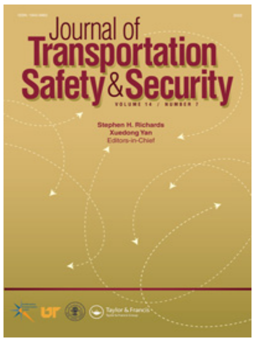

Taylor & Francis
Taylor & Francis Group

Journal of Transportation Safety & Security

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/utss20

Impact of operating speed measures on traffic crashes: Annual and daily level models for rural two-lane and rural multilane roadways

Subasish Das, Eun Sug Park & Sobhan Sarkar

To cite this article: Subasish Das, Eun Sug Park & Sobhan Sarkar (2022): Impact of operating speed measures on traffic crashes: Annual and daily level models for rural two-lane and rural multilane roadways, Journal of Transportation Safety & Security, DOI: 10.1080/19439962.2022.2098441

To link to this article: https://doi.org/10.1080/19439962.2022.2098441

Published online: 15 Jul 2022.

JOURNAL OF TRANSPORTATION SAFETY & SECURITY
https://doi.org/10.1080/19439962.2022.2098441

Impact of operating speed measures on traffic crashes: Annual and daily level models for rural two-lane and rural multilane roadways

Subasish Das[a], Eun Sug Park[a], and Sobhan Sarkar[b]

[a]Texas A&M Transportation Institute, San Antonio, Texas, USA; [b]Information Systems & Business Analytics, Indian Institute of Management Ranchi, Ranchi, Jharkhand, India

ABSTRACT
A significant association between crash severity and operating speed is known to exist. However, the findings related to the speed-crash association are inconclusive. Some studies found that higher speed is associated with a higher number of crashes, whereas other studies found the opposite result. Some of the critical issues in this research problem result from study design, the definition of operating speed measures, types and granularity of operating speed measures, spatial correlation, and design standards of different roadway facilities. The road safety profession will benefit greatly from informative research on the impact of vehicle operating speed, roadway design elements, and traffic volume on crash outcomes. This study investigated the speed-crash association in both annual and daily level datasets to determine how roadway characteristics interact with various speed measures to impact the likelihood of crash occurrences on both annual and daily levels. For annual models, the average operating speed is positively associated with both fatal and injury and property damage only (PDO) crashes. However, for daily models, this association is mostly negative and insignificant. The standard deviation of operating speed is positively associated with crash occurrences for both daily and annual models. The findings of this study can provide additional insights into the speed-crash association literature.

KEYWORDS
operation speed; rural two-lane; rural multilane; path analysis

1. Introduction

Vehicle operating speed influences both crash frequency and severity. Vehicles being driver over the posted speed limit (PSL) on the designated roadways is defined as speeding. According to the Fatality Analysis Reporting System (FARS), the frequency of fatal crashes and the percentages of speeding-related fatal crashes have remained relatively constant between 1990 and 2006 (National Highway Traffic Safety Administration

CONTACT Subasish Das s-das@tti.tamu.edu Texas A&M Transportation Institute, 3500 NW Loop 410, San Antonio, TX 78229, USA

(NHTSA),), 2020). However, there is an agreement that crash severity increases with higher operating speeds. Understanding driver's speed choices based on different roadway geometries has been an active research area for many years. The surroundings of the roadway and spatio-temporal scenarios while driving provide indications to drivers that influence their speed selections. However, due to the dynamic nature of operational speed, a comprehensive analysis of speed and safety requires control of a wide range of conditions in addition to the geometry of the facility, such as traffic volume and density of the flow, before conclusions about safety can be drawn and defended.

There is a need, therefore, to use data-driven procedures to better understand the relationship between motor vehicle travel speed, roadway geometry, and operational measures such as traffic volumes on crash outcomes. Conventional assessments are mostly dependent on the corridor traffic volume and physical site characteristics (American Association of State Highway and Transportation Officials (AASHTO),), 2010). It is generally expected that the operating speed of a vehicle during a crash affects the injury severity of crash victims and that the speed differential between vehicles impacts the crash frequency potential. It is generally anticipated that speed plays a major role in traffic safety. However, the speed-crash association is still inconclusive. To address this critical research gap, there is an urgent need to analyze new data and gain an understanding of how to efficiently measure highway safety in different temporal durations such as annual and daily levels.

The lack of reliable operating speed data, however, has limited the scope for developing this relationship, especially on rural roadways. This study used two national databases: 1) the National Performance Management Research Data Set (NPMRDS) version 1, and 2) The Highway Safety Information Systems (HSIS), a cooperative endeavor funded by FHWA, which is a roadway-based system that provides quality data on a large number of crashes and roadway and traffic variables from a group of selected states. This study developed a conflated dataset using these two databases to quantify the association between crash outcomes and speed, volume, and roadway geometrics.

2. Literature review

Many previous studies explored the speed-crash association issue. Some studies showed that operating speed is positively associated with traffic crashes, while other studies found the opposite association. The current literature review is limited to the studies associated with speed-crash relationship.

Hauer (2009) demonstrated the evolution of speed and the effect of implementing safety measures on speed evolution. He explored the effect of speed on crash likelihood, severity, and highway safety. While this study observed a clear positive correlation between speed and severity, no such relationship was obtained between speed and crash probability, leaving room for future studies. Elvik (2014) studied the relationship between road safety and traffic speed using the power and exponential model. Furthermore, from a review conducted by the author, it was found that the speed dispersion increase was positively correlated with higher crash risk. Anastasopoulos and Mannering (2016) surveyed drivers about their speed choices under different PSLs and identified several factors (age and gender of drivers, their education level, household income, and licensure age) governing drivers' speed choices.

Imprialou, Quddus, Pitfield, and Lord (2016) explored the speed-crash relationship using a precondition-based (data aggregation in terms of their pre-crash traffic and geometric condition) approach instead of the traditional link-based (data aggregation based on segment properties) approach. The authors demonstrated that the link-based approach led to an unexpected negative relationship between speed and crash frequency, whereas the proposed approach gave a positive relationship. By performing a meta-analysis of 108 research papers, Musicant, Bar-Gera, and Schechtman (2016) found that changing the speed limit slightly changes the average observed speed in the same direction. Furthermore, the authors found that speed limit changes were nearly proportional to the average safety changes on interurban roads. Dimaiuta (2018) attempted to study vehicle speed and its effect on safety using the linked Roadway Information Database (RID) and the National Performance Management Research Data Set (NPMRDS). For this purpose, the authors devised a methodology for linking the databases followed by data preparation and a two-phase analysis. In the first phase, the effect of PSL and degree of curvature on the observed vehicle speed was evaluated. In the second phase, the effect of the difference between posted and observed speeds on crash severity was examined. For the first phase, the authors found that as the sharpness of the curve and PSL increases the drivers tend to decrease their speed. For the second phase, however, a counterintuitive result was obtained. The results indicated that as the difference between the PSL and the operating speed increases, the crash severity decreases.

