to model speed and crashes while accounting for other attributes.
- Direct and indirect relationships among all variables were analyzed.
- Many variables including average speed significantly affect collisions.
- The effects of some variables on safety are mediated through speeds.
- Outcomes reveal the importance of considering indirect effects when managing speeds.

1. Introduction

Speeding is an issue on roads worldwide, and its impacts on road safety are well documented in the safety literature (NHTSA, 2012). Speeding can be defined as simply exceeding posted speed limits or driving at speeds deemed too high for road or climate conditions during a certain period or at a certain location (e.g., foggy conditions or icy roads). In fact, the role that the road environment plays in a driver's decision to violate speed limits has been extensively researched in recent years (OECD/ECMT, 2006, Goldenbeld and van Schagen, 2007, Yannis *et al.*, 2013).

Whether it is driving at speeds higher than the speed limit or inappropriate for given conditions, speeding is a common factor contributing to collisions; in addition, speed-related collisions also represent a major share of severe collisions. Statistics show that, in 2012, 30% of all fatal collisions observed on roads in the U.S. were speed-related crashes (NHTSA, 2012); furthermore, statistics also show that, although the number of fatal collisions has dropped over the years, the portion attributed to speed-related collisions has not. Despite the statistics showing that speed is a common factor in collisions of all severity levels, the relationship remains one of the most debated topics in the field of traffic safety (Hauer, 2009).

Speeding is known to affect both the severity of a collision (Nilsson, 1982) and the likelihood of being involved in a collision (Aarts and Van Schagen, 2006). Although both relationships can be validated by means of basic laws of physics, research has shown that modelling the latter relationship is extremely complex. Common sense implies that the higher the speed, the less time (perception-reaction time) the driver has to react to an issue that could suddenly arise on the road and, thus, the higher the risk of not taking appropriate action.

However, in reality, modelling the relationship is not as straightforward as it might seem (Elvik *et al.*, 2004, Hauer, 2009).

Previous studies have attributed this complexity to the fact that collisions are random events that cannot be accredited to a single factor. This means that even though there is a possible high correlation between speed and collisions, it is impossible to ignore the other variables that affect collisions when modelling the relationship. Moreover, some of those variables could have masking effects, which conceal the relationship of interest (Elvik *et al.*, 2004).

In an attempt to classify the different types of variables influencing road accidents, Elvik *et al.* (2004), developed the causal diagram shown in Figure 1. As seen in the figure, the variables affecting collisions (i.e., the dependent variable) were divided into four different categories.

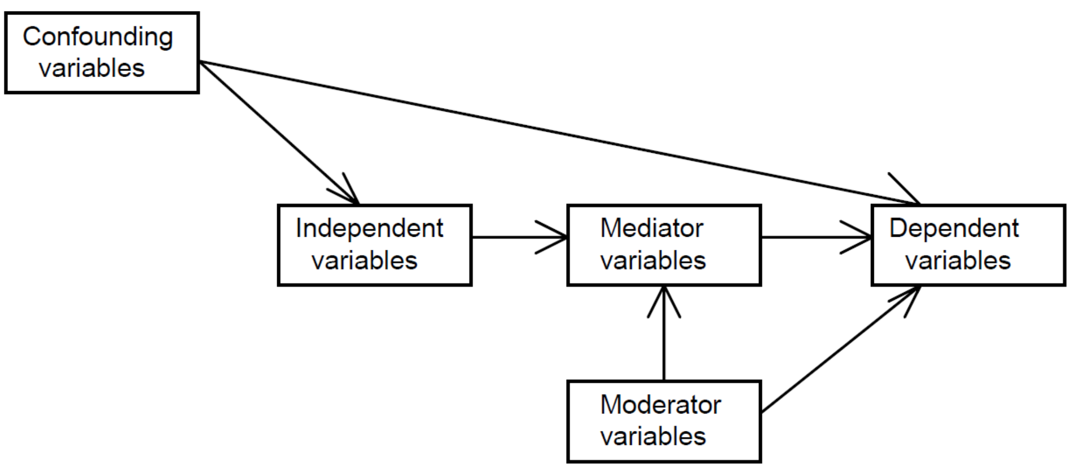

FIGURE 1 Main Categories of Variables Influencing Road Accidents (Elvik et al., 2004).

The dependent variable (DV) is the outcome variable to be predicted; in traffic safety, this is often the expected number of accidents. The mediator variable is a risk factor, which, if modified, can influence the dependent variable. It is also a variable through which other factors could impact the DV. The primary mediator when analyzing safety, according to Elvik *et al.* (2004), is changes in speeds. The independent variables (IV) are defined as the variables introduced to affect the DV through their influence on the mediator variable; examples of these variables include speed enforcement or other speed calming techniques. Moderating variables are seen as risk factors, which may include factors a safety measure does not intend to influence. One example is a road's surface condition (i.e., friction) moderating the effects of speed on collisions; the effects of speed on collisions could be greater on smooth surfaces than rough surfaces.

Confounding variables are the most challenging variables to account for; these variables, in addition to speed, have some influence on collision counts. Examples of these variables include different roadside features, elements of the road geometry, climate factors and temporal factors. Previous studies have found that these variables not only affect collision counts at a certain location but also the speeds (O'Flaherty and Coombe, 1971, Giles, 2004). As a result, accounting for those factors is essential when assessing the speed-collision relationship.

The diagram in Figure 1 helps create a rational framework for researchers attempting to model the relationship. Nevertheless, previous studies have assessed only individual paths of the diagram; thus, to the best of our knowledge, the diagram as a whole has not been validated using empirical data from any region. Therefore, the existence of those direct and indirect relationships illustrated in the path diagram remains entirely theoretical. This paper aims to address the aforementioned gap by performing a data analysis of two-lane urban roads in the city of Edmonton. Slight modifications to the diagram are applied and the path analysis is used to statistically test for the existence of the proposed relationships.

In addition to understanding the relationships among speed, safety and road environment, providing evidence of the existence of mediated effects between road geometry and collisions could also help authorities adopt a different approach when addressing speeding problems at collision hotspots. This matter is discussed in more detail in the results section.

2. Previous Work

Modelling the relationship between speed and the likelihood of a collision on a road section has often followed one of two approaches: cross-sectional analysis or longitudinal analysis. In cross-sectional analysis, data from several different locations are used to relate speed to collisions; whereas, in longitudinal analysis, time series collision data are analyzed. From a speed safety perspective, the majority of longitudinal studies involve before-after (BA) analysis of collision data where changes in speed are related to changes in collisions at specific locations.

