Accident Analysis and Prevention 150 (2021) 105896

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Exploration of the relationship among roadway characteristics, operating speed, and crashes for city streets using path analysis

Eun Sug Park *, Kay Fitzpatrick, Subasish Das, Raul Avelar

Texas A&M Transportation Institute, Texas A&M University System, 3135 TAMU, College Station, TX 77843-3135, United States

ARTICLE INFO

Keywords:
Operating speed
Crashes
Roadway geometry
Path analysis
Posted speed limit

ABSTRACT

Estimating the speed-crash relationship has long been a focus area of interest in roadway safety analysis. Because of many confounding factors that may influence both speeds and crashes, the relationship cannot be appropriately established without considering the corresponding roadway contexts and accounting for their effects on speeds and crashes. This paper investigates the speed-crash relationship for city streets by jointly modeling speed, roadway characteristics, and crashes using a path analysis approach that has been recently introduced into safety analysis while incorporating a wide range of roadway and traffic related variables and additional speed measures. The results from the coherent path analysis identified multiple speed measures of interest that have a statistically significant association with crashes as well as having intuitive and useful interpretation. The results also supported a positive relationship between speed variability and crash occurrence (i.e., larger spread/variability in operational speed is associated with more crashes).

1. Introduction

There has been great interest in assessing the relationship between speed and crash frequency for the past five decades as speed is considered as one of the key factors contributing to crash risk. The speed-crash relationship is often confounded by many other factors (road characteristics, weather, etc.) and as a result the estimated relationship has not been consistent across different studies, datasets, or speed measures. While speed variation was found to have an adverse effect on safety in most of past studies, the findings of previous research on the relationship between average speed and crashes have been conflicting with each other (Wang et al., 2013; Gargoum and El- Basyouny 2016). For example, a negative relationship between average speed and crash frequency/crash rates was found by some studies (see, e.g., Baruya, 1998; Quddus, 2013) while a positive relationship was found by other studies (e.g., Taylor et al., 2000; Aarts and van Schagen, 2006; Gitelman et al., 2017). Confounding factors have often been cited as possible reasons for such a disputable relationship (see., e.g., Elvik et al., 2004; Aarts and Van Schagen, 2006). The argument is that confounding factors may influence both speeds and crashes and consequently mask the true speed-crash relationship and interactions.

As a matter of fact, the speed-crash relationship cannot be appropriately established without considering the corresponding contexts (such as roadway type, roadway geometry, traffic, etc., which may confound the relationship between speeds and crashes if not considered) and how different factors interact. A holistic approach incorporating underlying relationships among different types of variables into estimation of the speed-crash relationship is desired.

There are two primary relationships among variables to be considered: the relationship between speed and roadway characteristics including traffic volume and other roadway geometry and traffic control device variables; and the relationship between crashes and speed. In most previous analyses, the relationship between speed and roadway characteristics variables and the relationship between crashes and speed-related variables (e.g., average speed, variance or standard deviation of speeds) along with roadway characteristics variables were assessed separately. A notable exception is the study by Gargoum and El-Basyouny (2016).

Gargoum and El-Basyouny (2016) is an important study that attempted incorporating those different relationships simultaneously into modeling to explore the association between speed and safety for urban streets. They employed path analysis (Bollen, 2014) to identify and control for the different variables confounding the speed-crash relationship based on data for 353 two-lane urban roads in the city of Edmonton, Canada. The average speed data at each location, computed using a total of 35 million individual vehicle speed records collected for

* Corresponding author.
E-mail addresses: e-park@tamu.edu (E.S. Park), K-Fitzpatrick@tamu.edu (K. Fitzpatrick), s-das@tti.tamu.edu (S. Das), r-avelar@tamu.edu (R. Avelar).

https://doi.org/10.1016/j.aap.2020.105896
Received 14 July 2020; Received in revised form 15 October 2020; Accepted 10 November 2020
Available online 4 December 2020

a five-year period (2009–2013), were linked to the crash frequency at each location during the same time frame along with other factors such as road, traffic, and climate. The results showed that average speed, standard deviation of speed, traffic volume, segment length, presence of median, and horizontal alignment have statistically significant (at $\alpha = 0.1$) direct effects on crashes and posted speed limit, shoulder lanes, and bus stop have statistically significant indirect effects on crashes. While the outcomes of the analysis are encouraging, some of the estimated effects are counterintuitive. For example, while average speed was found to be positively correlated with crashes, standard deviation of speed was found to be inversely related to crashes (i.e., negatively correlated), which contrasts with many of previous studies. The authors noted that the standard deviation negative relationship with crashes was statistically significant only at the 10 % significance level (p-value = 0.088). The presence of a median, horizontal alignment (presence of curve), and shoulder lanes (presence of a shoulder) were also determined to be positively associated with crash frequency, which deserve further investigations. Several key urban roadway features such as land use development, driveway density, school zone, and presence of curb were not included in the study conducted by Gargoum and El-Basyouny (2016). Additionally, only two speed measures (average speed and standard deviation of speed) were used in their study. To gain a broader understanding of speed-crash relationships for city streets, the proposed approach is to conduct an extended study that explores speed-crash association with inclusion of a wide range of potential contributing factors.

The objective of this paper is to continue the investigation into the relationship between speed and safety for city streets by using a wide range of roadway and traffic related variables and speed measures. This paper is to coherently assess the speed-crash relationship by exploring the underlying relationships among speed, crashes, and roadway characteristics simultaneously based on the roadway characteristics-speed-crash databases and identifying important variables that influence speed and crashes. In addition to conventional speed measures such as average speed and standard deviation (or variance) of speed, several additional speed measures that can quantify various aspects of speed distributions are examined to identify most relevant speed variable(s) that can effectively describe speed-crash relationship.

2. Previous studies on speed-crash relationship

Current state-of-the-art knowledge suggests that decreasing speed variance on the road decreases crash likelihood and that if all vehicles traveled at the same speed, the opportunity for crashes would be low. This belief may be based on the Solomon (1964) study, which found that on rural highways crash likelihood increases as the discrepancy between an individual driver's speed and the mean speed on the roadway increases. However, the study had limitations in the methodology, the precision of the data used, and in being only for one functional classification (rural highways). In addition to these limitations, the data collected for Solomon's study are now over six decades old; since that time, there have been substantial changes in driver demographics; driving behavior; vehicle safety features and performance; and, perhaps to a lesser extent, highway design.

Researchers have used various operating speed measures to predict the association between the speed choice and its safety effect. The most frequent measures used for this purpose are the individual speeds, space mean speed, and the speed variance (Aarts and Schagen, 2006). By observing vehicle speeds and conducting surveys or questionnaires with individual drivers, previous research efforts have determined that crash liability increases as the individual driving speeds increase (Fildes et al., 1991; Maycock et al., 1998).