Dimitriou, Stylianou, and Abdel-Aty (2018) estimated the rear-end collision potential and various factors affecting it using multinomial logit regression. The authors found that 66.4% of the total observations had rear-end collision potential. Also, it was shown that with an increase in speed deviation and traffic flow there was an increase in the potential for rear-end collisions and hence a decrease in safety. van Schagen, et al. (2018) stated that speeding negatively affects road safety by increasing both crash frequency and crash severity. The researchers further discussed a combination of measures to address the problem of speed exceedance such as public education, use of dynamic speed limits, speed enforcement, and road design. Dutta and Fontaine (2019) showed how the inclusion of speed data in crash prediction models increased their accuracy by 10-20%. Fitzpatrick and Das (2019) studied the effect of vehicle and bicycle volume on the average vehicle operating speed. The papers also explored the possibility of including crowdsource data for the estimation of unconstrained vehicle speed. The authors indicated that the crowdsource data closely matched the on-tube collected data in non-congested periods and weekends. The similarity of the two databases was also found to be hugely dependent on the signal and driveway numbers. In the study conducted by Gonzales, Fontaine, and Dutta (2019), the effect of installing variable speed limit (VSL) systems on operational speed in foggy weather was observed. The authors found from their preliminary analysis that the reduction of speed limits via VSL systems reduced the number of crashes and decreased crash severity as greater compliance existed between operating speed and the speed limit.

Das and Geedipally (2020) determined crash probability on rural roads by including operating speeds and weather data which was acquired by fusion of multiple big data sources. The authors found that as speed variability increased, crash frequency generally increased, whereas no such clear general association between operating speed and crash count was observed. Ederer, Rodgers, Hunter, and Watkins (2020) used probe speed data to establish relationships between speed and crash parameters. The authors also proposed the difference between using the 85th percentile and the median speed for calculations of speed dispersion, as the traditional speed dispersion formula treats higher speeds and lower speeds similarly. The authors demonstrated that their proposed speed dispersion formula was able to better estimate road safety. Instead of using accident data from the past, Hossain and Medina (2020) predicted severe or fatal crash frequency based on road and traffic characteristics. The developed methodology indicated that crash frequency increased with an increase in speed and traffic flow. Hutton (2020) examined the relationship between operating speed and crash statistics. The researchers found that several roadway characteristics influence operating speed which in turn affects crash frequency. Newnam, Mulvihill, and Muir (2020) stated that incident responders serving on highspeed roads have a higher probability of getting severely injured and hence aimed at providing safe movements to them by developing a training manual. Pratt, Geedipally, and Le (2020) developed a model for estimating vehicle operating speed on four-lane rural highway horizontal curves.

It is important to note that many recent studies addressed common barriers such as unobserved heterogeneity and homeostasis issues by using advanced statistical models (Ahmad, Ahmed, Wali, & Saeed, 2020; Chen, Saeed, & Labi, 2017; Chen, Saeed, Alinizzi, Lavrenz, & Labi, 2019; Saeed, Qiao, Chen, Gkritza, & Labi, 2017; Saeed, Hall, Baroud, & Volovski, 2019; Saeed, Burris, Labi, & Sinha, 2020a; Saeed, Nateghi, Hall, & Waldorf, 2020b; Waseem, Ahmed, & Saeed, 2019; Volovski, Grillo, Varga, Saeed, & El-Hakim, 2021). For example, Saeed et al. (2019) developed both uncorrelated and correlated random parameter count models to predict traffic crash frequencies on multilane highways considering. Waseem et al. (2019) conducted an empirical assessment of factors affecting motorcyclists' injury severities in Pakistan using random parameters logit model with heterogeneity in means and variances. In the recent years, researchers have been using different advanced data mining techniques to identify the key contributing factors or patterns of key contributing factors to solve difference roadway safety issues (Rahman, Sun, Das, & Khanal, 2021; Carney, Harland, & McGehee, 2018; Das, 2021; Das et al., 2022, Das, Wei, Kong, & Xiao, 2021d, Das, Kong, Lavrenz, Wu, & Jalayer, 2021c, 2021b, 2021a, Das, Ashraf, Dutta, & Tran, 2020a, Das, Dutta, & Sun, 2020b, Das, Dutta, Kong, & Sun, 2019, Das, Mudgal, Dutta, & Geedipally, 2018; Kong, Das, Zhou, & Zhang, 2021a, 2021b; Liu et al., 2015; Wen, Zhang, Sun, Wang, & Xu, 2019; Wu, Dadashova, Geedipally, Pratt, & Shirazi, 2021)

The literature review shows that the speed crash association is not consistent. This is partly due to the nature of the study designs, the definitions of operating speed measures, types and granularity of operating speed measures, spatial correlation, and design standards of different roadway facilities. This study used large-scale granular (5-minute) operating speed data and developed a rigorous database by conflating two linear networks. Additionally, both annual and daily level models have been developed to examine the temporal effect of speed measures on crash occurrences.

3. Methodology

3.1. Data sources

This study used two databases (NPMRDS and HSIS) of Washington in this analysis. The HSIS data is a multi-state safety database for a select group of states. It contains roadway inventory, traffic volume, and crash data. The NPMRDS data includes 5-minute intervals of probe vehicle-based travel time data (for passenger vehicles, trucks, and both passenger vehicles and trucks) for all national highway system (NHS) facilities. 'Version 1′, or

'HERE NPMRDS', is the first version of the NPMRDS and is used in this study. This dataset contains a database file and a static GIS file with a common identifier known as a Traffic Message Channel (TMC) or TMC segment. The final database contains a set of files with information about the average travel times of passenger vehicles, trucks, and both passenger vehicles and trucks to identify roadways geo-referenced to TMC segment IDs.

3.2. Data conflation

To develop the final dataset, this researcher conflated the HSIS roadway network with the NPMRDS network. As the NPMRDS contains roadway and speed information in both directions, the centerlines of the HSIS network have been conflated in each of the directions of the NPMRDS. As segment length and start/end-points of segments for these two networks vary widely, a lengthy conflation procedure has been applied. As NPMRDS is considered the base network, some of the long NPMRDS segments have also been included in the model. Additional re-segmentation of the long segments was not conducted in this study. After conflating the networks, 5-minute interval operating speed measures were converted into annual and daily level speed measures. The following four speed measures were considered in the analysis:

- Average operating speed (annual level): Average of all 5-minute interval operating speeds in 2015
- The standard deviation of operating speed (annual level): Standard deviation of all 5-minute interval operating speeds in 2015
- Average operating speed (daily level): Average of all 5-minute interval operating speeds on each day of 2015
- The standard deviation of operating speed (annual level): Standard deviation of all 5-minute interval operating speeds on each day of 2015

3.3. Descriptive statistics

Table 1 displays the descriptive statistics of rural two-lane roadways for annual and daily crash data, respectively. Table 2 displays the descriptive statistics of rural multilane roadways for annual and daily crash data, respectively. The average operating speeds on rural two-lane roadways are usually lower than the average operating speeds of rural multilane roadways. On the other hand, the standard deviation of the operating speed is higher than the standard deviation of the operating speed of rural multilane roadways. The summary statistics also show that rural two-lane had some

Table 1. Summary statistics of rural two-lane roadways.

