Rural Speed Safety Project for USDOT Safety Data Initiative:
Findings and Outcomes

By Subasish Das, Ph.D. (M) and Srinivas Geedipally, P.E., Ph.D. (M)

shutterstock/ Sascha Burkard

To save more lives and reduce injuries from roadway crashes, agencies must identify sections of the highways that have an increased risk of crash occurrences. Toward that end, the U.S. Department of Transportation's (USDOT) vision for the Safety Data Initiative (SDI) includes the integration of big data sources as a focus area to enhance the general understanding of crash risks and mitigate future crash occurrences. Current crash estimation or prediction methods, such as those in the first edition of the *Highway Safety Manual* (HSM) use annual average daily traffic (AADT) data along with geometric characteristics to predict the annual average crash frequency of roadway segments and intersections. One inadequacy of the HSM is the limitation of speed-related factors in crash prediction. Recent research has made limited progress in incorporating speed measures (i.e. average daily speed, standard deviation of hourly operating speed) into crash predictive models. To advance the state of the practice, this study begins the work of investigating the association between crash risk and traffic speeds using traffic speed information from big data.

Related Work

Although speed is considered as a major contributing factor of roadway crashes, research findings are inconsistent. Abdel-Aty and Radwan studied speed by capturing the magnitude of speeding relative to the posted speed limit.[1] This speeding indicator variable was shown to affect the crash involvement of male and young drivers. The preliminary analysis of a study conducted by Taylor et al. based in the United Kingdom revealed that the average speed was negatively related to crash frequency.[2] Pei et al. showed that crash risk is negatively associated with mean speed when controlling for distance exposure, which goes with the argument that roadway segments designed for higher speeds should deliver better road safety performance.[3] However, for time exposure, this association was positive. Imprialou et al. showed that speed-crash relation is positive for condition-based approach.[4] However, the outcome of the link-based model is the opposite. Several other studies examined the similar research questions.[5-10] However, the overall findings are not similar in nature. Few recent studies investigated the association between weather and crash outcomes.[11, 12] The literature review reveals that additional research is needed to investigate the association between crash, operating speed, weather, roadway, and traffic measures.

USDOT SDI Pilot Study on Rural Safety

This paper is an abridged version of Das et al. study that developed Safety Performance Functions (SPFs) by using geometric and operational characteristics that include speed measures.[13, 14] SPFs are the statistical "base" models used to estimate the average crash frequency for a facility type with specific base conditions. The research addressed two research questions: 1) do different speed measures contribute to crash outcomes, and 2) is there more variability in speeds just prior to a crash?

This study developed an interactive decision support tool that provides annual expected number of crashes with colored lines based on the number of crashes per year. The transferable framework of this tool was developed using open source R software and its 'Shiny' (an interactive web technology) framework.[14, 15]

Intended Audience

The main audience for this report is practitioners who want to assess risk on rural roadways. Some analytic methods in the report require transportation safety modeling knowledge and skills. However, the tool is easy to use to perform annual risk analysis by the facility types.

Methodology

Data Sources

The databases used for the conflation are: 1) the National Performance Management Research Dataset (NPMRDS), and 2) the Highway Safety Information Systems (HSIS) data. The project team assigned the weather station data from the National Oceanic and Atmospheric Administration (NOAA) on the conflated database. Figure 1 illustrates the overall approach developed by Das et al.[13]

Data Integration

Given the list of the data sources and the purpose of the data analysis, the project team developed conflated datasets by integrating information from different sources. The data integration work has major three steps:

- Conflate the HSIS roadway network data to NPMRDS directional network
- Determine different speed measures by temporal segregation (for example, annual, month or daily)
- Conflate average precipitation (annual and daily) data (from NOAA) to the NPMRDS network

It is important to note that the speed data on TMC segments are recorded by epoch (5-minute bins in the raw NPMRDS data). The project team averaged the annual, monthly, and daily level. For example, monthly average speed (MAS) for a TMC segment can be calculated as:

$$MAS_{epoch_e, TMC_i} = \frac{1}{n}\sum_{n=1}^{31} Speed_{day_n, epoch_e, TMC_i} \quad (1)$$

where:

MAS_{epoch_e, TMC_i} = the average epoch e speed at segment i over a month

n = days in a month

$Speed_{day_n, epoch_e, TMC_i}$ = the NPMRDS speed on day n and epoch e at segment i