Different statistical modelling techniques have also been considered in previous research; likewise, different measures of speed (e.g., average speed, speed variance) have all been modelled when analyzing the relationship. Despite the variation in techniques and analysis criteria, not all previous work has been able to reach perfectly consistent conclusions regarding the relationship between speed and safety. In studies where such a relationship was not concluded, researchers often cite confounding factors masking the relationship between speed and collisions as a possible reason; these factors are typically characteristics of the road that define its quality (Taylor *et al.*, 2002). In the next few paragraphs, a review of relevant work is presented.

Taylor *et al.* (2002), developed a Poisson model to assess the effects of several variables, including speeds, on injury crashes observed at rural single carriage roads across the U.K. Data used in the study included traffic flows, spot speeds and information about road geometry and other features. Preliminary analysis of the data revealed that, for the compiled dataset, average speed was negatively related to accident frequency. The authors attributed this finding to the difference in road quality at the road segments sampled; therefore, they created homogenous groups through which the effects of road quality on the relationship between collisions and speed could be captured. Indeed, further analysis of the data, while including a group variable, revealed that average speed positively correlated with collisions.

The reasoning behind such a phenomenon, as expressed by Elvik *et al.* (2004), is that the "best roads tend to have the highest speed limits," with the best roads being those with the highest quality attributes. In fact, even in the study by Taylor *et al.* (2002), the variables on which the grouping was based were all attributes of road quality. In order to further elaborate on the concept, Elvik *et al.* (2004) used hypothetical data to compare the differences between running a bivariate analysis on all road groups (Figure 2a) and running separate analyses for the different groups (Figure 2b).

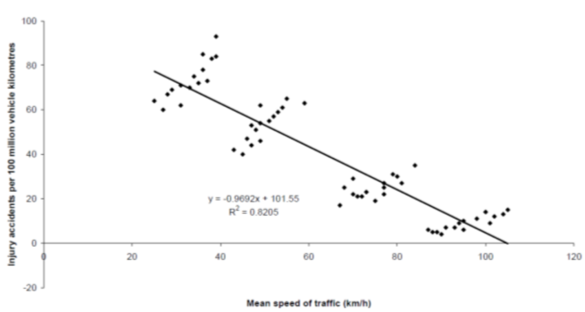

FIGURE 2a: Bivariate Analysis of Speed and Safety (Elvik et al., 2004)

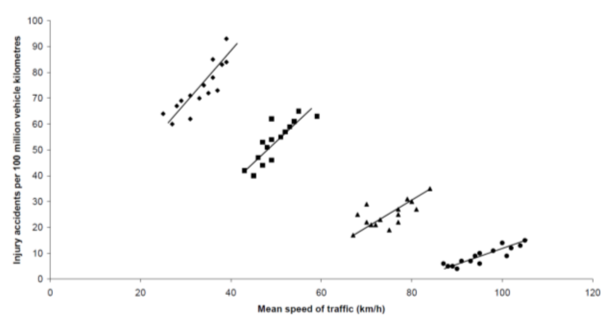

FIGURE 2b: Separate Analysis of Speed and Safety for Different Groups (Elvik et al., 2004)

Despite the statistical validity in creating the homogenous groups, the models developed in (Taylor *et al.*, 2002) were subject to some criticism. The reason was that the classification of the data into groups was based on variables that were also considered as explanatory variables in the model, which was seen as a limitation of the work by some researchers (Wang *et al.*, 2013). For instance, traffic flow was a variable used to classify data into groups, but both flow and the grouping ID were used in the model, indicating a potential duplicate effect.

In an earlier study, Taylor *et al.* (2000) studied the effects of speed on urban classified roads. Data were collected from 300 different road sections and linked with 1590 injury crashes. In this case, statistical cluster analysis was used to classify 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. The findings of the analysis revealed that, among other variables, average speeds were strongly related to accident frequencies. The relationships between average speed and frequency were positive in all four groups, wherein, higher average speed was associated with more accidents. A 1mph reduction in speed was associated with a 3-6% decrease in accidents depending on the road group.

In a before-after study assessing the impacts of speed limit changes on average speed and crash rates, Baruya and Finch (1994) used data from four countries to develop three models (simple linear model, power model and asymptotic function). The study found that average speeds were positively correlated with crash rates, with a speed reduction of 1 mph typically corresponding to a 5% decrease in crash rates. Nevertheless, the authors realize the limitations of their study, including the failure to account for national differences. In another before-after study, Islam and El-Basyouny (2015) used a full Bayesian approach to assess the effects of reducing speed limits in urban areas on safety. The authors found that reductions in speeds were associated with reductions in crashes of all severity levels.

One of the early studies that found a negative relationship between average speed and collisions was Baruya (1998). A multiplicative Poisson model was developed using data from three European countries. The findings revealed that traffic flow was the primary predictor of collisions; moreover, the results also indicated that locations with a higher Posted Speed Limit (PSL) are associated with a higher crash frequency. In contrast, higher average speed was found to result in lower collisions. Aarts and Van Schagen (2006) attributed Baruya's findings to three reasons, which were seen as limitations of the study: (1) interactions between the variables considered in Baruya's model, (2) existence of national differences, and (3) the time frame for flow data (24-hour) and speed data (off-peak periods) not matching.

Using two years of data from high-speed roads in the U.S., Lave (1985) fitted 12 different regression models to understand the effects of speed on fatalities. Average speed was found to have a statistically insignificant and negative effect on fatalities in almost all models, while speed variance was found to have positive and statistically significant effects on the fatality rate, compelling the author to conclude that it is variance, not speed, that kills.

It is worth mentioning that Solomon (1964) also reached the conclusion that vehicles travelling significantly faster/slower (+/- 30 mph) than the modus speed on a road had higher crash rates than vehicles travelling within a 6 mph margin. This conclusion was illustrated through the renowned U-curve. Nevertheless, the RTI (Research Triangle Institute (1970)) found that the U-curve was actually due to the inclusion of maneuver crashes that are not typically considered as speed-related crashes, and so this was seen as a limitation of the study. Just like Solomon (1964), Lave (1985) was also subject to criticism by Levy and Asch (1989), who found a positive and statistically significant relationship between average speed and fatalities while conducting an assessment similar to that of Lave. Lave's work was subject to further criticism by Davis (2002), who attributed Lave's findings to ecological fallacy due to the failure to distinguish between individual and aggregated measures of risk.