Variables	Code	Mean	SD	Min	Max
Annual level data					
Fatal and Injury Crashes per segment	FI	1.26	1.76	0	12
PDO Crashes per segment	PDO	3.00	3.00	0	24
Segment Length (mi.)	Length	5.11	3.97	0.02	25.21
Annual Average Daily Traffic (vehicle per day)	Aadt	5818	4490	5	26493
Posted Speed limit (mph)	PSL	30	70	30	70
Surface Width (ft.)	Surfwid	24.54	3.98	20.00	66.79
Shoulder Width (ft.)	shdwid	5.97	2.06	0.00	10.62
Average Speed (mph) [annual]	spdavg	47.23	10.99	4.52	65.44
Standard Dev. of Operating Speeds (mph) [annual]	spdavgsd	8.53	2.99	1.75	18.30
Daily level data					
Fatal and Injury Crashes per segment	FI	0.00	0.06	0	2
PDO Crashes per segment	PDO	0.00	0.00	0	4
Segment Length (mi.)	length	5.11	3.97	0.02	25.21
Annual Average Daily Traffic (vehicle per day)	aadt	5818	4487	5	26493
Posted Speed limit (mph)	PSL	30	70	30	70
Surface Width (ft.)	surfwid	24.54	3.98	20	66.79
Shoulder Width (ft.)	shdwid	5.97	2.06	0	10.62
Average Speed (mph) [daily]	spdavg	47.87	11.18	5.23	80.24
Standard Dev. of Operating Speeds (mph) [daily]	spdavgsd	8.49	4.45	0	54.72

Table 2. Summary statistics of rural multilane roadways.

Variables	Code	Mean	SD	Min	Max
Annual level data					
Fatal and Injury Crashes per segment	FI	1.68	2.09	0	11
PDO Crashes per segment	PDO	4.00	5.00	0	26
Segment Length (mi.)	Length	3.88	2.98	0.02	12.03
Annual Average Daily Traffic (vehicle per day)	Aadt	18492	12600	5	77827
Posted Speed limit (mph)	PSL	30	70	30	70
Surface Width (ft.)	Surfwid	49.39	5.81	29.32	76.00
Shoulder Width (ft.)	shdwid	6.14	2.15	0.00	8.00
Median Width (ft.)	medwid	46.21	34.20	0.00	150.00
Divided	divided	0.81	0.40	0.00	1.00
Average Speed (mph) [annual]	spdavg	51.95	12.63	14.53	62.97
Standard Dev. of Operating Speeds (mph) [annual]	spdavgsd	6.57	2.95	3.11	15.87
Daily level data					
Fatal and Injury Crashes per segment	FI	0.00	0.07	0	4
PDO Crashes per segment	PDO	0.00	0.00	0	4
Segment Length (mi.)	length	3.88	2.96	0.02	12.03
Annual Average Daily Traffic (vehicle per day)	aadt	18492	12550	5	77827
Posted Speed limit (mph)	PSL	30	70	30	70
Surface Width (ft.)	surfwid	49.39	5.78	29.32	76.00
Shoulder Width (ft.)	shdwid	6.14	2.14	0.00	8.00
Median Width (ft.)	medwid	46.21	34.06	0.00	150.00
Divided	divided	0.81	0.40	0.00	1.00
Average Speed (mph) [daily]	spdavg	52.55	12.36	3.23	71.82
Standard Dev. of Operating Speeds (mph) [daily]	spdavgsd	6.49	3.67	0.00	51.58

long segments (max = 25.21 miles) compared to rural multilane roadways (max = 12.03 miles).

3.3.1. Path analysis

While speed is presumed to be an important variable that affects crashes, other variables may also affect crashes (e.g., roadway type, roadway geometry, traffic, PSL, etc.). Some of those variables, such as PSL, may directly affect operating speed and subsequently affect crashes indirectly through operating speed. Other variables, such as traffic volume, may directly affect both the operating speed and crashes. The multiple relationships among variables cannot be described by a single equation, but rather a set of equations describing those relationships is needed.

Path analysis is a methodology for analyzing systems of structural equations (Bollen, 2014). Path analysis can be considered as an extension of multiple regression. The modeling framework can examine the chains of influence among the selected variables to determine whether the data are consistent with the model. It is a holistic approach that enables the simultaneous estimation of multiple relationships among crash, speed, and other roadway characteristic variables. Through path analysis, it is possible to estimate not only the effects of variables directly affecting crashes (direct effects) but also the effects of variables indirectly affecting crashes (indirect effects) through an intervening variable (mediator variable) such as speed. Although path analysis possess certain advantages, it has been less explored in the transportation safety analysis. The studies by Gargoum and El-Basyouny (2016) and Park, Fitzpatrick, Das, and Avelar (2021) are among the handful of studies that employed the path analysis approach in the safety analysis. Path analysis models in this study consist of two sub-models:

- Crash model providing the relationship between crash counts (outcome variable) and operating speed measures (mediator variables) as well as other road geometric variables (independent variables).
- Speed model providing the relationship between operating speed measures and other roadway characteristic variables (including PSL).

Different crash models were employed for modeling annual crashes and daily crashes. For annual crashes, a negative binomial model with the mean given in Equation (1) was adopted:

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

where μ_i is the expected number of annual crashes that occurred on segment i $(i = 1, \ldots, I)$, m_i is the mediator variable (a speed variable), $X_{1i}, \cdots, X_{Ki}$ are K covariates, and $\beta_0, \beta_m, \beta_1, \ldots, \beta_K$ denote regression coefficients for annual crashes.

Unlike annual crash frequency, daily crash frequency per segment was mostly zeros or ones. Thus, for daily crashes, a crash occurrence (which has a value of one if a crash occurred and zero if no crash occurred), rather than a crash count, at each segment was modeled. The logistic regression model given in Equation (2), which expresses the log-odds of the probability of a crash occurrence as a function of a speed variable and other geometric and operational variables, was adopted for the daily crash data:

$$g(\mathbf{x}) = \ln\left[\frac{P(Y_{it} = 1|\mathbf{x})}{1 - P(Y_{it} = 1|\mathbf{x})}\right]$$
$$= \beta_0 + m_{it}\beta_m + X_{1i}\beta_1 + \cdots X_{K-1,i}\beta_{K-1} + X_{Kit}\beta_K \quad (2)$$

where Y_{it} denotes a crash occurrence on segment i $(i = 1, \ldots, I)$ and day t, m_{it} is the mediator variable (a speed variable), $X_{1i}, \cdots, X_{K-1,i}, X_{K,it}$ are K covariates, and $\beta_0, \beta_m, \beta_1, \ldots, \beta_K$ denote regression coefficients for daily crashes.