Some earlier studies determined that the average operating speed is positively correlated with the crash rate and more so with the crash severity (Hauer, 1971; Nilsson, 2004; Elvik et al., 2004). Taylor et al. (2000), a study based on the United Kingdom, showed that the average speed measure was not positively related to crash frequency at aggregate level. Using different homogenous groups, however, the findings showed that crash counts are positively associated with the average operating speed.

Pei et al. (2012) performed a study exploring the link between speed and crash risk. They found several explanatory factors of the association, including road design, weather types, and temporal patterns. The study showed negative association between crash and average operating speed. In another study, Yu et al. (2013) analyzed crash data from one year on I-70 in Colorado. The results show that weather condition has a significant relationship with crash occurrence. Like previous studies, this study found that roadway segments with low posted speed limit and high upstream traffic volume (5–10 min before the crash time) has association with crash occurrences. This finding could be a result of traffic congestion; it could also be affected by severe weather conditions which would then be a confounding variable. The findings from a pre-crash condition-based study conducted by Imprialou et al. (2016) indicated positive association between annual average operating speed and crash frequency. However, link based finding shows the opposite association. The study mentioned that a link-based approach might be associated with significant errors due to the way of data aggregation conducted in the study.

Yu et al. (2018) conducted a study using advanced traffic sensing data to investigate aggregation approaches and their impact on crash-speed relationship analysis. Using data from Shanghai urban expressway system, the results show that during congested traffic conditions, average operating speed is negatively associated with crash frequency. In another study, Wang et al. (2018) used taxi-based high frequency GPS data from eight arterials in Shanghai to investigate the impact of speed on traffic crashes The findings show that a 1% increase in mean operating speed was associated with a 0.70 % increase in total crash counts.

Dutta and Fontaine (2019) conducted a study using probe data and continuous count station data on 6-lane urban freeway segments in Virginia. The study found that speed variation shows an increasing effect on crash frequency on urban freeways. Banihashemi et al. (2019) found that an increase in the severity of crashes (a ratio of fatal and injury crashes to total crashes) was related to an increase in speed differential.

Hutton et al. (2020) examined the speeds of individual drivers along 100 study segments to evaluate variations among drivers on the same roadway segment as well as variance within individual trips. The findings show that road characteristics have impact on speed-crash relationship. The findings also show that speed variation between trips is associated with multi-vehicle crash incidents. Most of the other evaluated speed measures (for example, space mean speed, average operating speed) either had a negative correlation or they were not found to be related to crash frequency. Das et al. (2020) study examined the speed-crash relationship in USDOT Safety Data Initiative (SDI) rural speed safety pilot project. The uniqueness of this study is the usage of aggregated speed measures of 5-minute interval operating speed and rainfall data in the modeling framework. The findings show that speed-related operational information (for example, average operating speed, standard deviation of operating speed, difference between weekend and weekday operating speed) can be used to understand safety outcomes on rural highways better.

As presented, the findings of previous studies are not consistent, sometimes even conflicting, which may have been caused by the issue of *ecological fallacy* or *aggregation bias* (see, e.g., Davis, 2002, 2004; Hauer, 2005) in addition to other confounding factors. This again calls for a holistic approach considering the corresponding roadway contexts, as well as taking into account confounding factors and how different factors interact, in deriving a speed-crash relationship.

3. Data

The database was developed as a fusion of several datasets which contains data for operating speed, crashes, and roadway characteristics including posted speed limit and vehicle volume. The researchers determined that the most efficient approach to building the database was to investigate existing sources to obtain the needed data and supplement those sources as resources permit. Operating speed data were obtained from 663 on-road tube sites (RTS) in the city of Austin collected between 2015 and 2017. Crash data for 2011–2017 were extracted from the Crash Records Information System (CRIS) data assembled and maintained by the Texas Department of Transportation (TxDOT). TxDOT also maintains a database that includes a variety of roadway characteristics and AADT values for the corresponding roadway sections as the Roadway Highway Inventory Network Offload (RHiNO). To determine the association among operating speed, roadway characteristics, and crash outcomes, the researchers developed a database where the RTS was conflated with RHiNO and CRIS to build the database for analysis. In addition, the researchers used Google® Earth to gather the additional necessary roadway and traffic control device data for each segment. The segment boundaries were identified as intersections with traffic control or roadway geometry that would affect speed, generally being intersections with traffic control signals or stop-control on the segment and in a few cases a horizontal curve or the end of the street. More details on the compilation of speed data, crash data, and roadway characteristics data can be found in Fitzpatrick et al. (2021).

Tables 1 and 2 list the distribution of number of segments and total segment length by posted speed limit and number of lanes, respectively. The dominant groups for these variables are roadways with a posted speed limit of 30 mph and two-lane roadways.

Tables 3 and 4contain descriptions of the specific geometric variables considered in the analyses and the descriptive statistics for them, respectively. These variables were primarily chosen based on the findings from relevant studies as discussed in the literature review.

4. Methods and results

To examine the relationships among crashes, speed, and roadway characteristics including posted speed limit, and volume, the researchers selected non-intersection or segment crashes rather than all crashes. All crashes would have included intersection crashes, especially signalized intersection crashes. The non-intersection or segment crashes were identified as those coded as not intersection (NI) or driveway (D) related, and the abbreviation of NID was selected to indicate that the crashes reflected segment crashes. The researchers conducted evaluations using crashes with injuries (KABC) and crashes with all severity levels (KABCO; where K = Fatal, A = Incapacitating Injury, B = Non-Incapacitating Injury, C = Complaint, and O = No Injury) in case the evaluations that only included the fatal/injury crashes was limited by sample size.

In this study, researchers considered several speed measures that can quantify various aspects of speed distributions at each segment including newly developed measures as well as speed measures explored previously in other studies such as average speed and standard deviation of speed. Table 5 contains the speed measures considered in the current analysis.

Fig. 1 shows some of the speed measures (PSL, SpdAve, PSL – Avg, Pace, PSL – S85) of Table 5 to help illustrate several of the speed measures considered in this study. Fig. 1(a) and Fig. 1(b) show the speed distribution for a segment with a posted speed limit of 60 mph and 30 mph, respectively. The calculated values of the average, 50th percentile, and 85th percentile speeds are called out within the plots. Within traffic engineering, the 10-mph pace, which is defined as the 10-mph range containing the most vehicles, is sometimes considered when evaluating the operations on the roadway. For example, the process in the Florida Department of Transportation (DOT) Speed Zoning for Highways, Road and Streets in Florida manual (2018) is to post the speed limit at or near the upper limit of the 10-mph pace when the observed 85th percentile speed falls above the upper limit of the 10-mph pace. The roadway segment shown in Fig. 1(a), has a speed distribution where the upper limit of the 10-mph pace (62 mph) is slightly less than the 85th percentile speed of 62.5 mph. The percent of vehicles within that 10-mph range was considered within the analyses discussed in this paper.

Table 1
Number of segments by posted speed limit.

Posted Speed Limit (mph)	Number of Segments	Length (mi)
25	169	52
30	318	138
35	68	36
40	51	37
45	43	28
50	12	13
55	2	2
Grand Total	**663**	**305**

Table 2
Number of segments by number of lanes.

