Behavior-Based Predictive Safety Analytics Phase II

June 2023 | Final Report

SAFE-D
SAFETY THROUGH DISRUPTION

Texas A&M Transportation Institute

San Diego State University
Leadership Starts Here

Virginia Tech Transportation Institute
Virginia Tech

Disclaimer

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. 04-114	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Behavior-Based Predictive Safety Analytics Phase II		5. Report Date June 2023 6. Performing Organization Code:
7. Author(s) Andrew Miller (VTTI)* Abhijit Sarkar (VTTI) Tony McDonald (TTI/TAMU) Sahar Ghanipoor-Machiani (SDSU) Arash Jahangiri (SDSU)		8. Performing Organization Report No. 04-114
9. Performing Organization Name and Address: Safe-D National UTC Virginia Polytechnic Institute and State University Virginia Tech Transportation Institute 3500 Transportation Research Plaza, Blacksburg, VA 24061, USA		10. Work Unit No. 11. Contract or Grant No. 69A3551747115/04-114
12. Sponsoring Agency Name and Address Office of the Secretary of Transportation (OST) U.S. Department of Transportation (US DOT)		13. Type of Report and Period Final Research Report Start Date: 3/2019 End Date: 6/2023 14. Sponsoring Agency Code
15. Supplementary Notes This project was funded by the Safety through Disruption (Safe-D) National University Transportation Center, a grant from the U.S. Department of Transportation – Office of the Assistant Secretary for Research and Technology, University Transportation Centers Program.		
16. Abstract This project addressed the emerging field of behavior-based predictive safety analytics, focusing on the prediction of road crash involvement based on individual driver behavior characteristics. This has a range of applications in the areas of fleet safety management and insurance, but may also be used to evaluate the potential safety benefits of an automated driving system. This project continued work from a pilot study that created a proof-of-concept demonstration on how crash involvement may be predicted on the basis of individual driver behavior utilizing naturalistic data from the Second Strategic Highway Research Program. The current project largely focuses on understanding and identifying the risks from a driver based on their driving behaviors, personal characteristics, and environmental influences. This project analyzed large scale continuous naturalistic data as well as event data to study the role of different driving behaviors in the buildup of risk related to a safety-critical event or crash. This research can be used structure the development of real-time crash risk that accounts for those identified driver behaviors to be evaluated across the contextualized information on a roadway.		
17. Key Words Crash risk, individual differences, predictive safety, safety metrics, crashes, near-crashes		18. Distribution Statement No restrictions. This document is available to the public through the Safe-D National UTC website, as well as the following repositories: VTechWorks, The National Transportation Library, The Transportation Library, Volpe National Transportation Systems Center, Federal Highway Administration Research Library, and the National Technical Reports Library.
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 17 — 22. Price $0

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Abstract

This project addressed the emerging field of behavior-based predictive safety analytics, focusing on the prediction of road crash involvement based on individual driver behavior characteristics. This has a range of applications in the areas of fleet safety management and insurance, but may also be used to evaluate the potential safety benefits of an automated driving system. This project continued work from a pilot study that created a proof-of-concept demonstration on how crash involvement may be predicted on the basis of individual driver behavior utilizing naturalistic data from the Second Strategic Highway Research Program. The current project largely focuses on understanding and identifying the risks from a driver based on their driving behaviors, personal characteristics, and environmental influences. This project analyzed large scale continuous naturalistic data as well as event data to study the role of different driving behaviors in the buildup of risk related to a safety-critical event or crash. This research can be used structure the development of real-time crash risk that accounts for those identified driver behaviors to be evaluated across the contextualized information on a roadway.

Acknowledgements

This project was funded by the Safety through Disruption (Safe-D) National University Transportation Center, a grant from the U.S. Department of Transportation – Office of the Assistant Secretary for Research and Technology, University Transportation Centers Program.

Table of Contents

List of Figures

List of Tables

Introduction

In recent decades, research on driving-related outcomes has expanded rapidly, with heavy emphasis on understanding driving- and non-driving-related behaviors that lead to the least desirable driving outcome—serious crashes (Hakkert & Gitelman, 2014). Crashes are the cause of approximately 1.35 million annual deaths worldwide and anywhere from 20 to as many as 50 million people are injured in road-related crashes each year (World Health Organization, 2018). Although only 3% of those fatalities occur in the U.S., in 2010, the economic and societal impact of crashes was estimated to cost $836 billion (National Highway Traffic Safety Administration [NHTSA], 2015). Although there has been a steady decrease in the rate of road-related fatalities from the inception of recorded crash statistics, according to the Fatality Analysis Reporting System, the rate has been hovering around 1.1 deaths per 100 million vehicle miles traveled since 2009 and has even increased in the past few years (NHTSA, 2023). Additionally, estimates suggest that as many as 94% of crashes are the result of driver-related errors—specifically, recognition, decision, performance and non-performance errors—while the remaining 6% occur due to vehicle, environmental, or unknown reasons (NHTSA, 2015). This has led researchers, regulators, and practitioners to heavily focus on crash-related aspects of individual drivers.