For the speed model, a normal linear model given in Equation (3) was employed for both annual and daily level data.

$$m_{it} = \alpha_0 + X_{1i}\alpha_1 + \cdots + X_{Li}\alpha_L + \varepsilon_i. \quad (3)$$

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

This study used two performance measures to describe the model performances. As a performance metric, Akaike's information criteria (AIC) is widely used. The goal of AIC is to choose the model with the fewest parameters, as measured by the negative likelihood penalty in the following equation (Akaike, 1973; Acquah, 2010).$AIC = -2\log p(L) + 2p$ (4)

Where p is the number of model parameters and L is the likelihood under the fitted model. Specifically, AIC is aimed at finding the best approximating model to the unknown true data generating process.

The Bayesian Information Criterion (BIC) is another widely used performance metric. While AIC is based on information theory, the BIC is an estimate of the Bayes factor for two competing models (Schwarz, 1978; Kass & Raftery, 1995). BIC can be expressed as:$BIC = -2\log p(L) + p\log(n)$ (5)

.

Estimation of model coefficients was performed by Mplus version 8.3 (Muthen & Muthen, 2017). For daily crash data, to account for potential correlation in the outcomes obtained for 365 days from the same segment in the estimation, a cluster variable TMC was included in the Mplus analysis. Unobserved heterogeneity is one of the critical issues in crash data modeling. To address this issue, the common approach is the application of cluster analysis. After performing the cluster analysis, we have divided crash based on crash severity to perform path analysis.

4. Results and discussions

4.1. Annual level analysis

The estimated regression coefficients for annual crashes in Equations (1) and (3), having average operating speed as a mediator variable and roadway characteristic variables and standard deviation of operating speed as covariates, are given in Table 3. The mediator, average operating speed, was found to have statistically significant effects on annual crash frequency for all four cases considered (i.e., for fatal and injury crashes and PDO crashes for each of the rural two-lane and rural multi-lane roadways). As expected, traffic volume and segment length also have statistically significant direct effects on crash frequency. Surface width has statistically significant direct effects on rural two-lane fatal and injury crashes and rural multi-lane PDO crashes and statistically significant indirect effects on rural two-lane fatal and injury crashes and rural two-lane PDO crashes. Shoulder width has statistically significant direct effects on rural multi-lane fatal and injury crashes and rural multi-lane PDO crashes while statistically significant indirect effects on crash frequency for all four cases. The standard deviation of operating speed has statistically significant direct effects on rural two-lane fatal and injury crashes, rural two-lane PDO crashes, and rural multi-lane fatal and injury crashes. This finding is in line with Newnam et al. (2020) and Imprialou et al. (2016). Median width and being divided both have statistically significant indirect effects on rural multi-lane fatal and injury crashes and rural multi-lane PDO crashes. Figure 1 shows the path analysis plots developed for annual models.

4.2. Daily level analysis

The estimated regression coefficients for daily crashes in Equations (2) and (3), having average operating speed as a mediator variable and roadway characteristic variables and standard deviation of operating speed as covariates, are given in Table 4. The mediator, average operating speed, was found to have statistically significant effects on the odds of PDO crash occurrence for rural two-lane roadways. Traffic volume and segment length have statistically significant direct effects on the odds of crash occurrence for all four cases considered (i.e., for fatal and injury crashes and PDO crashes for each of rural two-lane and rural multi-lane roadways). Surface width has statistically significant direct effects on rural two-lane fatal and injury crashes and rural multi-lane PDO crashes. Shoulder width has statistically significant direct effects on rural multi-lane fatal and injury crashes and rural multi-lane PDO crashes. The standard deviation of operating speed has statistically significant direct effects on rural two-lane fatal and

Table 3. Modeling results of annual level analysis.

Variables	Rural Two-lane: Fatal and Injury Crashes Estimates (S.E.*)	Rural Two-lane: PDO Crashes Estimates (S.E.)	Rural Multi-lane: Fatal and Injury Crashes Estimates (S.E.)	Rural Multi-lane: PDO Crashes Estimates (S.E.)
Outcome variable: Crash Frequency				
Intercept	−7.702 (0.963)	−7.450 (0.757)	−8.911 (2.210)	−4.324 (1.953)
ln(AADT)	0.805 (0.072)	0.740 (0.060)	0.673 (0.167)	0.659 (0.196)
Length	0.141 (0.013)	0.166 (0.012)	0.222 (0.035)	0.206 (0.042)
PSL	0.001[1] (0.002)	0.007 (0.002)	0.222 (0.035)	0.012 (0.007)
Median width	NA	NA	0.001 (0.004)	0.001 (0.004)
Divided	NA	NA	−0.696 (0.462)	−0.334 (0.463)
Surface width	−0.042 (0.018)	−0.001 (0.012)	−0.024 (0.024)	−0.056[1] (0.031)
Shoulder width	−0.02 (0.027)	−0.021 (0.022)	−0.194 (0.068)	−0.268 (0.084)
Average operating speed	0.018 (0.007)	0.011 (0.006)	0.053 (0.020)	0.034 (0.020)
Standard deviation of operating speed (mph)	0.055 (.019)	0.037 (.016)	0.106 (.053)	0.088 (.055)
Outcome variable: Average operating speed				
Intercept	52.235 (3.708)	52.235 (3.708)	24.375 (11.455)	24.375 (11.455)
ln(AADT)	0.051 (0.358)	0.051 (0.358)	−0.713 (0.800)	−0.713 (0.800)
PSL	0.051 (0.019)	0.051 (0.019)	−0.100 (0.047)	−0.100 (0.047)
Median width	NA	NA	0.136 (0.032)	0.136 (0.032)
Surface width	−0.701 (0.098)	−0.701 (0.098)	0.237 (0.166)	0.237 (0.166)
Shoulder width	1.607 (0.188)	1.607 (0.188)	1.725 (0.464)	1.725 (0.464)
Divided	NA	NA	13.600 (2.569)	13.600 (2.569)
Performance measures				
Loglikelihood	−3521.853	−3881.291	−614.271	−706.365
AIC	7073.706	7792.582	1266.541	1450.729
BIC	7141.843	7860.719	1320.126	1504.315
Sample-size adjusted BIC	7094.215	7813.091	1260.047	1444.236

*S.E.= standard error.
[1]shaded cell indicates that the variable is significant at 90% confidence level.

injury crashes, rural two-lane PDO crashes, and rural multi-lane fatal and injury crashes. This finding is in line with Newnam et al. (2020) and Imprialou et al. (2016). Median width has a statistically significant direct effect on rural multi-lane fatal and injury crashes. Figure 2 shows the path analysis plots developed for annual models.