In a before-after analysis, Kweon and Kockelman (2005) used a random-effects negative binomial (NB) model among several panel and non-panel models to evaluate the effects of changing the speed limits on major roads in Washington State on crashes of different severities. Four years of data from 63,937 roadway segments were used. The results indicated that speed limits had statistically insignificant effects on fatal crashes. The failure to account for road features, such as presence of driveways and interchanges, and weather conditions is seen as a possible limitation of their analysis.

In a study analyzing the effects of point-to-point (P2P) speed enforcement systems, Montella *et al.* (2015) used a before-after analysis to identify how the enforcement program affected speeds and, in turn, impacted safety. The fact that the study found that the enforcement system affects both speed and safety led the authors to develop a crash modification function for the enforcement strategy. The function was used to relate changes in speed parameters (i.e., mean speed, 85th percentile speeds, and standard deviations of speed) to crash reduction. The relationship between speed measures and collisions in the function was based on the power model. The authors found that decrease in mean speed, 85th percentile speeds, and standard deviations of speed were all associated with a crash reduction; however, the authors concede that this reduction shrunk over time, which was attributed to enforcement management.

Using data from the Greater London region, U.K., Quddus (2013) developed both a random-effects NB and a mixed-effects spatial model to understand the effects of speeds on slight injury and killed-seriously-injured collisions. In the later model, the aim was to account for potential spatial correlation between neighbouring segments. Both models indicated a negative, yet statistically insignificant, relationship between average speed and collisions. Speed variation, however, was found to have a positive and statistically significant association with collisions.

The authors concede that their failure to find a significant relationship between average speed and safety was a limitation of their work; this was attributed to the failure to account for segment characteristics, such as speed difference between lanes (related to the occurrence of overtaking maneuvers) and within lanes (related to the occurrence of rear-end collisions) and direction of the vertical grade (as speeding occurs at downgrades).

In a thorough review of the most significant empirical studies that addressed the speed-safety relationship, Aarts and Van Schagen (2006) concluded that crash rates increase with the speed. Moreover, the paper also concluded that higher speed variance is associated with higher crash rates.

Although the existence of a relationship between speed and safety is indisputable, the findings of previous research remain inconsistent. While some studies have found that higher speeds increase the likelihood of collisions, other studies have found the opposite, stating that higher speeds are associated with a lower probability of collisions, with the majority of these studies also citing speed variance as the main risk factor in the speed-safety relationship.

3. Data Description

The data used in this paper were collected from 353 different two-lane urban road segments across the city of Edmonton, covering a five-year period, 2009-2013. The average speed at each location was computed using a total of 35 million individual vehicle speed records. These records were obtained through speed surveys conducted at the sites of interest during the five-year period. The instrument used by the City for data collection is the Vaisala Nu-Metrics Portable Traffic Analyzer NC200. These devices have built-in sensors through which they detect, count and classify vehicles, and measure their speeds. The City of Edmonton speed survey data collection could extend up to a week, during which the devices are placed at a certain location and data is collected 24 hours a day.

The average speed data were linked to the crash frequency at each location during the same time frame (i.e. collisions observed during the period 2009-2013 at those locations), along with the traffic and road characteristics of interest and climate information.

In order to ensure speed modelling was performed on free flow speeds, it was essential to exclude vehicles experiencing traffic congestion from the dataset before calculating the average speeds at each location. Since the majority of existing studies considered vehicles with a headway of five seconds or less as not operating under free flow conditions (Hassan and Sarhan, 2011), this threshold was used in filtering the data. This meant that only vehicles keeping a headway of six seconds or more were kept in the dataset.

3.1 Predictor Variables

Several variables were investigated as potential predictors; selection of the variables was based on the size of their perceived effects on speed and collisions. This was mainly determined by observing the outcomes from previous studies. A statistical summary of all variables is found in Table 1.

TABLE 1 Sumary Statistics of Data

Variable	Minimum	Maximum	Mean	Std. Deviation
Vehicle Length*(m)	4.26	8.43	5.38	0.45
Temperature*(°C)	-7.53	22.84	11.84	6.76
Wind Speed*(kph)	7.78	32.25	15.36	3.96
Visibility*(km)	8.95	25.46	20.64	5.10
Median	0.00	1.00	0.02	0.15
Horizontal Alignment (HA)	0.00	1.00	0.44	0.50
Bus Stop	0.00	1.00	0.38	0.49
Posted Speed Limit (kph)	40.00	90.00	51.02	5.58
Shoulders	0.00	1.00	0.03	0.17
Collision Frequency	1.00	206.00	9.25	18.29
Ln (Volume)	1.10	6.94	3.98	1.07
Average Speed (kph)	26.34	77.86	46.64	7.80
Segment Length (m)	63.69	2100.71	588.44	373.01
Standard Deviation of Speed	6.17	16.71	11.32	1.57
Percentage of Speed Limit Violators	0.01	0.99	0.66	0.22
Pedestrian Crossing	0.00	1.00	0.08	0.28
Bike Lanes	0.00	1.00	0.03	0.18

*Average Observed Per Location

3.2 Traffic Characteristics

3.2.1 Average Speed and Standard Deviation of Speed

In addition to average speed, which is considered both an independent and dependent variable, the standard deviation of speeds was also considered as an IV when modelling collisions. The standard deviation is expected to account for the potential effects of speed variation on collision frequency. Standard deviation was present in preceding work as well, and was found to have significant effects on collisions (Quddus, 2013).

3.2.2 Traffic Flow

Traffic flow has been a primary predictor when estimating collisions at a certain location (Baruya, 1998, Abdel-Aty and Radwan, 2000, Quddus, 2013) In fact, Aarts and Van Schagen

(2006) mentioned traffic flow as one of the most common factors which confound the relationship between speed and safety. Unfortunately, traffic volume (Average Annual Daily Traffic) data were not readily available for locations considered in this study since the City of Edmonton does not collect the data for all urban roads. Hence, traffic flow was manually estimated by collecting vehicle counts at the speed survey locations continuously for several days during the data collection period. The data were then used to compute the average flow per hour at the locations of interest. The variable representing traffic flow was log transformed to ensure that the model does not produce any estimates when traffic volume is zero.