Characterizing individuals' driving performance has been a large focus of transportation safety research, dating back to the 1930s when Ryan and Warner (1936) measured drivers' mental capacity and physiological reactions during a period of fatigued driving, followed by McClintock (1936), who measured driver judgments on speed and distance in an attempt to correlate it with individual differences. Succeeding these efforts, researchers were interested in identifying underlying characteristics that led to an individual's predisposition to being in an accident; specifically, accident proneness was adopted from industrial research and applied to driving outcomes (see: Shinar, 2017).

Understanding an individual driver's contribution to a crash has resulted in an extensive review of differential crash involvement by a number of authors (e.g., Cullen et al., 2021; Guo & Fang, 2013). Further, a small proportion of drivers often account for a major proportion of crashes (Sagberg et al., 2015), a phenomenon often referred to as the Pareto principle, or the 80–20 rule. Due to their high involvement in crashes, it is of great value to be able to identify these risky drivers before crashes occur.

Background

The basic idea underlying differential crash involvement is that some drivers have certain characteristics that make them more likely to become involved in crashes. For example, drivers with a stronger propensity for risk-taking behaviors may tend to look away from the road for longer or more frequent periods than the average driver, increasing the risk that a distraction-related driver behavior will coincide with an unexpected external event, thus leading to a crash. In another scenario, a vigilant driver may habitually double- or triple- check opposing or oncoming traffic at a traffic signal before entering the intersection to reduce the likelihood of an incident with other vehicles.

These characteristics often reference an individual's demographics and disposition (Knipling, 2009). Research across the transportation safety domain frequently focuses on specific driver characteristics, such as gender (Cullen et al., 2021), age (Hu et al., 2020; Bharadwaj et al., 2021), personality (de Winter & Dodou, 2010; Braitman & Braitman, 2017), driving experience (Horswill et al., 2020), aggression (Velazquez, 2020; Demir et al., 2016), medical conditions and health (Filtness et al., 2020), and self-regulation (Wong et al., 2015; Delvin & McGillivray, 2016).

Further, personal factors may also include temporal facets. These typically relate to individualized changes that occur day-to-day or hour-to-hour, like mood, impairment, comfort, and fatigue (Knipling, 2009). Additional personal factors may manifest themselves in terms of observable driver behavior patterns (e.g., speeding, close following, hard braking, engagement in distraction) as well as in behavioral history (e.g., the number of crashes or violations in the past 3 years, criminal records).

While many research efforts have focused primarily on drivers' dispositions as related to crash proneness, many other studies evaluated the effects of situational factors (Knipling, 2009). Situational factors play an important role in shaping what behaviors are possible (e.g., driving in dense traffic), driving constraints (e.g., speed limits), or shifts in usual driving conditions (e.g., bad weather), among others.

Among these efforts, identification of risky drivers has been performed by examining individual driver characteristics, situational elements, and combinations of interacting effects. There is strong evidence that enduring personal factors influence crash involvement beyond mere chance (Simons-Morton et al., 2012). However, it is also clear that these personal factors often interact with situational factors and temporary personal factors in non-trivial ways in producing crashes. For example, while the occurrence of driver fatigue can be regarded a temporary factor, there is strong evidence that the susceptibility to fatigue is an enduring factor (Knipling et al., 2004). A similar argument may be made for the role of alcohol in crash causation, as the effect of alcohol on behavior may depend strongly on enduring personality-related factors (see review in Elander et al., 1993). Thus, crashes often occur through an interaction between dispositional and temporary personal factors and situational factors. However, the relative contributions of these different types of factors towards the crash genesis are often difficult to disentangle (Elander et al., 1993).

These temporal and enduring personal factors, along with reinforcement of behavioral selection, subsist in ones' driving style (Sagberg et al., 2015). Sagberg et al. (2015) refer to driving style as "a habitual way of driving, which is characteristic for a driver or a group of drivers." Further, driving style tends to occur in a consistent way, which may include both automatized and consciously chosen behaviors. A driver may have a repertoire of driving styles applied under different conditions, and driving styles that exclude behavior patterns determined exclusively by the driving context.

Current behaviors are influenced by situational factors as well as temporary and enduring personal factors. This results in behavioral outcomes that may be successful (i.e., the situation played out as expected) or unsuccessful, leading to, for example, crashes, traffic conflicts (e.g., near crashes), violations, and other safety-related events. Enduring personal factors are reflected in recurring observable behavioral patterns as well as behavioral history. Some of these recurring behaviors (e.g., tailgating, speeding, distraction) may be associated with increased crash risk.