The general findings from the two temporal level analyses are below:

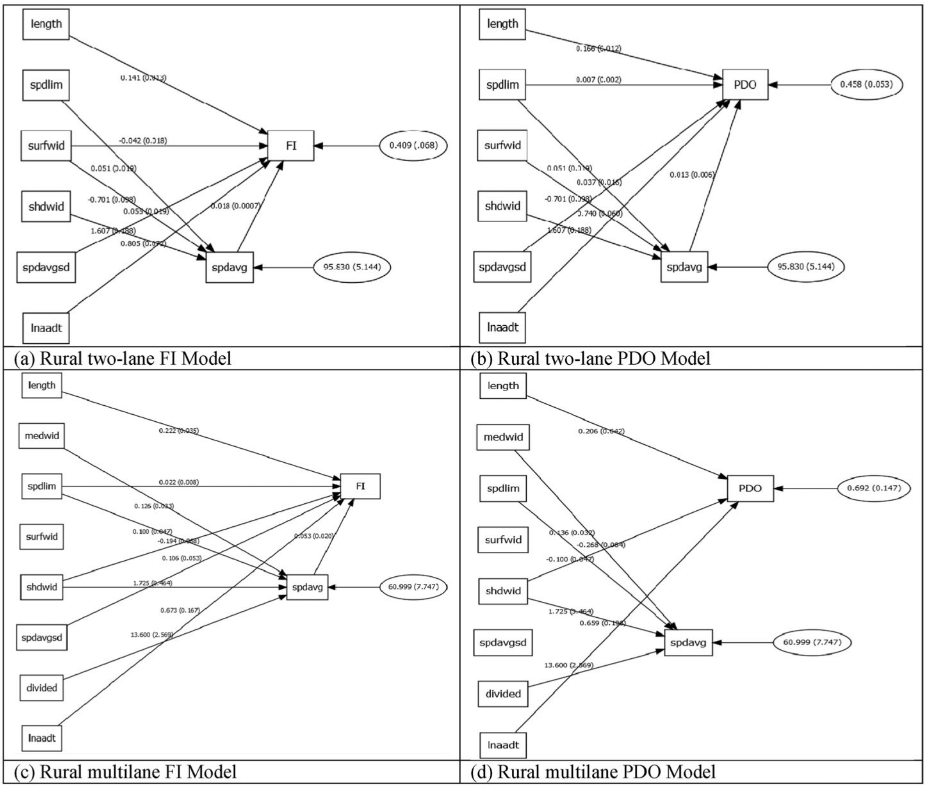

Figure 1. Path analysis plots for annual level analysis.
Note: The arrows with only significant coefficient estimates (at $\alpha = 0.05$) are shown.

- For annual models, the average operating speed measure is positively associated with both fatal and injury and PDO crashes. For daily models, this association is mostly insignificant and negative except for one daily level model (the rural two-lane PDO model). It is important to note that these speed measures are calculated based on their temporal (annual or daily) level. This study shows that the impact of operating speed measure and its calculation based on temporal limit can alter the speed-crash association.
- The standard deviation of operating speed is positively associated with crash frequencies for both annual and daily models.
- By considering average operating speed as a mediator variable, it is found that the impact of traffic volume on operating speed is not statistically significant. This might be due to other confounding factors such as the operating and design properties of these roadways.
- Path analysis is an applicable tool to determine direct and indirect association of factors in a complex dataset. It is found that speed and

Table 4. Modeling results of daily level analysis.

Variables	Rural Two-lane: Fatal and Injury Crashes Estimates (S.E.*)	Rural Two-lane: PDO Crashes Estimates (S.E.)	Rural Multi-lane: Fatal and Injury Crashes Estimates (S.E.)	Rural Multi-lane: PDO Crashes Estimates (S.E.)
Outcome variable: Crash Occurrence (1: Crash, 0: No crash)				
Intercept	−12.262	−10.993	−11.638	−7.758
ln(AADT)	0.745 (0.069)	0.620 (0.057)	0.450 (0.183)	0.457 (0.164)
ln(Length)	0.748 (0.062)	0.868 (0.051)	0.788 (0.129)	0.684 (0.116)
PSL	0.002 (0.002)	0.009 (0.002)	0.013 (0.008)	0.016 (0.010)
Median width	NA	NA	0.009 (0.003)	0.002 (0.003)
Surface width	−0.052 (0.027)	−0.012 (0.019)	−0.020 (0.021)	−0.033 (0.020)
Shoulder width	−0.007 (0.026)	−0.007 (0.022)	−0.218 (0.070)	−0.225 (0.062)
Divided	NA	NA	0.182 (0.555)	0.322 (0.486)
Average operating speed	−0.006 (0.007)	−0.016 (0.006)	0.016 (0.016)	−0.014 (0.012)
Standard deviation of operating speed (mph)	0.064 (0.011)	0.018 (0.008)	0.134 (0.031)	0.049[1] (0.027)
Outcome variable: Average operating speed				
Intercept	57.711 (5.574)	57.711 (5.574)	24.155 (11.096)	24.155 (11.096)
ln(AADT)	−0.567 (0.606)	−0.567 (0.606)	−1.073 (0.802)	−1.073 (0.802)
PSL	0.030 (0.020)	0.030 (0.020)	-.0116 (0.059)	-.0116 (0.059)
Median width	NA	NA	0.124 (0.025)	0.124 (0.025)
Surface width	−0.664 (0.133)	−0.664 (0.133)	0.336 (0.161)	0.336 (0.161)
Shoulder width	1.658 (0.174)	1.658 (0.174)	1.741 (0.528)	1.741 (0.528)
Divided	NA	NA	14.211 (3.082)	14.211 (3.082)
Logistic Regression Odds Ratios: Crash Counts (FI or PDO)				
ln(AADT)	2.106 (0.146)	1.858 (0.106)	1.568 (0.288)	1.580 (0.259)
ln(Length)	2.114 (0.132)	2.381 (0.122)	2.199 (0.284)	1.982 (0.230)
PSL	1.002 (0.003)	1.009 (0.002)	1.013 (0.008)	1.016 (0.010)
Surface width	0.949 (0.025)	0.988 (0.018)	0.980 (0.021)	0.968 (0.020)
Shoulder width	0.993 (0.026)	0.993 (0.022)	0.804 (0.057)	0.799 (0.049)
Divided	NA	NA	1.200 (0.667)	1.380 (0.671)
Median Width	NA	NA	1.009 (0.003)	1.002 (0.003)
Average operating speed	0.994[1] (0.007)	0.984 (0.006)	1.016 (0.016)	0.986 (0.012)
Standard deviation of operating speed (mph)	1.067 (0.012)	1.018 (0.008)	1.144 (0.036)	1.051 (0.028)

(continued)

Table 4. Continued.