3.2.3 Vehicle Length and Percentage of Speed Limit Violators

Traffic composition is also an important factor that was included in the models. The average vehicle length at each location was used as an IV with potential effects on both speed and collisions. In work modelling collisions, similar variables such as the percentage of trucks have been commonly present (Taylor *et al.*, 2000, Hauer *et al.*, 2004, Anastasopoulos *et al.*, 2008). Another factor representing traffic composition, which has been frequently used as a predictor of collisions in the past, is the percentage of speed limit violators (i.e., drivers operating their vehicles at speeds higher than the speed limit) (Baruya, 1998, Taylor *et al.*, 2000, Taylor *et al.*, 2002). Hence, this factor was considered in the models developed in this paper.

3.3 Road Characteristics

3.3.1 Segment Length

An extremely significant variable when modelling collisions on a road segment is the length of the segment. Along with traffic volume, segment length is an integral part of any collision prediction model; thus, it has been included in the majority of models developed in the past (Baruya, 1998, Kweon and Kockelman, 2005, Quddus, 2013). In this study, segment lengths were extracted from GIS maps kept by the City's Office of Traffic Safety (OTS). In addition to affecting collisions, traffic volume and segment length were also assumed to affect speeds.

3.3.2 Posted Speed Limits (PSL)

Another important road feature affecting speeds is the posted speed limit (PSL) on the road; as expected, the PSL is believed to be the most important predictor of speeds at a road segment (Fitzpatrick *et al.*, 2001). Hence, it was vital to include it as a variable in the analysis.

3.3.3 Other Geometric Attributes

Other road characteristics considered in the analysis were mainly represented using binary variables, where "1" and "0" represented the presence and absence of that attribute respectively. One important variable, which was investigated in this paper, was the horizontal alignment (HA) of the road segment, where 1 denotes curved and 0 denotes tangential road. Horizontal alignments and their geometric attributes have been included in most studies that have previously modelled speeds or collisions (Abdel-Aty and Radwan, 2000, Haynes *et al.*, 2007, Anastasopoulos *et al.*, 2012). Whether the two directions of the road segment were physically separated by a median barrier or not was another predictor variable in the model; this variable was found to have statistically significant effects on safety in previous work (Anastasopoulos *et al.*, 2008).

The presence or absence of shoulders on the side of the road (i.e., lateral clearance on the edge of the road for emergency usage), bus stops, bike lanes, and pedestrian crossing facilities was also analyzed in this study. As in the case of horizontal curves, past work shows a statistically significant association between shoulder lanes and collisions (Milton and Mannering, 1998, Abdel-Aty and Radwan, 2000). It is worth noting that the shoulders considered in this study were lateral shoulders available for emergency usage and not for travel.

All features mentioned in the previous paragraph were collected using Google Street View. Views from different years were compared to ensure that the features of interest did not experience significant change throughout the analysis period considered.

It is worth mentioning here that data on other variables, apart from those discussed, were collected. However, in order to ensure statistical power of the developed model, some variables had to be dropped from the analysis. This was done for one of two reasons: (i) Variables being under-represented by the dataset; for instance, a vertical grade was present only at two locations of those considered; (ii) Presence of high correlation between some variables and the main variables of interest (speed and collisions). Roadside parking was highly correlated with average speeds and, hence, was dropped from the models. Similarly, the number of lanes was also found to have a high correlation with average speeds; as a result, the analysis was performed only on two-lane urban roads.

3.4 Climate Information

Climate conditions were represented using the average temperature, wind speed and visibility at each location. Visibility is defined by Climate Canada as “the distance at which objects of suitable size can be seen and identified”. Weather data were collected for each speed survey observation, and then averages of the variables were computed for each site. The weather conditions at the date and time of each speed survey data point were matched with hourly weather records for the City of Edmonton, maintained by Environment Canada. As in the case of the other variables, climate conditions have been found to influence speeds and safety in previous research (Giles, 2004, Theofilatos and Yannis, 2014).

3.5 Hypothesis

The aim of this paper was to study the speed-safety relationship while accounting for factors mediating and confounding the relationship. The mediator was assumed to be the average speed on a road segment; whereas, confounding factors included traffic characteristics, road characteristics and climate conditions. A path diagram illustrating the hypothesized relationship can be seen in Figure 3a.

Based on outcomes of previous research and expert knowledge, some variables were assumed to have direct effects on collisions only, speed only or both speed and collisions. Furthermore, variables with effects on speed were later tested for indirect effects on collisions.

4. Methodology

The methodology used in this paper was path analysis. Path analysis is a form of Structural Equation Modelling (SEM) where all the variables are manifest variables (i.e., measureable). Structural Equation Modelling was chosen for two major reasons. First, SEM is a technique through which multiple relationships can be tested simultaneously, some of these relationships can be mediated or moderated. As discussed in the introduction and as highlighted in the work by Elvik et al., (2004), the speed-safety relationship is confounded, mediated and moderated by several different factors, and in order to properly model the relationship (speed-safety) all those factors must be accounted for or controlled in the models. Second, SEM enables testing for the existence of mediated effects. Understanding these effects could greatly impact the way responsible authorities manage speeds in their attempts to improve safety.

4.1 Structural Equation Modelling and Mediation

SEM is a series of statistical methods that enable the analysis of the relationships between a number of dependent variables (DV) and a set of independent variables (IV). Often used for confirmatory purposes, SEM is a technique where the main aim of the analysis is to test the validity of a certain relationship. When dealing with latent variables, performing the analysis usually includes a combination of confirmatory factor analysis and path analysis (Bollen, 2014).

Different types of variables can be considered in SEM. The IV can either be manifest (measurable/observed) or latent (unmeasurable/unobserved). Moreover, variables in a model can also be either exogenous (not influenced by any other variable in the model) or endogenous (influenced by another variable in the model). When variables in the model are all manifest, SEM simplifies the analysis to a path analysis, in which mediation, moderation, mediated moderation or moderated mediation can all be tested (Hayes, 2013).

In addition to the many advantages of SEM, the technique also makes certain assumptions that could be challenging to meet. It is typically required that SEM is only performed on a large sample size (n>200) (Kline, 2015); fortunately, this assumption is satisfied in the analysis performed in this paper. As already noted, SEM is a confirmatory analysis technique; this implies that model specification must be done a priori, presenting another potential challenge when dealing with relationships where it unknown which variable affects the other. Unlike other statistical tools, model estimation in the SEM framework involves modelling the covariance matrix of the observed variables as opposed to the observations themselves, due to this residual analysis being significantly different in SEM when compared with Generalized Linear Models.