Ultimately, relating these individual and situational factors to crash risk and crash involvement has been a long-standing goal in road safety research (e.g., Elander et al., 1993; Guo et al., 2010; McKenna, 1983). These factors have been related to a range of personal factors, such as gender, age, personality, and health, and may be manifested in recurrent patterns of observable driving behavior, such as speeding, close following, and secondary task engagement. Further, records of behavioral history, such as past violations, convictions, and crashes, have also been found to predict future crash involvement (Simons-Morton et al., 2012). Fatigue, as a temporal factor, is present in roughly 2% of fatal crashes within the U.S. (Stewart, 2022), and is noted as having similar inhibiting effects as driving while inebriated (Zhang et al., 2014). Figure 1 provides an illustration of the crash factors conceptualization.

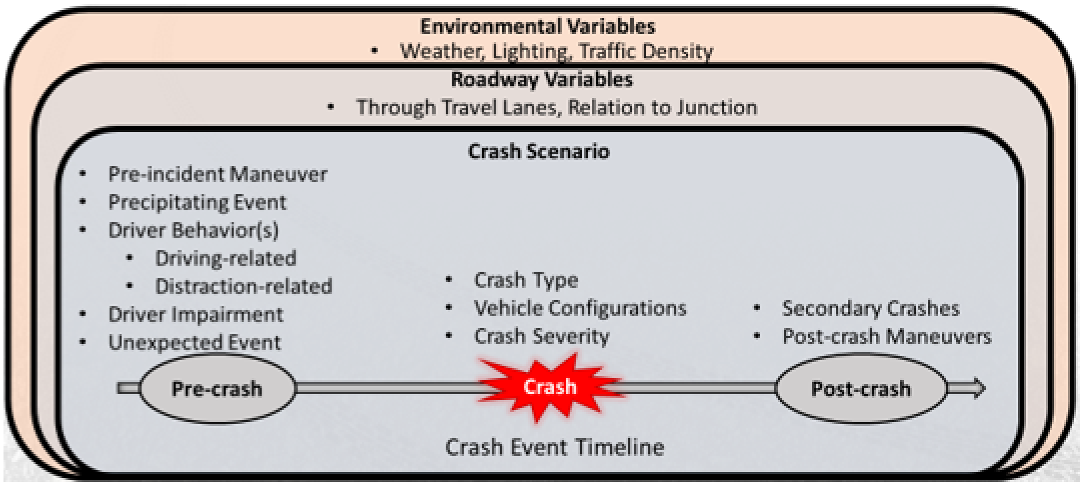

Figure 1. Conceptualization of person and situational factors involved in a crash.

The success of identifying risky drivers ultimately hinges on the establishment of models able to reliably relate individual driver characteristics to actual crash risk. This relationship is currently poorly understood (Sagberg et al., 2015; Engstrom et al., 2018). Traditionally, the main reason for this has been the lack of data containing enough detailed crash recordings and recorded driving behavior, demographics, and screening data collected over an extensive time period before the crash. In recent years, this picture has started to change due to the adoption of naturalistic driving studies. However, existing naturalistic driving analyses have typically focused on the relationship between drivers' engaging in potentially distracting behaviors that result in inattention to the driving task and crash risk, with the primary goal to identify risky tasks or behaviors (e.g., Dingus et al., 2016; Fitch et al., 2013).

Research Objectives

The goal of the current study was to provide insight into the calculation of behavioral indicators and associated metrics that would help clarify the relationship between driver-centric variables (i.e., driver behaviors, driving style, personality or risk-taking assessments, demographics, etc.) and crash risk. Further evaluation includes determining the effect that other situational components may have on driver-centric variables (within-person analyses) or crash risk (between-person analyses). The research team evaluated the Second Strategic Highway Research Project (SHRP 2) and large truck naturalistic databases to examine these interactions with crash risk (Dingus et al., 2015; Krum et al., 2016, Hammond et al., 2021).

Method

Data Collections

Two types of naturalistic data collections were used to produce results and analyses in exploring the relationship between situational, personal, and behavioral factors with crash risk. Each dataset serves a different population and subsequently has its own strengths and weaknesses that were explored throughout the project. For example, although the SHRP 2 dataset had a large number of drivers, it also had a large number of situational and temporary factors, reflecting the high-fidelity nature of the data collection, but reducing the ability to detect certain variance in driving style and nearly eliminating any claims of causation.