Variables	Rural Two-lane: Fatal and Injury Crashes Estimates (S.E.*)	Rural Two-lane: PDO Crashes Estimates (S.E.)	Rural Multi-lane: Fatal and Injury Crashes Estimates (S.E.)	Rural Multi-lane: PDO Crashes Estimates (S.E.)
Performance measures				
Loglikelihood				
AIC	1752344.229	1761352.018	305322.185	307726.643
BIC	1752489.382	1761497.171	305478.534	307882.992
Sample-size adjusted BIC	1752444.889	1761452.678	305421.330	307825.788

S.E.= standard error.
[1]*shaded cell indicates that the variable is significant at 90% confidence level.*

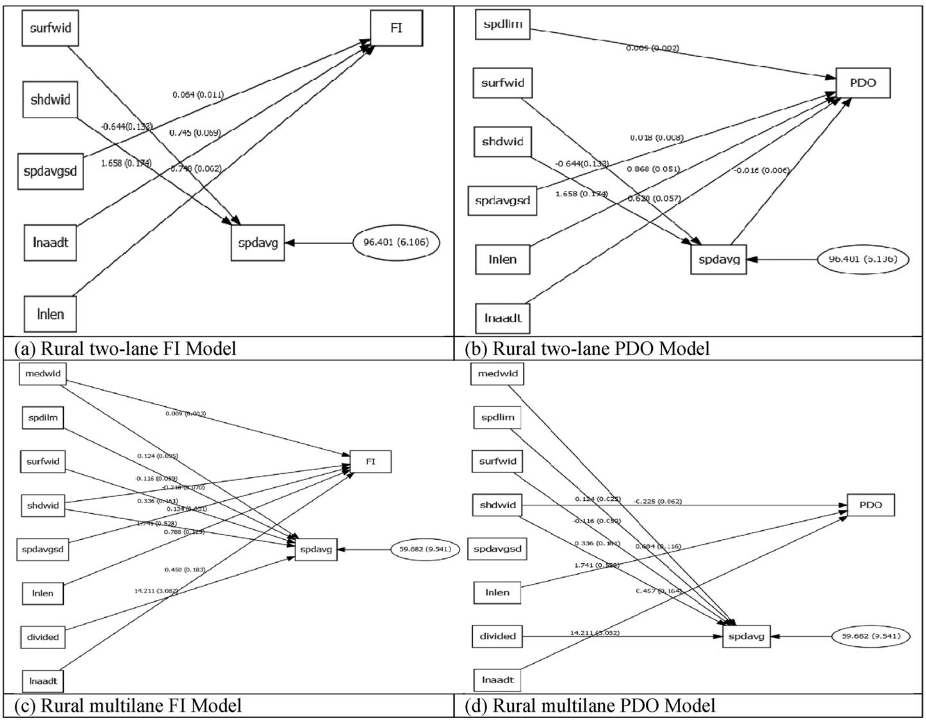

Figure 2. Path analysis plots for daily level analysis.
Note: The arrows with only significant coefficient estimates (at $\alpha = 0.05$) are shown.

associated speed related volatilities (e.g., standard deviation of operating speed) are highly associated with an increase in the probability of crashes, specifically fatal or serious injury crashes. Note thet statistically insignificant variables in the model can be associated with speed related volatilities and they may indirectly be correlated with the intensity outcome.

5. Conclusions

A definite association between operating speed and crash outcomes can help agencies to understand the impact of this important operational measure. Many studies examined this association. However, the findings are not conclusive (Das et al., 2020; Das & White, 2020; Hutton, 2020). It is mostly due to the impact of operating speed measures differ by roadway functional classes and other geometric characteristics. This study used two unique datasets to explore the association between crash and operating speed. Using data from Washington, this study developed path analysis models for rural two-lane and rural multilane roadways for both fatal and injury and PDO crashes at two temporal levels. Previously, Park et al. (2021) conducted a study a similar analysis at segment level using annual level data. The current study found that the average yearly operating speed is positively associated with crash frequencies. However, this association is not statistically significant for most of the models developed for daily level data. Both annual and daily level models show that the standard deviation of operating speed is positively associated with annual and daily crash counts, respectively. The results also show that several geometric features and PSL are associated with operating speed measures. One of the critical operation variables, traffic volume, shows no statistically significant association with the average operating speed. This may be due to the impact of other confounding factors, such as PSL and other geometric variables.

This study has some unique contributions. First, this study develops a data conflation framework that other DOTs and transportation agencies can utilize by conflating state specific roadway inventory network with NPMRDS network. Second, this study examines the impact of operating speed on both annual and daily level crash data. Third, the daily level models can be considered as the short-duration SPFs, which will provide significant application in real-time crash analysis frameworks. Findings of the current study can be incorporated in the ongoing effort by National Academies for developing SPFs for rural two-lane highways that incorporate speed measures (Dixon, 2022). Future update on the posted speed related guidance documents (Fitzpatrick, Das, Pratt, Dixon, & Gates, 2021a; Fitzpatrick, Das, Pratt, Dixon, & Gates, 2021b) can also utilize the findings from this current study.

Note that data aggregation has been done separately by different studies. The data aggregation used in this study is based on spatio-temporal patterns. This study used both annual and daily level model to understand the crash-speed association. There are no prior studies which applied path analysis on short duration (such as daily level) crash data. This is one of the unique contributions of this study. As operating speed is not static variable like geometric variables, innovative modeling techniques and data aggregation are needed to understand the real-world patterns. This study can be considered as one of the first attempts for developing path analysis model using short duration operation speed data.

This study also has several limitations. First, some critical geometric variables such as horizontal curve, vertical curve, curve radius, and roadside hazard ratings are considered to be the base condition. As this study aims to use similar variables for comparing annual and daily models, the static nature of other geometric variables will make many of the additional geometric variables statistically insignificant for daily level models. Second, NPMRDS data has missing values on rural roadways, specially low volume roadways. The current study only includes the available operating speed data to calculate the average operating speed. No imputation method has been applied to reduce the data missingness.

Acknowledgments

We like to thank three anonymous reviewers for their excellent suggestions.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Subasish Das http://orcid.org/0000-0002-1671-2753
Sobhan Sarkar http://orcid.org/0000-0001-6040-6051

References

Acquah, H. D. G. (2010). Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of an asymmetric price relationship. *Journal of Development and Agricultural Economics*, *2*(1), 001–006.

Ahmad, N., Ahmed, A., Wali, B., & Saeed, T. U. (2020). *Exploring factors associated with crash severity on motorways in Pakistan*. In Proceedings of the Institution of Civil Engineers-Transport, December (pp. 1–10). Scotland: Thomas Telford Ltd. doi:10.1680/jtran.18.00032

Akaike, H. (1973). *Information theory and an extension of the maximum likelihood principle* [Paper presentation]. In B. N. Petrov & F. Csaki (Eds.), 2nd International Symposium on Information Theory (pp. 267–81). Budapest: Akademiai Kiado.