Number of Lanes	Number of Segments	Length (mi)
2	529	222
4	126	78
6	8	6
Grand Total	663	305

Table 3
Roadway and traffic control device variables used in the analysis for City of Austin, Texas.

Variable	Description
Beg_IT_Legs	Number of legs for intersection at beginning of segment
Bike1yes	Bike lane presence: 1 = yes, 0 = no
Curb1yes	Is curb and gutter present on segment: 1 = yes or 0 = no
Develop2	Development: Resident = 1, Other = 0
DUpmBoth	Driveways and unsignalized intersections per mile in both directions
End_IT_Legs	Number of legs for intersection at ending of segment
Horz1tan	Horizontal alignment: 1=straight(tangent), 0=some horizontal curvature
Len_mi	Segment length (mi)
Median2	Median type: Raised = 1, NotRaised = 0
MedWidth	Typical or average median width for the segment (ft)
NSigInt	Number of signalized intersections along segment, including the signals at the begin or end of the segment
OnStPk2	On-street parking: OnStrPrk = 1, None = 0
PdCr1yes	Is a midblock marked pedestrian crossing present within the segment: 1 = yes or 0 = no
PedAuto	Typical or average distance between the sidewalk and the automobile lane for the segment, sum of the following (when present): parking width, bike width, bike-auto separation, and sidewalk to road separation (ft)
PSL	Posted speed limit (mph)
RoadSurf	Distance between the driving surface edges, calculated as number of through lanes multiplied by average lane width plus median width plus parking widths plus bike widths
RU_F_rev	Functional classification for street: Urban Local = 1, Not Local = 0
ScZn1yes	School zone presence: 1 = yes, 0 = no
SdWk1yes	Is a sidewalk present within the segment: 1 = yes or 0 = no
Site	Unique name for each site, consist of a segment number plus the primary direction for traffic (e.g., NB, SB, EB, or WB)
Vol_Day	Volume per day in both directions. Typically, the value is from TxDOT RHiNO's ADT_ADJ. If ADT_ADJ is not available or when ADT_ADJ = 405 (a known placeholder) the value is the average daily volume from the on-road counter

4.1. Path analysis

The focus of this analysis is to assess the effect of speed on crashes while accounting for the effects of other roadway characteristics variables on speed and crashes. Note that while roadway characteristic variables (e.g., traffic volume) may affect both speed and crashes, some

Table 4
Descriptive statistics of key variables.

Variable[a]	Variable Type[b]	Minimum	Maximum	Mean	Std. Deviation
Beg_IT_Legs	Numerical	1	5	3.48	0.59
End_IT_Legs	Numerical	1	5	3.48	0.58
DUpmBoth	Numerical	0	174.4	47.33	41.76
Len_mi	Numerical	0.06	2.77	0.46	0.29
MedWidth	Numerical	0	50	3.01	6.79
NSigInt	Numerical	0	2	0.68	0.85
PedAuto	Numerical	0	37	5.83	5.62
PSL	Numerical	25	55	31.42	6.20
RoadSurf	Numerical	18	100	41.67	15.07
Vol_Day	Numerical	92	44,673	6749.17	9404.19
Bike1yes	Dichotomous	0	1	0.22	0.42
Curb1yes	Dichotomous	0	1	0.95	0.22
Horz1tan	Dichotomous	0	1	0.35	0.48
PdCr1yes	Dichotomous	0	1	0.10	0.31
ScZn1yes	Dichotomous	0	1	0.09	0.28
SdWk1yes	Dichotomous	0	1	0.64	0.48
Develop2	Dichotomous	Residential (525), Other (138)			
Median2	Dichotomous	1: Raised (53), 0: NotRaised (610)			
OnStPk2	Dichotomous	1: OnStrPk(199), 0: None (464)			
RU_F_rev	Dichotomous	1: Local (360), 0: NotLocal (303)			

Notes:
[a] Variable descriptions are in Table 3.
[b] For dichotomous variables, '1′ indicates the presence of the feature and '0′ indicates its absence.

Table 5
Speed measures considered.

Speed Measures	Description
Abs(PSL − Avg)	Absolute value of posted speed limit minus average speed (mph)
CoefVar	Coefficient of variation of speed
Pace	Percent of vehicles in 10-mph pace for the site (%)
PerOvPSL	Percent of observations over the speed limit for the site (%)
PSL	Posted speed limit (mph)
PSL − Avg	Posted speed limit minus average speed (mph)
PSL − S85	Posted speed limit minus 85th percentile speed (mph)
S85−Avg	85th percentile speed minus average speed (mph)
SpdAve	Average speed (mph)
StdSpd	Standard deviation (mph)

variables such as the posted speed limit may affect crashes only through operating speeds and so only indirectly affect crashes. A speed variable plays the role of mediator variable (or intervening variable) between crashes and other variables that affect crashes only indirectly in this case. In addition to assessing the speed-crash relationship, it is also of interest to evaluate indirect effects of roadway characteristics on crashes through a mediator variable (speed) as well as direct effects of roadway characteristics on crashes.

To accommodate these general relationships among variables, researchers jointly modeled the relationship between crashes and speeds along with roadway characteristics and the relationship between speeds and roadway characteristics as well as the speed limit simultaneously based on a coherent Structural Equation Modeling (SEM) framework (Bollen, 2014), specifically using path analysis. Although SEM has been extensively used primarily in social and behavioral sciences (Tehrani and Yamini, 2020; Hollett et al., 2020; Hagger et al., 2018; Krishnan et al., 2018; Bellini et al., 2017; Seddig and Lomazzi, 2019; Özgür, 2020), a few recent transportation safety studies have also employed SEM (Choudhary et al., 2020; Ding et al., 2019; Mokarami et al., 2019; Lee et al., 2018). Path analysis is a special case of SEM where there is no latent variable in the model (i.e., all variables in the model are measured variables). Regardless of many advantages, path analysis has not been widely used in safety analysis yet. The study by Gargoum and El-Basyouny (2016) is one of the handful studies that employed the path analysis approach.

The path analysis model in this study consists of two submodels: 1. Crash model (Model 1) describing the relationship between crashes (outcome variable) and speed (mediator variable) as well as other roadway characteristic variables (independent variables); and 2. Speed model (Model 2) describing the relationship between speed and other roadway characteristic variables (including the posted speed limit). For the crash model (Model 1), a negative binomial model with the mean given in Eq. (1), which expresses the log mean crash frequency as a function of covariates corresponding to a speed variable and other roadway characteristic variables in Tables 4 and 5, was adopted.

$$\mu_i = \exp(\beta_0 + m_i\beta_m + X_{1i}\beta_1 + \cdots + X_{Ki}\beta_K) \quad (1)$$

where y_i is the observed outcome variable (the number of crashes that occurred on the segment in 7 years) on segment i ($i = 1,\cdots,I$), $\mu_i = E(y_i)$ is the expected number of crashes for 7 years, m_i is the mediator variable (a measure of speed), $X_{1i}, \cdots, X_{Ki}$ are K covariates, and $\beta_0, \beta_m, \beta_1, \cdots, \beta_K$ denote regression coefficients for the outcome variable.