Model Development

Certain speed measures incorporated into statistical models were found to be beneficial to quantifying safety risk. It includes all rural facility types, and the procedures developed can be applied to other states contributing to HSIS. Readers are referred to Das et al. study for variable details.[13] The project team developed both long-term (annual-level) and short-term (daily-level) SPFs with the inclusion of speed measures and weather information. For both (annual and daily) modeling frameworks, three roadway facilities are considered for separate SPF development:

- Rural two-lane roadways
- Rural multilane roadways (both divided and undivided)
- Rural interstate roadways

Annual-level analysis

The project team developed the following SPFs (Washington only, Ohio only, and Two-state model) by major facility types using

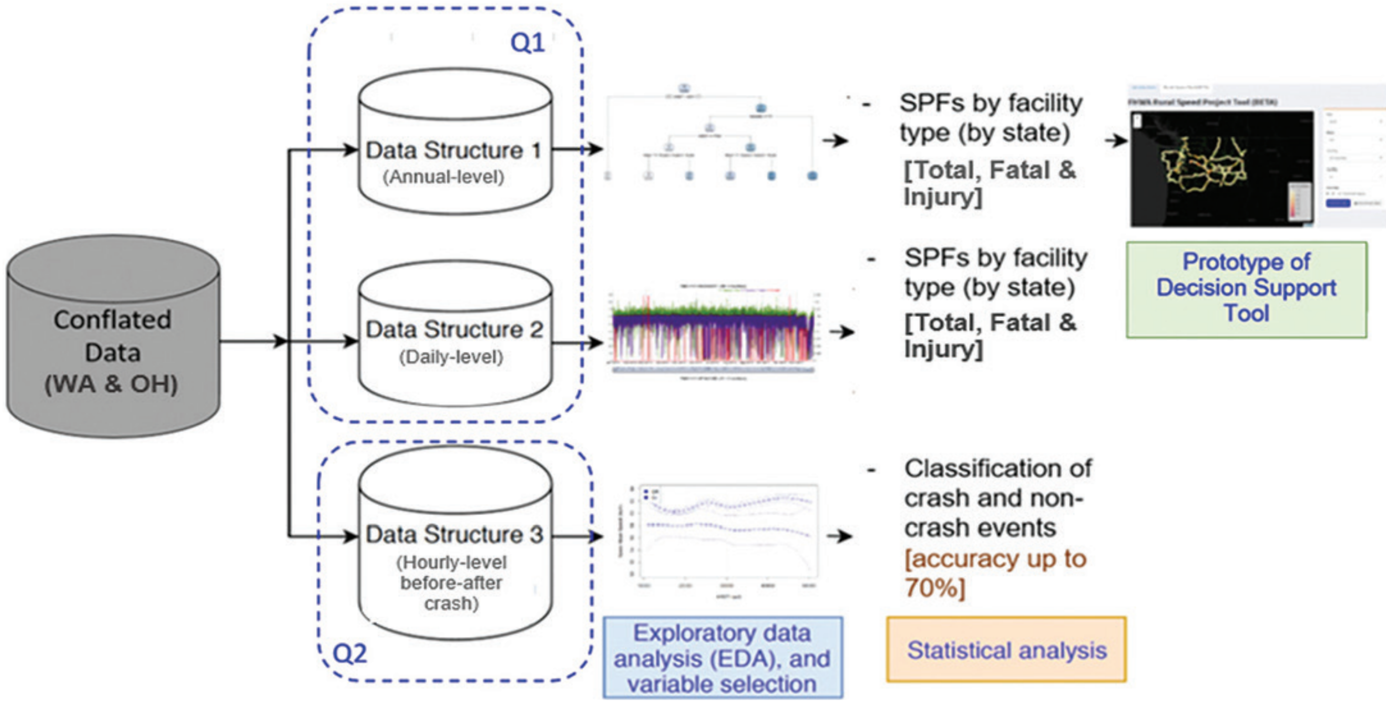


Figure 1. The Framework of "Rural Speed Safety Project."

aggregated annual data for: total (KABCO)* crashes, fatal and injury (KABC) crashes, and property damage only (PDO) crashes.

Certain speed measures were useful for annual crash prediction. It is important to note that the current study did not examine the speed variability between the vehicles as NPMRDS provides aggregated speed measures. Different variable combinations and various model forms were examined to identify the best possible relationship between the number of crashes and independent variables.