American Association of State Highway and Transportation Officials (AASHTO). (2010). *Highway safety manual* (1st ed.). Washington DC.

Anastasopoulos, P. C., & Mannering, F. L. (2016). The effect of speed limits on drivers' choice of speed: A random parameters seemingly unrelated equations approach. *Analytic Methods in Accident Research*, *10*, 1–11. doi:10.1016/j.amar.2016.03.001

Bollen, K. (2014). *Structural equations with latent variables*. New Jersey: John Wiley & Sons.

Carney, C., Harland, K. K., & McGehee, D. V. (2018). Examining teen driver crashes and the prevalence of distraction: Recent trends, 2007–2015. *Journal of Safety Research*, *64*, 21–27. doi:10.1016/j.jsr.2017.12.014

Chen, S., Saeed, T. U., & Labi, S. (2017). Impact of road-surface condition on rural highway safety: A multivariate random parameters negative binomial approach. *Analytic Methods in Accident Research*, *16*, 75–89. doi:10.1016/j.amar.2017.09.001

Chen, S., Saeed, T. U., Alinizzi, M., Lavrenz, S., & Labi, S. (2019). Safety sensitivity to roadway characteristics: a comparison across highway classes. *Accident; Analysis and Prevention*, *123*, 39–50. doi:10.1016/j.aap.2018.10.020

Das, S. (2021). Identifying key patterns in motorcycle crashes: findings from taxicab correspondence analysis. *Transportmetrica A: Transport Science*, *17*(4), 593–614. doi:10.1080/23249935.2020.1802362

Das, S., Dutta, A., & Rahman, M. (2021). Pattern recognition from light delivery vehicle crash characteristics. *Journal of Transportation Safety & Security*. https://doi.org/10.1080/19439962.2021.1995800

Das, S., & Geedipally, S. (2020). Rural speed safety project for USDOT safety data initiative: Findings and outcome. *ITE Journal*, *90*(9), 38–42.

Das, S., & White, L. D. (2020). RuralSpeedSafetyX: Interactive decision support tool to improve safety. *SoftwareX*, *11*, 100493. doi:10.1016/j.softx.2020.100493

Das, S., Ashraf, S., Dutta, A., & Tran, L.-N. (2020a). Pedestrians under influence (PUI) Crashes: Patterns from correspondence regression analysis. *Journal of Safety Research*, *75*, 14–23. doi:10.1016/j.jsr.2020.07.001

Das, S., Dutta, A., & Geedipally, S. R. (2021a). Applying Bayesian data mining to measure the effect of vehicular defects on crash severity. *Journal of Transportation Safety & Security*, *13*(6), 605–621. doi:10.1080/19439962.2019.1658674

Das, S., Dutta, A., & Sun, X. (2020b). Patterns of rainy weather crashes: Applying rules mining. *Journal of Transportation Safety & Security*, *12*(9), 1083–1105. doi:10.1080/19439962.2019.1572681

Das, S., Dutta, A., Kong, X., & Sun, X. (2019). Hit and run crashes: Knowledge extraction from bicycle involved crashes using first and frugal tree. *International Journal of Transportation Science and Technology*, *8*(2), 146–160. doi:10.1016/j.ijtst.2018.11.001

Das, S., Geedipally, S., Avelar, R., Wu, L., Fitzpatrick, K., Banihashemi, M., & Lord, D. (2020). *Rural Speed Safety Project for USDOT Safety Data Initiative*. Washington DC: USDOT.

Das, S., Kong, X. J., Lavrenz, S. M., Wu, L., & Jalayer, M. (2021c). Fatal crashes at highway rail grade crossings: A U.S. based study. *International Journal of Transportation Science and Technology*, *11*(1), 107–117. doi:10.1016/j.ijtst.2021.03.002

Das, S., Mudgal, A., Dutta, A., & Geedipally, S. R. (2018). Vehicle consumer complaint reports involving severe incidents: Mining large contingency tables. *Transportation Research Record: Journal of the Transportation Research Board*, *2672*(32), 72–82. doi:10.1177/0361198118788464

Das, S., Sun, X., Goel, S., Sun, M., Rahman, A., & Dutta, A. (2022). Flooding related traffic crashes: Findings from association rules. *Journal of Transportation Safety & Security*, *14*(1), 111–129. doi:10.1080/19439962.2020.1734130

Das, S., Wei, Z., Kong, X., & Xiao, X. (2021d). Mining crowdsourced data on bicycle safety critical events. *Transportation Research Interdisciplinary Perspectives*, *10*, 100360. doi:10.1016/j.trip.2021.100360

Dimaiuta, M., Banihashemi, M., Spear, S., Hull, R., Beer, P., Smadi, O., & Zach Hans, Z. (2018). *Speed–safety analyses using linked national performance management research* *data set (NPMRDS) and SHRP2 Roadway Information Database (RID) Data*. Washington DC: FHWA.

Dimitriou, L., Stylianou, K., & Abdel-Aty, M. A. (2018). Assessing rear-end crash potential in urban locations based on vehicle-by-vehicle interactions, geometric characteristics and operational conditions. *Accident; Analysis and Prevention*, *118*, 221–235. doi:10.1016/j.aap.2018.02.024

Dixon, K. (2022). Developing safety performance functions for rural two-lane highways that incorporate speed measures. NCHRP Research Project. https://apps.trb.org/cmsfeed/TRBNetProjectDisplay.asp?ProjectID=4574

Dutta, N., & Fontaine, M. D. (2019). Improving freeway segment crash prediction models by including disaggregate speed data from different sources. *Accident Analysis & Prevention*, *132*, 105253. doi:10.1016/j.aap.2019.07.029

Ederer, D. J., Rodgers, M. O., Hunter, M. P., & Watkins, K. E. (2020). Case study using probe vehicle speeds to assess roadway safety in Georgia. *Transportation Research Record: Journal of the Transportation Research Board*, *2674*(11), 554–562. doi:10.1177/0361198120948858

Elvik, R. (2014). Speed and road safety - new models. Institute of Transport Economics (TØI) (Fart og trafikksikkerhet - Nye modeller), Norway.

Fitzpatrick, K., Das, S. (2019). Vehicle operating speed on urban arterial roadways. https://www.vtti.vt.edu/utc/safe-d/wp-content/uploads/2019/04/TTI-01-04_Final-Research-Report.pdf.

Fitzpatrick, K., Das, S., Pratt, M. P., Dixon, K., & Gates, T. (2021a). *Posted speed limit setting procedure and tool: User guide*. Washington DC: Transportation Research Board.