Mediation analysis is a statistical method used to understand how a variable *x* transmits its effects 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, in case (iii) the effect is considered to be partially mediated.

Even though the technique has been often associated with causal inferences (Pearl, 2010), it is important to note SEM mediation analysis cannot be used to prove causality (Sobel, 2008). The main use of such techniques is to test a relationship, which is proposed based on a theoretical background, logical assumptions or research design. The analysis tests the existence and the magnitude of a relationship but, what is not possible, is to statistically identify the direction of the relationship (i.e., whether *x* affects *y* through *m*, or *y* affects *x* through *m*).

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*. Equations 1 and 2 are regression equations that represent a simple mediation model. Specifically, equation 1 represents the combination of the paths from *m* to *y* and *x* to *y*, and equation 2 represents the path from *x* to *m*.

$$y_i = \beta_0 + \beta_1 m_i + \beta_2 x_i + \varepsilon_{1i} \qquad (1)$$

$$m_i = \gamma_0 + \gamma_1 x_i + \varepsilon_{2i} \qquad (2)$$

Where, y_i denotes the outcome variable; m_i is the mediator variable; x_i represents all the IVs (exogenous); ε_{1i} and ε_{2i} are the errors; β_0 and γ_0 denote the intercepts of the models; and β_1, β_2 and γ_1 are all regression coefficients.

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)*.

4.2 Model Development

Since collisions are discrete, random and non-negative events, it is not possible to model the relationship among the IVs, including average speed and collisions, using the simple linear regression model shown in equation 1. Thus, collisions were modelled using a negative binomial (Poisson-Gamma) distribution, which also accounted for the overdispersion commonly observed in collision data (Lord and Mannering, 2010).

If Y_i were to denote the number of crashes at site *i*, then

$$Y_i \sim Poisson(\lambda_i) \tag{3}$$

It is also assumed that the number of accidents at all sites *n* is independent. Moreover, in order to account for overdispersion for unobserved heterogeneity, it is assumed that

$$\lambda_i \sim Gamma(\kappa, \kappa / \mu_i) \tag{4}$$

Where, κ represents the inverse dispersion parameter and μ_i is the predicted number of accidents, which is given by

$$\mu_i = \gamma_0 \theta_i^{\gamma_1} \tag{5}$$

This can also be written as

$$\ln(\mu_i) = \ln(\gamma_0) + \gamma_1 \ln(\theta_i) \tag{6}$$

Where, γ_0 and γ_1 represent the parameters of the model and θ_i represents the model covariates. The expected value and the variance of the negative binomial model can then be expressed as follows:

$$E(Y_i) = \mu_i, \;\; Var(Y_i) = \mu_i + \mu_i^2 / \kappa \tag{7}$$

The relationship between the exogenous variables and collision frequency and the relationship between those variables and the mediator were simultaneously modelled using SEM software Mplus version 6. The latter relationship was modelled using Ordinary Least Squares (OLS) regression, as presented in equation 2. Maximum likelihood estimation with robust standard errors was used in computing the model parameters.

After running multiple models, statistically insignificant paths were dropped. Moreover, the comparison of multiple models was also based on the Akaike Information Criterion (AIC) minimization criteria, where the model with the lowest AIC was assumed to have the best fit. To further verify the model fit, the AIC of the fitted model was compared to the baseline model (i.e. without predictor variables. Both the AIC (Baseline = 4762.9, Fitted = 4513.3) and Bayesian Information Criterion BIC (Baseline =4786.1, Fitted = 4575.2) were lower for the model used, indicating that the full model fits the data better. The final path diagram adopted is displayed in Figure 3b.

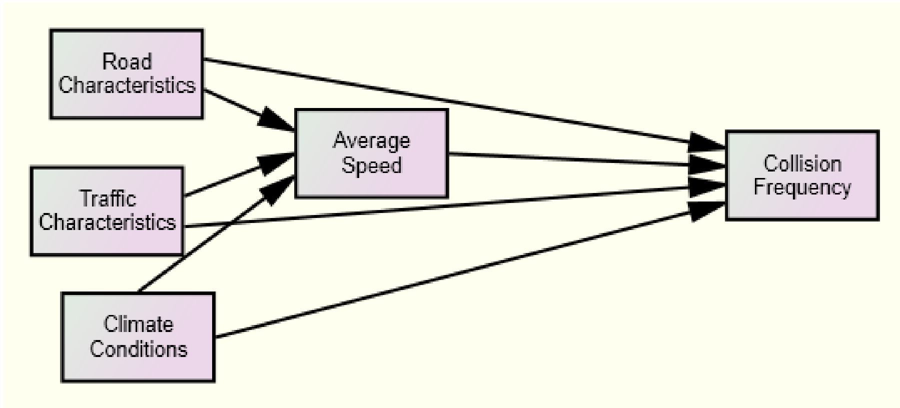

FIGURE 3a: Proposed Path Model

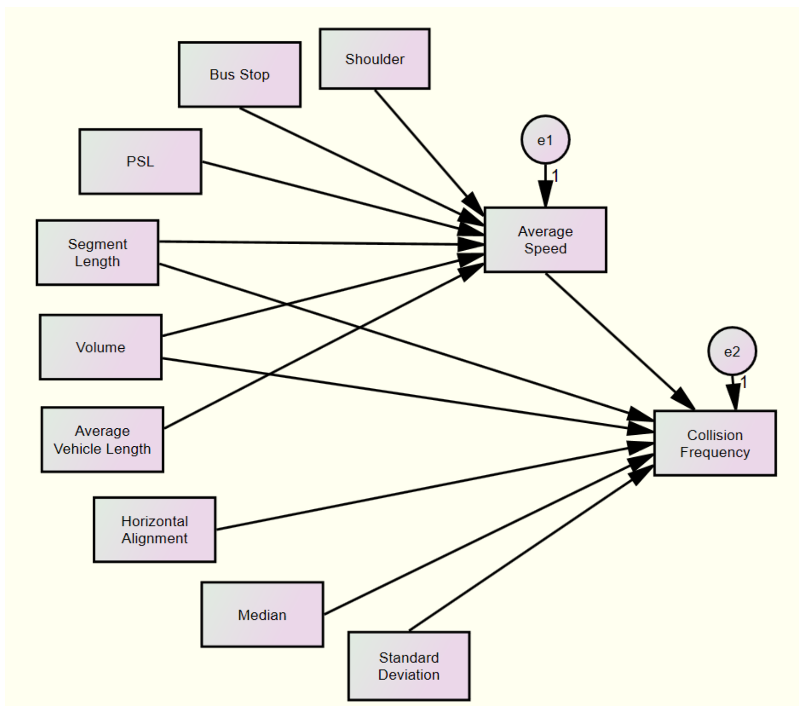

FIGURE 3b: Final Path Model

5. Results and Discussion

Several different variables were tested in the modelling stage; however, variables were kept in the model only if they were statistically significant. Overall, the outcomes shown in Table 2 seem reasonable and indicate both direct and indirect effects of different variables on collision frequency. Variables found to be statistically insignificant were dropped from the models in a stepwise manner using backward elimination.