Readers can consult Das et al.[13] report for the detailed modeling techniques. Table 1 shows a summary of the impacts on the crash frequency of the variables examined for the developed models. The key findings from annual level analysis are below:

- Increased variability in hourly operating speed within a day and an increase in monthly operating speeds within a year are both associated with increased crashes.
- Multilane, non-freeway roads with higher free-flow speeds are expected to experience a higher crash frequency than those with lower free-flow speeds. However, crash frequency decreases for interstate roadways, which is due to their more robust highway design standards.
- When operating speed difference between weekends and weekdays is greater, all roadway types experienced a higher number of crashes.
- Increased non-peak and non-event speed (average operating speed excluding peak hours and hours with events) is associated with an increase in crash frequencies on rural two-lane roadways. However, the opposite is true for the rural interstate model. This finding for decreased crash frequency on the interstate could be because of high design standards.
- Findings from precipitation-crash association is counter intuitive. Further investigation is needed to examine this finding by including other variables such as pavement condition, water on roadway, and skid number.

Daily-level analysis

Prediction based on annual information limits the SPFs' ability to quantify the effects of variables such as operating speeds, operating speed variance, or seasonal differences that fluctuate more often than year-to-year. Agencies require the ability to accurately assess what seasonal or daily changes could affect crash outcomes. To address this, the study developed statistical models for the segment daily level based on crash severity and roadway type. The Poisson-Tweedie statistical model is applied in the analysis due to the infrequent nature of crashes that requires zero inflation.** The project team used the Poisson distribution with rate λ to select a number n independent, and the identically distributed variables

* K= Fatal, A= Incapacitating Injury, B=Non-incapacitating Injury, C=Minor Injury, O= No Injury or Property Damage Only (PDO)

** TMCs with zero annual crashes are also included in the model

Table 1. Impact of Variable Changes on Annual Crash Frequency.

When	Crash Frequency On		
	Rural Interstate	Rural Two-Lane	Rural Multilane
Traffic volume increases	Increases	Increases	Increases
Segment length increases	Increases	Increases	Increases
Lane width increases	—	Decreases	—
Percentage of horizontal curves increases	Mostly Increases *(Decreases in OH model)*	Increases	Increases
Intersection is present	—	Increases	Mostly Increases *(Decreases in WA KABC model)*
Road is undivided	NA	NA	Increases
Percentage of days with precipitation increases	—	Decreases	Decreases
Operating speed difference between weekend and weekday increases	Increases	Increases	Increases
Average hourly operating speed variability within a day increases	—	Increases	Mostly Increases *(Decreases in WA KABCO model)*
Operating speed variability by month within a year increases	—	Increases *(OH PDO model only)*	Increases
Average hourly non-peak non-event operating speed increases (free flow) increases	Decreases		Increases
Average hourly speed increases	—	—	—

Note: at 95 percent confidence level: Increases (crash frequency goes up), Decreases (crash frequency goes down), –(not significant), NA= not applicable.

were then summed to generate a sample of the compound Poisson distribution. In the Tweedie case, these variables come from the gamma distribution with shape parameter α and scale parameter β.

Table 2 lists the variable changes and affects for the developed models. The general findings from daily level analysis are below:

- In all models, a segment with high variation in daily average speeds is expected to experience a higher number of crashes than a segment with a lower variation in daily speeds. ***The strength of this finding is one of the biggest insights gained from this study.***
- Average operating speed increases were associated with increased crashes for rural two-lane roadways. However, average operating speed increases were associated with decreased crashes in the Interstate models. This finding could be because high design standards for interstate highways.
- As the daily average precipitation increases, so do the number of daily crashes.

In addition, the project team developed another dataset (Data Structure 3 as shown in Figure 1) to examine speed variation before a crash event compared to a non-crash traffic flow condition.

Examination on a randomly selected sample dataset (with 150 crashes from Washington interstate roadways) shows that speed variability increased for the series just prior to a crash, which was also different from the comparison no crash series.