For the speed model (Model 2), a normal linear model given in Eq. (2) was employed.

$$m_i = \alpha_0 + X_{1i}\alpha_1 + \cdots + X_{Li}\alpha_L + \varepsilon_i \quad (2)$$

where $\alpha_0, \alpha_1, \cdots, \alpha_L$ are regression coefficients.

Estimation was performed by SEM software Mplus version 8.3 (Muthén and Muthén, 2017). Several different models (with different sets of independent variables) for each mediator variable in Table 5 were explored. Table 4 contains the roadway characteristic variables used in path analysis. Researchers kept variables in the model if the corresponding p-values were less than 0.2. It needs to be noted, however, that there may be multiple models that may be adequate for any given data set. Researchers used the Bayesian Information Criterion (BIC) which is a popular penalized-likelihood criterion used for model selection in various applications (see, for example, Kass and Raftery, 1995), to select an appropriate model. Although a model with lower BIC is preferred in general, a physically more meaningful model can be selected whenever there is not much difference in BIC values among competing models. Note that BIC can be used for comparing models with different independent variables but not models with different dependent variables. That is, BIC values should not be compared across different outcome variables or mediator variables as they represent different data sets.

4.2. Results using all segments available

The estimated regression coefficients for crashes in Eq.s (1) and (2) with having each of speed variables in Table 5 as a mediator variable and roadway characteristic variables selected from those in Table 4 are given in Tables 6,7. The results were split between Table 6 and Table 7 for space reasons. Table 6 provides results for mediator variables of PSL (posted speed limit), CoefVar (coefficient of variation), PerOvPSL (percent of observations over the speed limit), StdSpd (standard deviation), Pace, and SpdAve (average speed). Table 7 provides results for the mediator variables that reflected the difference between two speed measures and includes Abs(PSL − Avg), which is the absolute value of difference between the posted speed limit and average speed, PSL − Avg (posted speed limit minus average speed), PSL − S85 (Posted speed limit minus 85th percentile speed), and S85−Avg (85th percentile speed minus average speed). For space issues, only the results based on the fatal/injury crashes (KABC_NID crashes) are reported in this paper. The results based on crashes with all severity levels (KABCO_NID crashes) are not significantly different from those based on the fatal/injury crashes except that the effect sizes for mediator variables are smaller (and in some cases become statistically insignificant) compared to those based on the fatal/injury crashes. Interested readers can consult Fitzpatrick et al. (2021) for the detailed results on KABCO_NID crashes.

It can be seen from Table 6 and Table 7 that the mediator variables, standard deviation, Pace, absolute value of difference between the

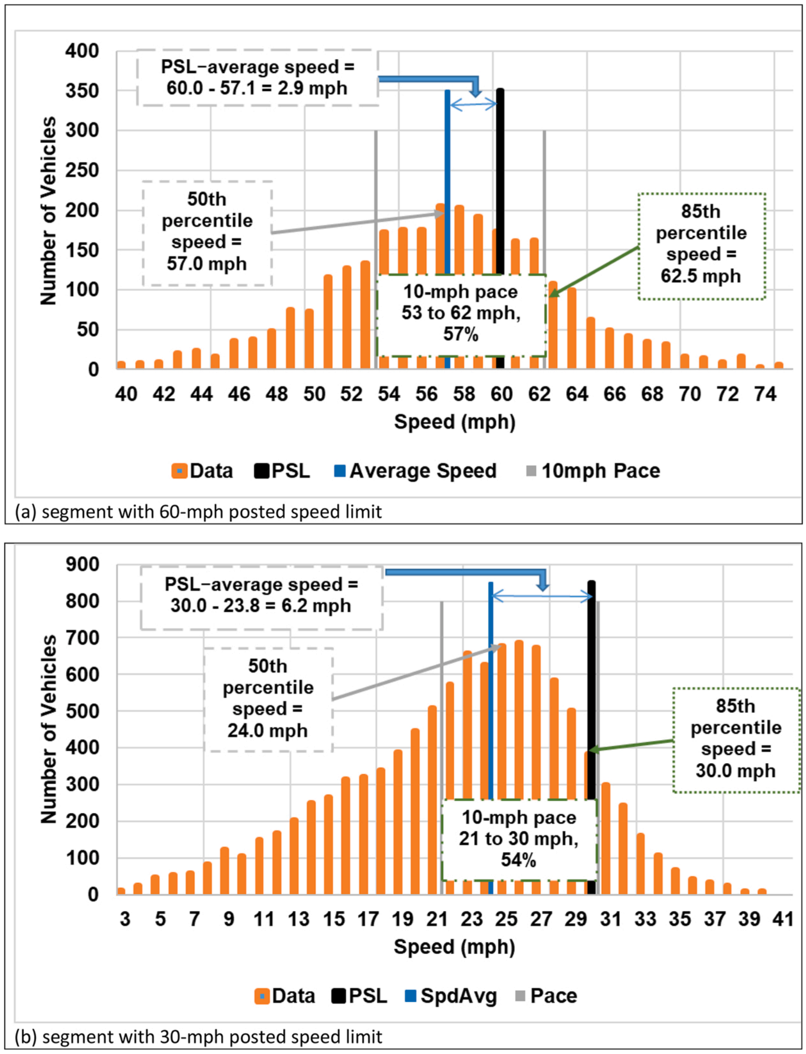

Fig. 1. Illustration of Several Speed Measures from Table 5.
(a) segment with 60-mph posted speed limit
(b) segment with 30-mph posted speed limit

posted speed limit and average speed, posted speed limit minus average speed, posted speed limit minus 85th percentile, and 85th percentile speed minus average speed, have statistically significant effects at $\alpha = 0.05$ on KABC_NID crash frequency. The association was positive for absolute value of difference between the posted speed limit and average speed, posted speed limit minus average speed, posted speed limit minus 85th percentile, 85th percentile speed minus average speed, and standard deviation of speed (i.e., as the values of those mediator increases, crash frequency increases), but negative for Pace (i.e., as the value of Pace increase, crash frequency decreases) which was expected. Also, the presence of curbs, being in a residential area, and presence of a raised median are associated with lower crash frequency. The number of signalized intersections, traffic volumes, and segment length were found to be positively correlated with crash frequency. As expected, higher Posted Speed Limit is associated with higher absolute value of difference between the posted speed limit and average speed, posted speed limit minus average speed, posted speed limit minus 85th percentile, 85th percentile speed minus average speed, standard deviation of speed, and average speed, but with lower Pace and Percent of observations over the speed limit. This implies that the posted speed limit has indirect effects on crashes through its effect on those speed measures.