Fitzpatrick, K., Das, S., Pratt, M. P., Dixon, K., & Gates, T. (2021b). *Development of a posted speed limit setting procedure and tool*. Washington DC: Transportation Research Board.

Gargoum, S. A., & El-Basyouny, K. (2016). Exploring the association between speed and safety: A path analysis approach. *Accident; Analysis and Prevention*, *93*, 32–40. doi:10.1016/j.aap.2016.04.029

Gonzales, D. E., Fontaine, M. D., & Dutta, N. (2019). Impact of variable speed-limit system on driver speeds during low-visibility conditions. *Journal of Transportation Engineering, Part A: Systems*, *145*(12), 04019050. doi:10.1061/JTEPBS.0000282

Hauer, E. (2009). Speed and safety. *Transportation Research Record: Journal of the Transportation Research Board*, *2103*(1), 10–17. doi:10.3141/2103-02

Hossain, F., & Medina, J. C. (2020). Effects of operating speed and traffic flow on severe and fatal crashes using the U.S. road assessment program methodology and field data verification. *Transportation Research Record: Journal of the Transportation Research Board*, *2674*(11), 30–41. doi:10.1177/0361198120943587

Hutton, J. M. (2020). Research utilizing SHRP2 data to improve highway safety: Development of speed–safety relationships (p. 110). https://www.fhwa.dot.gov/publications/research/safety/20035/20035.pdf.

Imprialou, M.-I. M., Quddus, M., Pitfield, D. E., & Lord, D. (2016). Re-visiting crash–speed relationships: A new perspective in crash modelling. *Accident; Analysis and Prevention*, *86*, 173–185.

Kass, R. E., & Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795. doi:10.1080/01621459.1995.10476572

Kong, X., Das, S., Zhou, H. "., & Zhang, Y. (2021a). Patterns of near-crash events in a naturalistic driving dataset: Applying rules mining. *Accident Analysis & Prevention*, *161*, 106346. doi:10.1016/j.aap.2021.106346

Kong, X., Das, S., Zhou, H., & Zhang, Y. (2021b). Characterizing phone usage while driving: Safety impact from road and operational perspectives using factor analysis. *Accident; Analysis and Prevention*, *152*, 106012. doi:10.1016/j.aap.2021.106012

Liu, T., Yang, Y., Huang, G.-B., Lin, Z., Klanner, F., Denk, C., & Rasshofer, R. H. *Cluster regularized extreme learning machine for detecting mixed-type distraction in driving*. 18th International IEEE Conference on Intelligent Transportation Systems (ITSC), Spain.

Musicant, O., Bar-Gera, H., & Schechtman, E. (2016). Impact of speed limit change on driving speed and road safety on interurban roads: Meta-analysis. *Transportation Research Record: Journal of the Transportation Research Board*, *2601*(1), 42–49. doi:10.3141/2601-06

Muthen, L., & Muthen, B. (2017). *Mplus user's guide* (8th ed.). Los Angeles, CA.

National Highway Traffic Safety Administration (NHTSA). (2020). Fatality analysis reporting system encyclopedia. Washington DC. https://www-fars.nhtsa.dot.gov/Main/index.aspx

Newnam, S., Mulvihill, C., & Muir, C. (2020). Safety when operating on high-speed roads: Protecting our incident responders. *Safety Science*, *131*, 104910. doi:10.1016/j.ssci.2020.104910

Park, E. S., Fitzpatrick, K., Das, S., & Avelar, R. (2021). Exploration of the relationship among roadway characteristics, operating speed, and crashes for city streets using path analysis. *Accident; Analysis and Prevention*, *150*, 105896. doi:10.1016/j.aap.2020.105896

Pratt, M. P., Geedipally, S. R., & Le, M. (2020). Cross-sectional study of vehicle speeds on rural four-lane highway curves. *Transportation Research Record: Journal of the Transportation Research Board*, *2674*(10), 15–27. doi:10.1177/0361198120939096

Rahman, M. A., Sun, X., Das, S., & Khanal, S. (2021). Exploring the influential factors of roadway departure crashes on rural two-lane highways with logit model and association rules mining. *International Journal of Transportation Science and Technology*, *10*(2), 167–183. doi:10.1016/j.ijtst.2020.12.003

Saeed, T. U., Burris, M. W., Labi, S., & Sinha, K. C. (2020a). An empirical discourse on forecasting the use of autonomous vehicles using consumers' preferences. *Technological Forecasting and Social Change*, *158*, 120130. doi:10.1016/j.techfore.2020.120130

Saeed, T. U., Hall, T., Baroud, H., & Volovski, M. J. (2019). Analyzing road crash frequencies with uncorrelated and correlated random-parameters count models: An empirical assessment of multilane highways. *Analytic Methods in Accident Research*, *23*, 100101. doi:10.1016/j.amar.2019.100101

Saeed, T. U., Nateghi, R., Hall, T., & Waldorf, B. S. (2020b). Statistical analysis of area-wide alcohol-related driving crashes: a spatial econometric approach. *Geographical Analysis*, *52*(3), 394–417. doi:10.1111/gean.12216

Saeed, T. U., Qiao, Y., Chen, S., Gkritza, K., & Labi, S. (2017). Methodology for probabilistic modeling of highway bridge infrastructure condition: Accounting for improvement effectiveness and incorporating random effects. *Journal of Infrastructure Systems*, *23*(4), 04017030. doi:10.1061/(ASCE)IS.1943-555X.0000389

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464.

van Schagen, I., et al. (2018). Speed and speed management. Washington DC: FHWA. https://ec.europa.eu/transport/road_safety/sites/roadsafety/files/pdf/ersosynthesis2018-detail-speedspeedmanagement32_en.pdf.

Volovski, M., Grillo, N., Varga, C., Saeed, T. U., & El-Hakim, M. (2021). Subway ridership: Accounting for regional variation across land-use and socioeconomic settings. *Journal of Infrastructure Systems*, *27*(2), 04021010. doi:10.1061/(ASCE)IS.1943-555X.0000613

Waseem, M., Ahmed, A., & Saeed, T. U. (2019). Factors affecting motorcyclists' injury severities: An empirical assessment using random parameters logit model with heterogeneity in means and variances. *Accident; Analysis and Prevention*, *123*, 12–19.
Wen, F., Zhang, G., Sun, L., Wang, X., & Xu, X. (2019). A hybrid temporal association rules mining method for traffic congestion prediction. *Computers & Industrial Engineering*, *130*, 779–787. doi:10.1016/j.cie.2019.03.020
Wu, L., Dadashova, B., Geedipally, S., Pratt, M. P., & Shirazi, M. (2021). Using naturalistic driving study data to explore the association between horizontal curve safety and operation on rural two-lane highways. *Journal of Transportation Safety & Security*, *13*(8), 896–913. doi:10.1080/19439962.2019.1697776