Decision Support Tool

The tool (RuralSpeedSafetyX) is hosted as a shinyapps.io webpage.[16] The tool has a dashboard with different dropdown menus to assess annual risk scoring (in terms of crashes by different injury levels) at the directional segment level. There are several flexible options to generate heatmaps at different levels (see Figure 2a for the interface and Figure 2b-e for different drop-down options). The outcomes of this tool will provide estimates of expected annual crashes (total and fatal/injury) at different geographic scales, such as state, county, and facility type. The estimates are graphically displayed in a color-coded heatmap format. This tool also allows to download the final queried data in comma separated value (CSV) format. Readers can consult Das et al. study for the framework and source codes developed for this tool.[17]

Conclusions

This study examined the prevailing operating speeds and weather data on a large scale and quantified how traffic speed and weather condition interact with roadway characteristics to affect the likelihood of crashes. The inclusion of speed information expanded upon the existing state of practice by incorporating exposure data as risk variables. This study has three unique contributions:

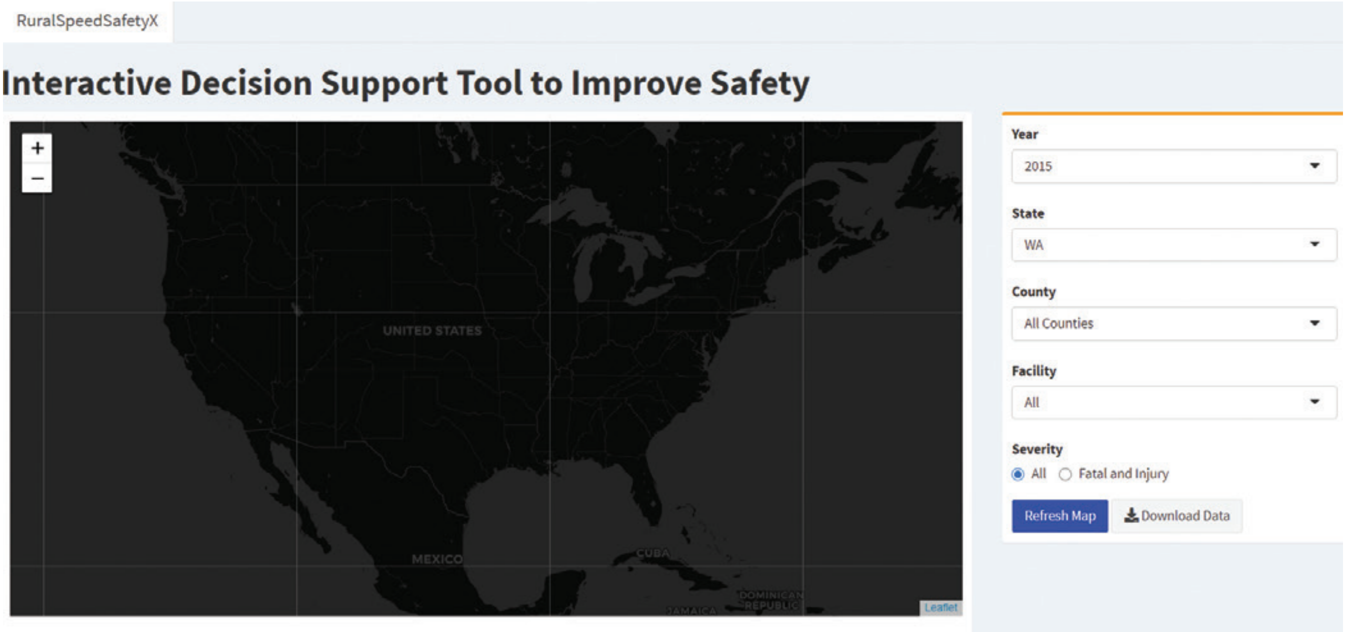
- Developed the data conflation framework using HSIS, NPMRDS, and NOAA.
- Quantified the targeted relationship between crashes and influential variables by developing best-fit models that address the impact of operating speed and weather to measure safety risk alongside traditional highway safety variables.
- Developed a scalable, flexible, and transferable decision support tool that can be reproduced by using newer datasets.

It is important to note that this study is a starting point in evaluating the effect of operating speed on crash outcomes. It is recommended that other states can use the current framework and develop similar SPFs with inclusion of speed and weather data. This study is not without limitations. First, this research used roadway segments based on TMC segment lengths, which are sometimes very long. Further examination of the effects of segment length would improve modelling reliability. Second, missing values in the NPMRDS travel time data are higher in lower functional classes. More robust NPMRDS data with fewer missing values would provide more insightful knowledge on how operating speeds