4.3. Results Using Segments with PSL of 25–45 mph

Researchers also refitted the path analysis model for KABC crashes with absolute value of difference between the posted speed limit and average speed as a mediator variable after excluding the road segments with posted speed limits 50 and 55 mph because streets with those speed limits may be considered as streets not having typical city character. There was one segment with 55 mph as the posted speed limit and 6 segments with 50 mph as the posted speed limit in the original data. Excluding those segments resulted in removal of 14 sites out of 663 sites, leaving 649 sites corresponding to 25–45 mph segments in the data. Initially, a full model with all of roadway characteristic variables in Table 4 was fitted to the dataset consisting of 649 sites, and then the model was refitted after removing variables that were insignificant at $\alpha = 0.1$.

The crash rate (in Million Vehicle Miles Traveled, MVMT) for the Austin data was calculated and graphed with the speed metric of PSL − Avg (difference between posted speed limit and average speed) to

Table 6
Estimated regression coefficients for KABC_NID crashes by path analysis (mediator variable: PSL, CoefVar, PerOvPSL, StdSpd, Pace, or SpdAve).

Outcome Variable	Independent Variable	Mediator / Coefficient	PSL	CoefVar	PerOvPSL	StdSpd	Pace	SpdAve
KABC_NID crashes	Intercept	β_0	-3.259	-3.076	-2.144	-3.386	-0.856	-3.359
	Mediator	$\beta_{Mediator}$	0.032	2.059	-0.004	0.143	-0.023	-0.001
	Curb1yes	$\beta_{Curb1yes}$	-0.827	-1.037	-1.069	-0.916	-0.872	-0.921
	Develop2	$\beta_{Develop2}$	-0.362	-0.429	-0.438	-0.399	-0.368	-0.402
	Median2	$\beta_{Median2}$	-0.723	-0.641	-0.632	-0.684	-0.668	-0.680
	NSigInt	$\beta_{NSigInt}$	0.602	0.644	0.642	0.638	0.629	0.639
	SdWk1yes	$\beta_{SdWk1yes}$	0.263	0.249	0.260	0.246	0.260	0.245
	OnStPk2	$\beta_{OnStPk2}$	0.194	0	0	0.173	0.167	0.171
	LnVol	β_{LnVol}	0.488	0.584	0.545	0.542	0.526	0.545
	LnLen	β_{LnLen}	0.720	0.877	0.853	0.828	0.836	0.836
	StdSpd	β_{StdSpd}	0	0	0	NA	0	0.144
Speed	Intercept	α_0	27.534	0.336	66.776	5.218	76.743	8.817
	PSL	α_{PSL}	0	0	-2.004	0.061	-0.413	0.433
	Bike1yes	$\alpha_{Bike1Yes}$	-0.615	0	0	0	1.482	0
	Curb1yes	$\alpha_{Curb1yes}$	-2.585	-0.029	0	-0.780	4.196	0
	Develop2	$\alpha_{Develop2}$	-1.940	-0.025	0	-0.335	3.905	0
	DUpmBoth	$\alpha_{DUpmBoth}$	-0.020	0.000	-0.099	0.003	0	-0.022
	Horz1tan	$\alpha_{Horz1tan}$	0	0.023	0	0.265	-1.259	-0.941
	Median2	$\alpha_{Median2}$	-1.601	0	0	0	0	0
	NSigInt	$\alpha_{NSigInt}$	0.656	-0.019	6.480	-0.145	0	1.342
	OnStPk2	$\alpha_{OnStPk2}$	-2.731	0.021	-11.864	0.335	0	-2.167
	RoadSurf	$\alpha_{RoadSurf}$	0.150	-0.001	0.487	0	0	0.095
	PedAuto	$\alpha_{PedAuto}$	0	0	0	-0.025	0	0
	Vol_Day/1000	$\alpha_{Vol\_Day/1000}$	0.179	-0.001	0	0	0	0.044
	Len_mi	α_{Len_mi}	4.167	-0.061	36.210	0	0	7.084
Model fit	BIC		5968.8	777.5	8471.8	4424.7	6915.4	6210.5

Notes: 1. LnVol = Log(Vol_Day), LnLen = Log(Len_mi); 2.The coefficient '0' denotes that the corresponding variable was excluded from the model; 3. Cells are highlight in light gray when the p-value is between 0.05 and 0.1.; 4. Cells are highlighted in dark gray when the p-value is less than 0.05.

provide an appreciation of the potential relationship, see Fig. 2. Note that crash frequency along with segment length and volumes were used in the statistical analyses rather than the crash rate shown in Fig. 2. Fig. 2 also has a simple trend line to help illustrate the relationship. The minimum crash rate appears to be near the point when posted speed limit equals average speed. When Posted speed limit minus 85th percentile speed is compared to crash rate, the low point is below the zero point which creates challenges in interpreting relationships. Therefore, the absolute value of difference between posted speed limit and average speed may be a better speed measure to identify and understand how variables are affecting safety on city streets.

The estimated model coefficients for KABC crashes from path analysis based on data from 649 sites corresponding to 25–45 mph segments are presented in Table 8. It can be seen from the table that the mediator variable Abs(PSL − Avg), absolute value of difference between posted speed limit and average speed, is statistically significant at $\alpha = 0.1$ having a positive association with KABC crashes (i.e., as the values of Abs(PSL − Avg) increases, KABC crash frequency increases). Also, residential area and the presence of raised median are associated with lower crash frequency. The number of signalized intersections, traffic volumes, and segment length were found to be positively correlated with crash frequency. As expected, higher PSL (posted speed limit) is associated with higher Abs(PSL − Avg), which implies that PSL has indirect effects on KABC crashes through its effect on Abs(PSL − Avg). Note that the presence of bike lane, the presence of school zone, and functional classification for street (with 1 for Urban Local, 0 otherwise) also have indirect effects on KABC crashes as they have statistically significant positive associations with Abs(PSL − Avg). Distance between the driving surface edges (RoadSurf) has a statistically significant negative association with Abs(PSL − Avg), which also has an indirect effect on KABC crashes, subsequently. The relationships of the variables with either KABC_NID or Abs(PSL − Avg) is shown in Fig. 3. Note that the standard errors of model coefficient estimates are provided in parentheses.

5. Conclusions

This study assessed the speed-crash relationship by using path analysis which is a coherent approach jointly modeling the relationship between speed and roadway characteristics and the relationship between crash frequency and speed taking into account roadway contexts. In addition to conventional speed measures such as average speed and standard deviation of speed, additional speed measures were investigated to quantify various aspects of speed distributions on crash frequency.

The results showed that, regardless of the mediator variable (speed measure), the presence of a median or presence of curb and gutter has statistically significant (at $\alpha = 0.05$) negative direct effects on KABC crashes (i.e., the presence is associated with less crashes). Other variables having statistically significant (at $\alpha = 0.05$) direct effects on KABC crashes regardless of mediator variables are the number of signalized

Table 7
Estimated regression coefficients for KABC_NID crashes by path analysis (mediator variable: Abs(PSL − Avg), PSL − Avg, PSL − S85, or S85−Avg).