TABLE 2: Modelling Results

	Variable	**Estimate**	**S.E.~**	**Est./S.E.**	***p*-value***
Effects on Average Speed (AVE)	PSL	0.415	0.128	3.236	0.001
	Shoulder Lane	7.435	2.574	2.888	0.004
	Bus Stop	4.474	0.655	6.831	<0.001
	Volume	1.667	0.353	4.723	<0.001
	Vehicle Length	2.692	1.09	2.47	0.014
	Segment Length	0.003	0.001	2.958	0.003
Direct Effects on Collisions (TC)	Average Speed	0.018	0.008	2.084	0.037
	Median	1.753	0.552	3.179	0.001
	Segment Length	0.001	0	3.395	0.001
	Standard Deviation	-0.072	0.042	-1.707	0.088
	HA	-0.359	0.147	-2.435	0.015
	ln(Volume)	1.89E-01	5.90E-02	3.204	0.001
Indirect Effects on Collisions	PSL	0.007	0.005	1.567	0.117
	Shoulder Lane	0.131	0.07	1.863	0.062
Indirect Effects on Collisions Intercepts	Bus Stop	0.029	0.015	2.005	0.045
	Ln(Volume)	2.90E-02	0.015	2.005	0.045
	Vehicle Length	0.047	0.03	1.564	0.118
	Segment Length	0.000045	0	1.975	0.048
	AVE	0.723	7.84	0.092	0.927
	TC	0.951	0.626	1.52	0.128
Residual Variances Dispersion	AVE	38.432	3.56	10.795	0
	TC	0.781	0.092	8.528	0

NOTE: *90% confidence level is used (i.e., *p*-value <0.1 is statistically significant), ~ S.E.: Standard Error

Above all, the mediator, average speed, was found to have statistically significant effects on crash frequency. As expected, the two variables were found to be positively correlated, which indicates that a higher collision frequency is anticipated at road segments with higher average speeds. In contrast, the model shows that the standard deviation of speed seems to be inversely related to collisions (i.e., increases in the deviation of speeds from the average were related to decreases in collision frequency, and vice versa); however, this relationship was only statistically at the 10% significance level (*p-value = 0.088*). It is worth noting that, in most cases where a positive association between standard deviation of speed and collisions was found, a negative and, in some cases, statistically insignificant relationship between average speed and collisions was also observed (Quddus, 2013).

As anticipated, the measure of traffic volume had statistically significant effects on both crash frequency and average speed. Traffic volume was found to be positively correlated with both average speeds and collision frequency. Although the positive relationship with collisions is fairly reasonable and is consistent with previous work (Wang *et al.*, 2013), the finding with respect to average speed is not so. Typically, increases in traffic volume would be expected to force drivers to reduce speeds due to congestion, but the reasoning here could lie in the fact that congestion effects have been filtered out of the speed data. It is also important to note that the speed-flow relationship included in the developed model represents the relationship between average speed and corresponding average flow at multiple locations. This is not the case for a typical speed flow diagram where the speeds of all vehicles at a certain location are plotted against the flow at that location. The difference here is that when data from multiple locations are used, other factors, in addition to flow, might influence average speeds collected at different locations. In addition to the direct effects, traffic volume also had an indirect and positive statistically significant effect on collision frequency, indicating partial mediation in the relationship.

Segment length was also a parameter with a highly significant effect on both average speed and crash frequency, recording *p*-values less than 0.01 in both cases. As in the case of traffic flow, increases in segment length were associated with increases in both average speeds and crash frequency, which are reasonable findings. In terms of indirect effects, segment length was another variable with statistically significant effects on crash frequency mediated through average speed (*p-value* = 0.048).

Locations with higher PSL, as expected, are associated with higher average speeds, a relationship that was found to be statistically significant and in which, on average, a 10kph increase in the PSL is associated with a 3.7kph increase in average speed. In addition to its effects on speed, the PSL also had indirect effects on collisions, which were statistically significant at the 90% level. Unlike the indirect effects, the direct effects of PSL on collision frequency were statistically insignificant; according to Baron and Kenny (1986), this is known as a full mediation effect. This is perfectly intuitive considering that it is the higher speeds on roads with higher speed limits that cause collisions and not the speed limits (i.e., PSLs affect collisions through their effects on speeds).

Another variable with statistically significant effects on collision frequency but not average speed was HA. The model shows that, on average, significantly fewer collisions are expected at locations of horizontal curvature compared to tangential segments. Although this is counterintuitive, the finding has also been observed in previous work (Hauer *et al.*, 2004, Anastasopoulos *et al.*, 2012), and is often attributed to one of two reasons: (i) drivers being more alert due to varying road geometry, known as highway hypnosis (Anastasopoulos *et al.*, 2008), or

(ii) drivers being more cautious due to a decrease in subjective safety, known as risk compensation (Assum *et al.*, 1999).

As with the presence of HA, the presence of a physical median also had a positive and statistically significant effect on crash frequency. Physical medians were also found to be associated with an increase in collisions in previous work as well (Anastasopoulos *et al.*, 2012). Nevertheless, what is interesting with respect to medians is that, even though the presence of a median would be expected to increase subjective safety, and hence result in less cautious driving, no significant relationship was found between their presence and average speeds in this paper.

A feature that is quite common on urban roads and has not been considered much in previous work is the presence of bus stops. Table 2 shows the variable to be statistically significant, and the model shows that it was positively related to both average speed and collision frequency (only indirectly in the case of the latter). Although this is quite surprising considering that more conservative driving is expected at those locations, similar findings were previously concluded at the City of Edmonton (Barua *et al.*, 2015). One reason for this finding could be the wider design of roads at locations with bus facilities.