Table 2. Impact of Variable Changes on Daily Crash Frequency

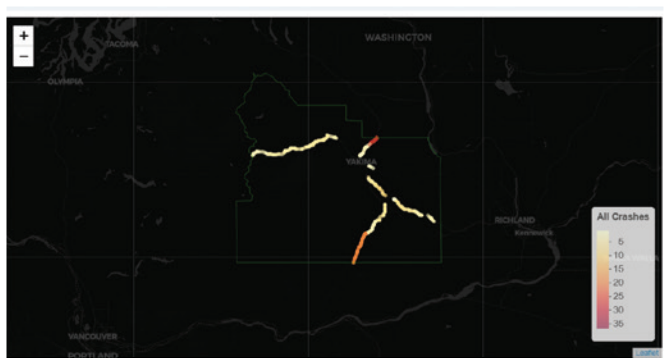
When	Crash Frequency On		
	Rural Interstate	Rural Two-Lane	Rural Multilane
Traffic volume increases	Increases	Increases	Increases
Segment length increases	Increases	Increases	Increases
Number of lanes increases	Decreases*(OH KABC model only)*	NA	—
Lane width increases	Increases *(OH KABC model only)*	Decreases *(WA model only)*	Increases *(WA model only)*
Number of curvatures increases	Decreases *(OH KABCO model only)*	Decreases *(WA KABCO model only)*	Increases *(OH model only)*
Total length of curvatures increases	—	Increases *(WA KABCO model only)*	—
Percentage of days with precipitation increases	Increases	Increases	—
Variability of daily average speed increases	Increases	Increases	Increases
Daily average speed increases	Decreases	Increases	—

Note: at 95 percent confidence level: Increases (crash frequency increases), Decreases (crash frequency increases), –(not significant), NA= not applicable.

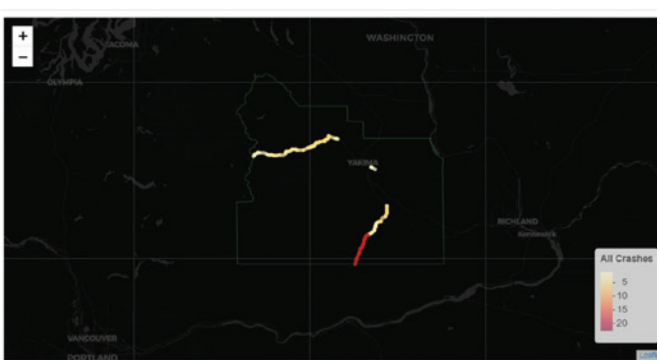


(a) Interface of the Framework of the Decision Support Tool.

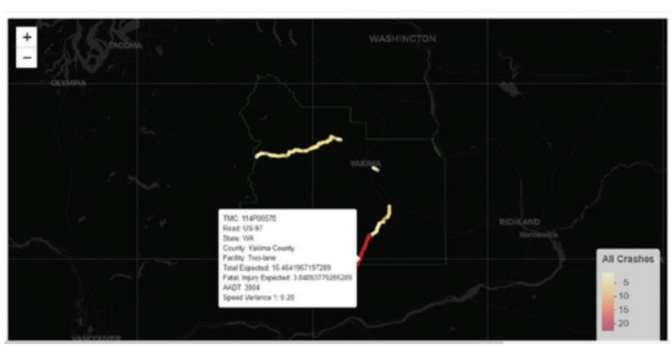
(b) Selection at State Level



(c) Selection at County Level



(d) Selection at County and Facility Type Level



(e) Hovering Option

Figure 2. Interface and Selection Options in the Interactive Tool.

affect crashes. Third, the current study used a limited number of variables. Subsequent study may examine some limitations found in this study, particularly some missing data in the current version of NPMRDS and zero inflation in short-term crash prediction, to see if those limitations can be overcome by revised versions of the data and more robust modeling techniques. **itej**

Acknowledgements

This paper has been generated from the final report of "Rural Speed Safety Project," which is one of the five pilot projects of the U.S. Department of Transportation (USDOT) Safety Data Initiative (SDI). The project was sponsored by the Office of the Secretary of Transportation (OST), USDOT (contract no. DTFH6116D00039L). The datasets used in this study are owned by the USDOT and the U.S. Department of Commerce.

Resources Available

The project team developed a weblink*** that includes descriptive statistics and data visualization tools (both static and interactive). The team members also developed an interactive decision support tool**** to show segment level risk scoring.