Outcome Variable	Independent Variable	Mediator / Coefficient	Abs(PSL−Avg)	PSL−Avg		PSL−S85	S85−Avg
KABC_NID crashes	Intercept	β_0	-2.877	-2.400	-3.021	-2.199	-3.171
	Mediator	$\beta_{Mediator}$	0.052	0.036	0.025	0.031	0.127
	Curb1yes	$\beta_{Curb1yes}$	-0.883	-0.941	-0.858	-0.992	-0.881
	Develop2	$\beta_{Develop2}$	-0.378	-0.427	-0.402	-0.438	-0.412
	Median2	$\beta_{Median2}$	-0.615	-0.599	-0.630	-0.603	-0.687
	NSigInt	$\beta_{NSigInt}$	0.635	0.632	0.627	0.635	0.639
	SdWk1yes	$\beta_{SdWk1yes}$	0.224	0.245	0.247	0.250	0.249
	OnStPk2	$\beta_{OnStPk2}$	0.144	0	0	0	0
	LnVol	β_{LnVol}	0.565	0.541	0.528	0.544	0.535
	LnLen	β_{LnLen}	0.875	0.909	0.887	0.887	0.811
	StdSpd	β_{StdSpd}	0	0	0.103	0	0
Speed	Intercept	α_0	-1.346	-8.817	-8.817	-12.625	5.623
	PSL	α_{PSL}	0.357	0.567	0.567	0.497	0.044
	Bike1yes	$\alpha_{Bike1Yes}$	0.039	0	0	0	0
	Curb1yes	$\alpha_{Curb1yes}$	-0.633	0	0	0	-0.950
	Develop2	$\alpha_{Develop2}$	-1.291	0	0	0	-0.424
	DUpmBoth	$\alpha_{DUpmBoth}$	0.008	0.022	0.022	0.019	0.003
	Horz1tan	$\alpha_{Horz1tan}$	0.576	0.941	0.941	0.732	0
	Median2	$\alpha_{Median2}$	-0.194	0	0	0	0
	NSigInt	$\alpha_{NSigInt}$	-0.793	-1.342	-1.342	-1.188	-0.127
	OnStPk2	$\alpha_{OnStPk2}$	0.634	2.167	2.167	1.931	0.420
	RoadSurf	$\alpha_{RoadSurf}$	-0.058	-0.095	-0.095	-0.087	0
	PedAuto	$\alpha_{PedAuto}$	0.031	0	0	0	-0.023
	Vol_Day/1000	$\alpha_{Vol\_Day/1000}$	-0.078	-0.044	-0.044	-0.050	0
	Len_mi	α_{Len_mi}	-3.367	-7.084	-7.084	-6.914	0
Model fit	BIC		5907.6	6200.4	6200.4	6172.7	4489.2

Notes: 1. LnVol=Log(Vol_Day), LnLen=Log(Len_mi); 2. The coefficient '0' denotes that the corresponding variable was excluded from the model; 3. Cells are highlight in light gray when the p-value is between 0.05 and 0.1.; 4. Cells are highlighted in dark gray when the p-value is less than 0.05.

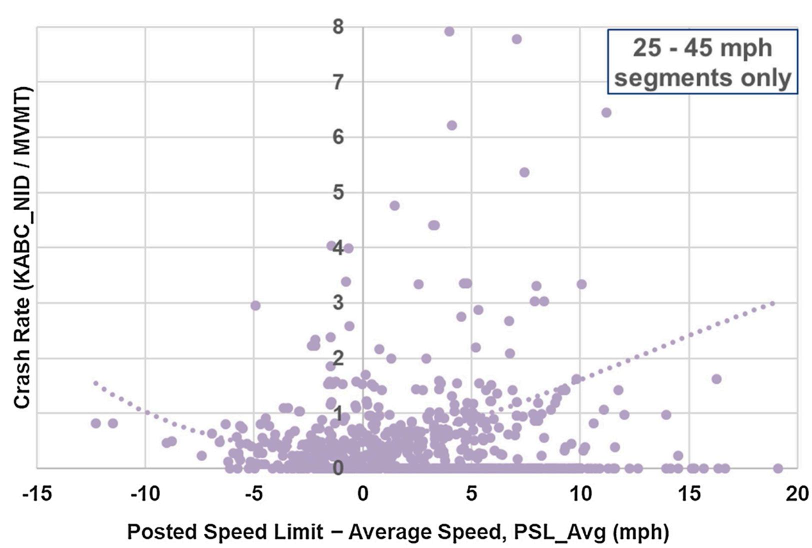

Fig. 2. Comparison of crash rate to the difference between posted speed limit and average speed.

Table 8
Estimated regression coefficients for KABC_NID crashes by path analysis with a mediator variable: Abs(PSL − Avg) based on 649 sites (PSL from 25 to 45 mph).

Outcome Variable	Independent Variable	Mediator / Coefficient	Abs(PSL−Avg)
KABC_NID crashes	Intercept	β_0	-3.648
	Mediator	$\beta_{Mediator}$	0.044
	Dvelop2	$\beta_{Develop2}$	-0.483
	Median2	$\beta_{Median2}$	-0.618
	NSigInt	$\beta_{NSigInt}$	0.542
	LnVol	β_{LnVol}	0.599
	LnLen	β_{LnLen}	0.852
Speed	Intercept	α_0	-0.471
	PSL	α_{PSL}	0.327
	Bike1Yes	$\alpha_{Bike1Yes}$	1.123
	Develop2	$\alpha_{Develop2}$	-0.870
	RoadSurf	$\alpha_{RoadSurf}$	-0.044
	ScZn1yes	$A_{ScZn1yes}$	1.497
	RU_F_rev	$\alpha_{RU\_F\_rev}$	1.189
	LnVol	α_{LnVol}	-0.770
	LnLen	α_{LnLen}	-2.030
Model fit	BIC		5568.8

Notes: 1. LnVol=Log(Vol_Day), LnLen=Log(Len_mi); 2. Cells are highlight in light gray when the p-value is between 0.05 and 0.1.; 3. Cells are highlighted in dark gray when the p-value is less than 0.05.