Average vehicle length was found to significantly affect average speed. On average, increases in vehicle length were linked with increases in average speed at a specific location. It is worth noting here that previous work, which had assessed the relationship, also found that longer vehicles tend to make higher speed choices, compared to shorter passenger cars (Giles, 2004). As in the case of PSL, average vehicle length was also positively associated with crash frequency; however, only indirect effects were found to be statistically significant. In other words, longer vehicles make higher speed choices, which results in a higher collision frequency at a particular location.

The presence of a shoulder on the road is another variable that had fully mediated effects on crash frequency. As in the case of PSL, vehicle length and bus stops, the effects of the presence of a shoulder lane, when regressed on crash frequency, were found to be statistically insignificant (*p-value* of 0.549, not shown in the paper). However, the indirect effect of the variable on crash frequency through average speed was found to be statistically significant (*p-value = 0.062*), as seen in Table 2. This implies that, although the presence of a shoulder has no statistically significant effects on crash frequency directly, it still affects crash frequency through its effects on speed. In other words, the presence of a shoulder is associated with increases in speeds, which, in turn, result in an increase in crash frequency.

The importance of such a finding lies in the fact that it could help authorities identify alternative means of influencing the mediator (speed), and in the long term the dependent variable (collisions). For instance, consider a road segment where shoulders are present, 85th percentile speeds are high and the location is also prone to collisions. In this scenario, authorities might be tempted to intensify enforcement at this location as a means of reducing speed (the mediator), which should also reduce collisions. However, in some cases, a more effective way to manage the problem might be to consult with a road designer on removing or narrowing shoulder lanes. Even though the presence of a shoulder has no direct effects on collisions, according to the analysis, removing the shoulder could help in managing speeds and, consequently, managing collisions. In other cases, enforcement might already be at maximum levels at a certain location; therefore officials need to consider other means of influencing speeds and consequently collisions.

It is important to note that climate variables considered in the analysis (average wind speed, average temperature and average visibility) had no statistically significant effects; this was true for both independent variables (collisions and average speeds). Similarly, the effects of the percentage of speed limit violations, the presence of bike lanes, and the presence of pedestrian crossing are also variables that were considered in the analysis but found to have no statistically significant effects on average speed and/or collisions.

6. Conclusion

Regardless of the relationship analyzed, understanding how different factors interact and how one variable influences the other is vital from a scientific perspective. Such analysis is extremely important, yet quite challenging, especially when dealing with a complex relationship like the one observed between speed and the likelihood of collision occurrence. This complexity indicates the need for more research where the different interactions and confounding factors are considered and where the different factors mediating and moderating the relationship are accounted for.

This study proposes an approach through which all these different relationships could be modelled and, consequently, important inferences can be made. SEM and path analysis have been recently adopted in many fields, particularly in social science (Hauer *et al.*, 2004). In this paper, path analysis was used to identify and control for the different variables confounding the speed-safety relationship. The outcomes of the analysis are quite encouraging, both in terms of the reasonable findings with respect to the effects of the different confounding factors on speed and safety and the mediated effects of some variables on safety through their effects on speeds.

Average speed was observed to significantly affect crash frequency, with increases in speed being associated with an increase in collisions. Traffic volume and segment length were found to have both direct and indirect effects (only slight statistical significance) on crash frequency, as well as affect average speeds. Among other statistically significant variables were shoulder lanes, horizontal alignments and medians. Shoulders were found to have effects on average speed and collisions, albeit only indirectly. On average, it was found that horizontal alignments were associated with a lower number of collisions, while the presence of a median resulted in an increase of collisions at a certain location. Another important finding from this study was the high correlation between the number of lanes and average speeds within the considered dataset. This significant correlation masked the relationship between average speed and collisions, and hence, the analysis had to be performed only on two-lane urban roads to overcome this issue. A similar masking effect due to pedestrian activity has previously been observed (Baruya and Finch, 1994).

Despite these promising results, some limitations exist in the analysis. Estimating the indirect effects using the product-of-coefficient estimator could be seen as a drawback of this study. Since the outcome of the analysis is a count variable, estimating the indirect effects using this technique could be susceptible to potential bias. However, the large sample size used in the study is expected to minimize any potential biases (MacKinnon *et al.*, 2007, Coxe and MacKinnon, 2010). Moreover, previous research recommends using the product-of-coefficient estimator over using the difference-in-coefficient method, since it tends to give less biased estimates for the indirect effects (MacKinnon *et al.*, 2007).

In terms of the variables modelled, although attempts were made to consider most variables potentially confounding the speed-safety relationship, there could be other factors that were not included due to dataset limitations. These could be latent variables representing driver psychology and mood; such variables could be considered in future work. Similarly, using geometric attributes of horizontal curves, such as degree of curvature and radius of curve, and lane widths could also provide more insight into variables confounding the speed-safety relationship; hence, it is recommended to include these variables in future work.

Future research could also extend the analysis performed in this paper by estimating the effects using causal mediation. This is a more advanced technique that has been developed in recent years to estimate causally-defined direct and indirect effects; a further advantage of this technique is that it could also be used to analyze non-linear and nonparametric models (Muthén, 2011).

Acknowledgements

This study was sponsored jointly by a NSERC Discovery Grant and the City of Edmonton's Office of Traffic Safety (OTS).

References

Aarts, L., Van Schagen, I., 2006. Driving speed and the risk of road crashes: A review. Accident Analysis & Prevention 38 (2), 215-224.

Abdel-Aty, M.A., Radwan, A.E., 2000. Modeling traffic accident occurrence and involvement. Accident Analysis & Prevention 32 (5), 633-642.

Anastasopoulos, P.C., Mannering, F.L., Shankar, V.N., Haddock, J.E., 2012. A study of factors affecting highway accident rates using the random-parameters tobit model. Accident Analysis & Prevention 45, 628-633.

Anastasopoulos, P.C., Tarko, A.P., Mannering, F.L., 2008. Tobit analysis of vehicle accident rates on interstate highways. Accident Analysis & Prevention 40 (2), 768-775.

Assum, T., Bjørnskau, T., Fosser, S., Sagberg, F., 1999. Risk compensation—the case of road lighting. Accident Analysis & Prevention 31 (5), 545-553.