References

1. Abdel-Aty, M., and A. Radwan. Modeling traffic accident occurrence and involvement. Accident Analysis and Prevention 32, 2000, 633–642.
2. Taylor, M., D. Lynam, and A. Baruya. The effects of drivers' speed on the frequency of road accidents. TRL REPORT. 421, 2000.
3. Pei, X., S. Wong, and N. Sze. The roles of exposure and speed in road safety analysis. Accident Analysis and Prevention 48, 2012, 464– 471.
4. Imprialou, M., M. Quddus, D. Pitfield, and D. Lord. Re-visiting crash–speed relationships: A new perspective in crash modelling. Accident Analysis and Prevention 86, 2016, 173-185.
5. Yu, R., M. Abdel-Aty, and M. Ahmed. Bayesian random-effect models incorporating real-time weather and traffic data to investigate mountainous freeway hazardous factors. Accident Analysis and Prevention 50, 2013, 371–376.
6. Gargoum, S., and K. El-Basyouny. Exploring the association between speed and safety: A path analysis approach. Accident Analysis and Prevention 93, 2016, 32–40.
7. Yu, R., M. Quddus, X. Wang, and K. Yang. Impact of data aggregation approaches on the relationships between operating speed and traffic safety. Accident Analysis and Prevention 120, 2018, 304–31.
8. Banihashemi, M., M. Dimaiuta, A. Zineddin, B. Spear, O. Smadi, and Z. Hans. Using Linked SHRP2 RID and NPMRDS Data to Study Speed-Safety Relationships on Urban Interstates and Major Arterials. The 98th TRB Annual Meeting, 2019, Washington DC.
9. Lamm, R., E. Choueiri, T. Mailaender. Comparison of Operating Speeds on Dry and Wet Pavements of Two-Lane Rural Highways. In Transportation Research Record: Journal of the Transportation Research Board, No. 1280, 1990, 199-207.
10. Alhomaidat, F., V. Kwigizile, J. Oh, and R. Houten. How does an increased freeway speed limit influence the frequency of crashes on adjacent roads? Accident Analysis and Prevention 136, 2020.
11. Jackson, T. L., and H. O. Sharif. Rainfall Impacts on Traffic Safety: Rain-Related Fatal Crashes in Texas. Geomatics, Natural Hazards & Risk, Vol. 7, No. 2, pp. 843-860, 2016.
12. Najafi, S., G. Flintsch, and A. Medina. Linking roadway crashes and tire-pavement friction: a case study. International Journal of Pavement Engineering, Vol. 18, No. 2, 2015, 119-127.
13. Das, S., Geedipally, S., Wu, L., Fitzpatrick, K., Avelar, R., Banihashemi, M., and Lord, D. Rural Speed Safety Project for USDOT Safety Data Initiative. USDOT Final Report. 2019.
14. Rstudio, R Shiny. 2019, https://shiny.rstudio.com/.
15. Open Source Software R project, R project. 2019, https://www.r-project.org/.
16. Das, S., and L. White. Decision Support Tool for Rural Speed Safety Project. https://ruralspeedsafety.shinyapps.io/rss_sdi/ Accessed: April 2020.
17. Das, S., and L. White. RuralSpeedSafetyX: Interactive Decision Support Tool to Improve Safety. SoftwareX, 2020. https://doi.org/10.1016/j.softx.2020.100493

*** http://subasish.github.io/pages/FHWA_Rural_Speed_T4_1/

**** https://ruralspeedsafety.shinyapps.io/rss_sdi/

Subasish Das, Ph.D. (M) *is an assistant research scientist with Texas A&M Transportation Institute in College Station, TX, USA. His major areas of expertise include statistical analysis and machine learning with an emphasis in transportation safety and operations, spatial analysis with web GIS tools, interactive data visualization, and deep learning for CV/AV technologies. Subasish completed his master of science and Ph.D. in civil engineering from the University of Louisiana at Lafayette. He is an Eno Fellow and the author of Artificial Intelligence in Highway Safety, which will be published by the CRC Press in 2021.*

Srinivas Geedipally, Ph.D., P.E. (M) *is an associate research engineer in the Crash Analysis and Modeling group of the Center for Transportation Safety. He received his Ph.D. in Civil Engineering from Texas A&M University in 2008, and he holds a master of science degree in traffic environment and safety management from Linköpings University, Sweden (2005) and a bacheor's in civil engineering from Osmania University, India (2002). Dr. Geedipally has more than 15 years of experience in traffic safety research and his areas of expertise include developing statistical models for analyzing crash data, highway geometrics, Bayesian statistics, before/after studies and evaluation of treatments.*