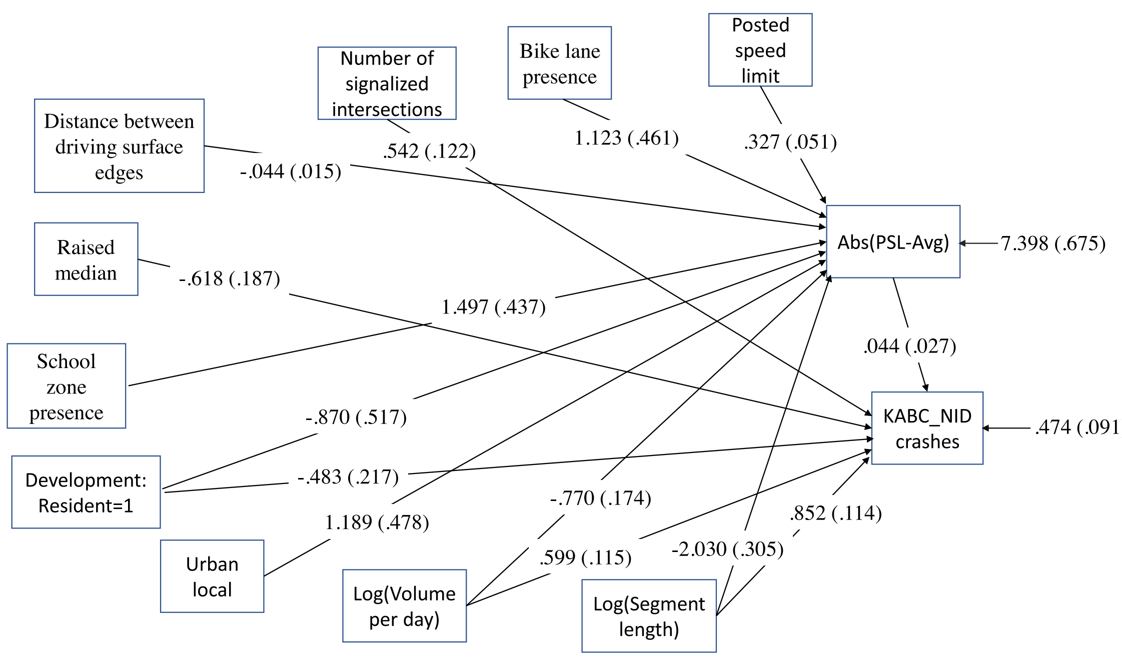

Fig. 3. Path analysis outcomes for segments with posted speed limits of 20 to 45 mph. Standard errors are given in parentheses.

intersections, traffic volume, and segment length, which are all found to be positively correlated with crashes as expected. Although the indirect effects of other roadway characteristic variables vary depending on the mediator variable in the model, it is worth noting that the posted speed limit (PSL) has a statistically significant positive indirect effect on KABC crash frequency through mediator variables absolute value of difference between the posted speed limit and average speed (Abs(PSL − Avg)), posted speed limit minus average speed (PSL − Avg), posted speed limit minus 85th percentile (PSL − S85), 85th percentile speed minus average speed (S85−Avg), average speed (SpdAve), or standard deviation of speed (StdSpd). Several of the speed measures considered (PSL, Coefficient of Variation, StdSpd, Pace, Abs(PSL − Avg), PSL − Avg, PSL − S85, and S85−Avg) showed statistically significant (at $\alpha = 0.05$) effects on crashes with injuries. Especially, Abs(PSL − Avg) provides a very intuitive and useful interpretation, suggesting association with lower KABC crashes when the average speed is closer to PSL for city streets.

Another important implication from this research is the confirmation of the relation between the speed variability and crash occurrence (unlike Gargoum and El-Basyouny (2016) which observed a counterintuitive direction of the standard deviation of speed with a p-value of 0.088). Stated in another manner, when the range of operating speeds is great within a segment, more KABC crashes occur. This is indicated by positive and statistically significant effects of mediator variables standard deviation and coefficient of variation of speeds in Table 6 and Abs (PSL-Avg), PSL-Avg, PSL-S85, and S85-Avg in Table 7 which can all be viewed as measures of spread/variability or speed differential for the speed distribution. The effects of all the aforementioned mediator variables are consistent with past research and intuition: larger spread/variability in operational speed is indicative of reduced smoothness in operations, with higher potential for speed differentials. That condition, in turn, may result in increased risk of crash occurrence as the results from this work clearly suggest. Another possible explanation is that the associations researchers found could indicate that sites with more speed variability tend to be those with mixed visual cues or prone to ambiguous contextual situations (e.g., wide streets in a residential setting) that may result in different drivers choosing different speeds, and perhaps by doing so a larger proportion of the driving population could be more likely to exceed roadway conditions and thus increase their risk of crashing. Nevertheless, those or other causal interpretations of this study results cannot be confirmed from this work. It needs to be noted that this study provides evidence of association (a suggestive evidence) between the speed variables and crashes, not a conclusive evidence for causal claims.

Future research would include considering two (or more) speed measures simultaneously as mediator variables or introducing latent variables related to other external factors that may affect safety in addition to speed variables into structural equation modeling and examine how the speed-safety relationships are affected. It is also of interest to examine the similarities and differences in speed-safety relationships by crash types and facility types (e.g., speed effects at rural highways vs. urban highways, and for roadway departure, overturn, and head on crashes, just to mention a few) based on path analysis. Finally, employing other crash models such as zero-inflated negative binomial models or rare event logistic regression models to deal with excess zeros in more disaggregated crash data (such as monthly or daily crash data) is another area of interest.

Author contribution statement

The authors confirm contribution to the paper as follows: study conception and design: Kay Fitzpatrick, Eun Sug Park, Subasish Das; data collection: Kay Fitzpatrick, Subasish Das; analysis and interpretation of results: Eun Sug Park, Kay Fitzpatrick, Subasish Das, Raul Avelar; manuscript preparation: Eun Sug Park, Kay Fitzpatrick, Subasish Das, Raul Avelar. All authors reviewed the results and approved the final version of the manuscript.

Funding acknowledgement

This work was supported by the National Cooperative Highway Research Program of the United States [contract number NCHRP Project 17–76].

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

Acknowledgments

This research was conducted under a contract with the National Cooperative Highway Research Program (Project 17-76). The authors appreciate the assistance provided by several Texas A&M students in developing the database used in the analysis and the City of Austin for collecting the wealth of speed data that was critical for this evaluation. The authors also appreciate the comments and feedback provided by Mike Pratt, Karen Dixon, and members of the 17-76 panel during this study.