Baron, R.M., Kenny, D.A., 1986. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of personality and social psychology 51 (6), 1173.

Barua, S., El-Basyouny, K., Islam, M.T., 2015. Factors influencing the safety of urban residential collector roads. Journal of Transportation Safety & Security (just-accepted), 00-00.

Baruya, A., Year. Speed-accident relationships on european roads. In: Proceedings of the 9th International Conference on Road Safety in Europe.

Baruya, A., Finch, D., Year. Investigation of traffic speeds and accidents on urban roads. In: Proceedings of the TRAFFIC MANAGEMENT AND ROAD SAFETY. PROCEEDINGS OF SEMINAR J HELD AT THE 22ND PTRC EUROPEAN TRANSPORT FORUM, UNIVERSITY OF WARWICK, ENGLAND, SEPTEMBER 12-16, 1994. VOLUME P381.

Bollen, K.A., 2014. Structural equations with latent variables John Wiley & Sons.

Coxe, S., Mackinnon, D.P., 2010. Mediation analysis of poisson distributed count outcomes. Multivariate Behavioral Research 45 (6), 1022-1022.

Davis, G.A., 2002. Is the claim that 'variance kills' an ecological fallacy? Accident Analysis & Prevention 34 (3), 343-346.

Elvik, R., Christensen, P., Amundsen, A., 2004. Speed and road accidents. An evaluation of the Power Model. TØI report 740, 2004.

Fitzpatrick, K., Carlson, P., Brewer, M., Wooldridge, M., 2001. Design factors that affect driver speed on suburban streets. Transportation Research Record: Journal of the Transportation Research Board 1751 (1), 18-25.

Giles, M.J., 2004. Driver speed compliance in western australia: A multivariate analysis. Transport Policy 11 (3), 227-235.

Goldenbeld, C., Van Schagen, I., 2007. The credibility of speed limits on 80km/h rural roads: The effects of road and person (ality) characteristics. Accident Analysis & Prevention 39 (6), 1121-1130.

Hassan, Y., Sarhan, M., 2011. Modeling operating speed: Synthesis report. Chapter 1: Introduction. Transportation Research E-Circular (E-C151).

Hauer, E., 2009. Speed and safety. Transportation Research Record: Journal of the Transportation Research Board 2103 (1), 10-17.

Hauer, E., Council, F., Mohammedshah, Y., 2004. Safety models for urban four-lane undivided road segments. Transportation Research Record: Journal of the Transportation Research Board (1897), 96-105.
Hayes, A.F., 2013. Introduction to mediation, moderation, and conditional process analysis: A regression-based approach Guilford Press.
Haynes, R., Jones, A., Kennedy, V., Harvey, I., Jewell, T., 2007. District variations in road curvature in england and wales and their association with road-traffic crashes. Environment and Planning A 39 (5), 1222.
Islam, M.T., El-Basyouny, K., 2015. Full bayesian evaluation of the safety effects of reducing the posted speed limit in urban residential area. Accident Analysis & Prevention 80, 18-25.
Kline, R.B., 2015. Principles and practice of structural equation modeling Guilford publications.
Kweon, Y.-J., Kockelman, K.M., 2005. Safety effects of speed limit changes: Use of panel models, including speed, use, and design variables. Transportation Research Record: Journal of the Transportation Research Board 1908 (1), 148-158.
Lave, C.A., 1985. Speeding, coordination, and the 55 mph limit. The American Economic Review, 1159-1164.
Levy, D.T., Asch, P., 1989. Speeding, coordination, and the 55-mph limit: Comment. The American Economic Review, 913-915.
Lord, D., Mannering, F., 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.
Mackinnon, D.P., Lockwood, C.M., Brown, C.H., Wang, W., Hoffman, J.M., 2007. The intermediate endpoint effect in logistic and probit regression. Clinical Trials 4 (5), 499-513.
Milton, J., Mannering, F., 1998. The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. Transportation 25 (4), 395-413.
Montella, A., Imbriani, L.L., Marzano, V., Mauriello, F., 2015. Effects on speed and safety of point-to-point speed enforcement systems: Evaluation on the urban motorway a56 tangenziale di napoli. Accident Analysis & Prevention 75, 164-178.
Muthén, B., 2011. Applications of causally defined direct and indirect effects in mediation analysis using sem in mplus. Manuscript submitted for publication.
Nhtsa, 2012. Traffic safety facts, nhtsa, us department of transportation, 2009.
Nilsson, G., Year. Effects of speed limits on traffic accidents in sweden. In: Proceedings of the The international symposium on the effects of speed limits on traffic crashes and fuel consumption, Dublin. Organisation for Economy, Co-operation, and Development (OECD), Paris.
O'flaherty, C., Coombe, R., 1971. Speeds on level rural roads: A multivariate approach. Traffic Engng Control 13 (1-3).
Oecd/Ecmt, 2006. Speed management. Paris: OECD, Joint Transport Research Committee, Paris.
Pearl, J., 2010. The foundations of causal inference. Sociological Methodology 40 (1), 75-149.
Quddus, M., 2013. Exploring the relationship between average speed, speed variation, and accident rates using spatial statistical models and gis. Journal of Transportation Safety & Security 5 (1), 27-45.
Research Triangle Institute, 1970. Speed and accident, . In: Bureau, N.H.S. ed., North Carolina.

Sobel, M.E., 2008. Identification of causal parameters in randomized studies with mediating variables. Journal of Educational and Behavioral Statistics 33 (2), 230-251.
Solomon, D., 1964. Crashes on main rural highways related to speed, driver and vehicle. Bureau of Public Roads. US Department of Commerce. United States Government Printing Office, Washington DC.
Taylor, M., Baruya, A., Kennedy, J., 2002. The relationship between speed and accidents on rural single-carriageway roads.
Taylor, M., Lynam, D., Baruya, A., 2000. The effects of drivers' speed on the frequency of road accidents Transport Research Laboratory Crowthorne.
Theofilatos, A., Yannis, G., 2014. A review of the effect of traffic and weather characteristics on road safety. Accident Analysis & Prevention 72, 244-256.
Wang, C., Quddus, M.A., Ison, S.G., 2013. The effect of traffic and road characteristics on road safety: A review and future research direction. Safety science 57, 264-275.
Yannis, G., Louca, G., Vardaki, S., Kanellaidis, G., 2013. Why do drivers exceed speed limits. European Transport Research Review 5 (3), 165-177.