References

Aarts, L., Van Schagen, I., 2006. Driving speed and the risk of road crashes: a review. Accid. Anal. Prev. 38 (2), 215–224.
Banihashemi, M., Dimaiuta, M., Zineddin, A., Spear, B., Smadi, O., Hans, Z., 2019. Using linked SHRP2 RID and NPMRDS data to study speed-safety relationships on Urban interstates and Major arterials. In: In the Proceeding of the 98th TRB Annual Meeting. Washington, DC.
Baruya, A., 1998. Speed-accident relationships on european roads. Proceedings of the 9th International Conference on Road Safety in Europe.
Bellini, S., Cardinali, M.G., Grandi, B., 2017. A structural equation model of impulse buying behaviour in grocery retailing. J. Retail. Consum. Serv. 36, 164–171.
Bollen, K.A., 2014. Structural Equations With Latent Variables. John Wiley & Sons, New York.
Choudhary, P., Pawar, N.M., Velaga, N.R., Pawar, D.S., 2020. Overall performance impairment and crash risk due to distracted driving: a comprehensive analysis using structural equation modelling. Transp. Res. Part F Traffic Psychol. Behav. 74, 120–138.
Das, S., Geedipally, S., Fitzpatrick, K., 2020. Inclusion of speed and weather measures in safety performance functions for rural roadways. IATSS Res. https://doi.org/10.1016/j.iatssr.2020.05.001.
Davis, G.A., 2002. Is the claim that 'variance kills' an ecological fallacy? Accid. Anal. Prev. 34, 343–346.
Davis, G.A., 2004. Possible aggregation biases in road safety research and a mechanism approach to crash modeling. Accid. Anal. Prev. 36, 1119–1127.
Ding, N., Jiao, N., Zhu, S., Liu, B., 2019. Structural equations modeling of real-time crash risk variation in car-following incorporating visual perceptual, vehicular, and roadway factors. Accid. Anal. Prev. 133, 105298.
Dutta, N., Fontaine, M.D., 2019. Improving freeway segment crash prediction models by including disaggregate speed data from different sources. Accid. Anal. Prev. 132, 105253.
Elvik, R., Christensen, P., Amundsen, A., 2004. Speed and road accidents. An Evaluation of the Power Model. TØI Rep. 740, Institute of Transport Economics, Oslo.
Fildes, B.N., Rumbold, G., Leening, A., 1991. Speed Behaviour and Drivers' Attitude to Speeding. Monash University Accident Research Centre, Report, 16, Victoria, Australia.
Fitzpatrick, K., Das, S., Gates, T., Park, E.S., Pratt, M.P., Dixon, K., Kay, J., Chakraborty, M., 2021. Development of the Posted Speed Limit Setting Procedure and Tool. Web-Only NCHRP 17-76 Report (forthcoming in 2021). National Cooperative Highway Research Program, Washington, DC.
Florida Department of Transportation Traffic Engineering & Operations Office, 2018. Speed Zoning for Highways, Road & Streets in Florida. Topic #750-010-002. Accessed December 18, 2019. https://fdotwww.blob.core.windows.net/sitefinity/docs/default-source/traffic/speedzone/2019-01-28_speed-zoning-manual_august-2018.pdf?sfvrsn=ac20bad7_0.
Gargoum, S.A., El-Basyouny, K., 2016. Exploring the association between speed and safety: a path analysis approach. Accid. Anal. Prev. 93, 32–40.
Gitelman, V., Doveh, E., Bekhor, S., 2017. The relationship between free-flow travel speeds, infrastructure characteristics and accidents, on single-carriageway roads. Transp. Res. Procedia 25, 2026–2043.
Hagger, M.S., Polet, J., Lintunen, T., 2018. The reasoned action approach applied to health behavior: role of past behavior and tests of some key moderators using meta-analytic structural equation modeling. Soc. Sci. Med. 213, 85–94.
Hauer, E., 1971. Accidents, overtaking and speed control. Accid. Anal. Prev. 3 (1), 1–13.
Hauer, E., 2005. Cause and Effect in Observational Cross Section Studies on Road Safety. Draft of March 1, 2005.
Hollett, R.C., Gignac, G.E., Milligan, S., Chang, P., 2020. Explaining lecture attendance behavior via structural equation modeling: self-Determination Theory and the Theory of Planned Behavior. Learn. Individ. Differ. 81, 101907.
Hutton, J.M., Cook, D.J., Grotheer, J., Conn, M., 2020. Research Utilizing SHRP2 Data to Improve Highway Safety: Development of Speed—Safety Relationships (No. FHWA-HRT-20-035). United States. Federal Highway Administration. Office of Safety Research and Development, McLean, Virginia.
Imprialou, M.I.M., Quddus, M., Pitfield, D.E., Lord, D., 2016. Re-visiting crash–speed relationships: a new perspective in crash modelling. Accid. Anal. Prev. 86, 173–185.
Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90 (430), 773–795.
Krishnan, S., Agrawal, K., Tryon, R.R., Welch, L.C., Horn, W.F., Newman, J.W., Keim, N. L., 2018. Structural equation modeling of food craving across the menstrual cycle using behavioral, neuroendocrine, and metabolic factors. Physiol. Behav. 195, 28–36.
Lee, J., Chae, J., Yoon, T., Yang, H., 2018. Traffic accident severity analysis with rain-related factors using structural equation modeling – a case study of Seoul City. Accid. Anal. Prev. 112, 1–10.
Maycock, G., Brocklebank, P.J., Hall, R.D., 1998. Road Layout Design Standards and Driver Behaviour. TRL REPORT 332, Transportation Research Laboratory, Berkshire, England.
Mokarami, H., Alizadeh, S.S., Rahimi Pordanjani, T., Varmazyar, S., 2019. The relationship between organizational safety culture and unsafe behaviors, and accidents among public transport bus drivers using structural equation modeling. Transp. Res. Part F Traffic Psychol. Behav. 65, 46–55.

Muthén, L.K., Muthén, B.O., 2017. Mplus User's Guide. Eighth Edition. Los Angeles, CA, 1998.
Nilsson, G., 2004. Traffic Safety Dimensions and the Power Model to Describe the Effect of Speed on Safety. Department of Technology and Society, Lund Institute of Technology, Lund, Sweden.
Özgür, H., 2020. Relationships between teachers' technostress, technological pedagogical content knowledge (TPACK), school support and demographic variables: a structural equation modeling. Comput. Human Behav. 112, 106468.
Pei, X., Wong, S.C., Sze, N.N., 2012. The roles of exposure and speed in road safety analysis. Accid. Anal. Prev. 48, 464–471.
Quddus, M., 2013. Exploring the relationship between average speed, speed variation, and accident rates using spatial statistical models and GIS. J. Transp. Saf. Secur. 5 (1), 27–45.
Seddig, D., Lomazzi, V., 2019. Using cultural and structural indicators to explain measurement noninvariance in gender role attitudes with multilevel structural equation modeling. Soc. Sci. Res. 84, 102328.
Solomon, D.H., 1964. Accidents on Main Rural Highways: Related to Speed, Driver, and Vehicle. FHWA, U.S. Department of Transportation, Washington, DC.
Taylor, M., Lynam, D., Baruya, A., 2000. The Effects of Drivers' Speed on the Frequency of Road Accidents. Crowthorne: Transport Research Laboratory Berkshire, England.
Tehrani, H.D., Yamini, S., 2020. Parenting practices, self-control and anti-social behaviors: meta-analytic structural equation modeling. J. Crim. Justice 68, 101687.
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. Saf. Sci. 57, 264–275.
Wang, X., Zhou, Q., Quddus, M., Fan, T., 2018. Speed, speed variation and crash relationships for urban arterials. Accid. Anal. Prev. 113, 236–243.
Yu, R., Abdel-Aty, M., Ahmed, M., 2013. Bayesian random effect models incorporating real-time weather and traffic data to investigate mountainous freeway hazardous factors. Accid. Anal. Prev. 50, 371–376.
Yu, R., Quddus, M., Wang, X., Yang, K., 2018. Impact of data aggregation approaches on the relationships between operating speed and traffic safety. Accid. Anal. Prev. 120, 304–310.