SHRP 2 Naturalistic Data Collection

The SHRP 2 collection consisted of 3,092 primary drivers and 454 secondary drivers across six collection sites within the contiguous U.S., for a collective 3,546 drivers (52.2% female). Primary drivers consisted of those who consented to the study and completed additional questionnaires, while secondary drivers had data collected inadvertently by driving in a primary driver's instrumented vehicle. Young and senior drivers were oversampled in order to have ample data to examine any age-related effects involving extreme subgroups.

SHRP 2 Subjective Measures

A series of subjective measures were collected before a data acquisition system (DAS) was installed in a participant's personal vehicle, heretofore referred to as ego vehicle. These measures comprehensively covered driving-relevant factors that may in some way affect driving style, though driving style was not directly measured.

Risk-taking Behavior

A risk-taking behavior questionnaire was created as a combination of the Cox Assessment of Risk Driving Scale (CARDS) and DeJoy Risk Perception Questionnaire (Transportation Research Board, 2014). The questions assessed self-report frequency of driving behaviors in the prior 12 months. Example items included the frequency of running red lights, making illegal turns, and getting very angry at other drivers.

Risk-Perception

A risk-perception questionnaire was created for the SHRP 2 data collection (Dingus et al., 2015). The questions assessed the perceptual risk with driving behaviors on a seven-point Likert scale ranging from "No Greater Risk" to "Much Greater Risk." An example item from the scale was "If you were to engage in changing lanes suddenly to get ahead in traffic, how do you think that would affect your risk of a crash?"

Thrill/Adventure Seeking

The Sensation Seeking Scale-form V (SSS-V; Zuckerman, 1994) presented a series of scenarios in which respondents chose their preference in a pair of options. Examples of scenarios involved mountain climbing, learning to fly an airplane, and social drinking. A composite score was created from the scale to represent sensation seeking.

Driver History

A driving history questionnaire was created for the SHRP 2 data collection effort. This included estimating annual mileage, years of driving, and number of police-reportable crashes and moving violations that occurred in the past 3 years.

Driving Knowledge

A custom 20-item questionnaire to assess driving knowledge was created for the data collection effort. This consisted of several state licensing practice tests on non-state specific driving regulations. A composite score ranging from 0 to 20 was created from correct answers.

Driver Behavior

A modified version of the Manchester Driver Behavior Questionnaire (Dingus et al., 2015) was used to assess driver style and driver behaviors. This questionnaire had 24 items that assessed behavioral frequency of improper driving-related behaviors. Example items included the frequency of speeding, missing environmental cues, or performing aggressive behaviors.

Large Truck Naturalistic Driving Data Collections

VTTI conducted several naturalistic driving studies with commercial motor vehicles (CMVs) that provide a wealth of vehicle data and driver behaviors. CMV collections at VTTI often have fewer participants than light vehicle studies, but the CMV operator will drive many more miles than their light vehicle counterpart. The 177 participants used from the large truck collections (FAST DASH 2 and OBMS 2 [Krum et al., 2016; Hammond et al., 2021]) averaged approximately 40,000 miles traveled during the study duration.

Due to the large number of miles traveled per CMV, there is ample data to examine within-person effects, largely, the variance in driving styles or deviations from normal behavior. However, there were a number of other considerations for evaluating CMVs compared to light vehicles, including the following.

- Naturalistic data collections contain some drivers from the same fleet, so fleet-level variables may influence driver behaviors (e.g., safety culture, training, dispatcher flexibility).
- CMV operators are trained professionals and may therefore exhibit different driving styles than light vehicle operators who do not drive professionally.
- Though dependent on vehicle operations, Class 8 tractor-trailer CMVs primarily travel on highway or interstate roadways. This reduces the variance in urban or last-mile contexts, but provides extensive driving metrics across high-speed travel.
- CMV operators must adhere to hours-of-service regulations, which limit the extent of time on-the-job and on-road. Fatigue may be the largest influence on a truck driver's driving style outside of the state-based characteristics (i.e., mood, anxiety).
- The CMV operator workforce is male-dominated. These drivers share other similar characteristics (i.e., demographics, personality) that may fail to be representative of the larger population of light vehicle drivers.
- The data collection efforts do not capture as many person-based assessments as SHRP 2 and therefore most investigations involve in situ driver behaviors.

- Fewer crashes occurred during the course of data collection, in magnitude and on a per mile basis. Recorded crashes that did occur were primarily low-speed incidents that neither involved other vehicles nor caused property damage.

Data Cleaning

The naturalistic datasets required preparation to ensure use of the appropriate data. We used a time series data dictionary and took the following steps to ensure appropriate data was used for the analyses:

- Trips were analyzed by driver instead of vehicle to mitigate multiple drivers within one vehicle.
- The first 60 secs of each drive were deleted to obtain consistent speed and GPS data. This incorporated the boot speed of the DAS. However, reduced events that occurred within 1 minute of the start of the DAS were not removed.
- To reduce the file size and the number of missing values, we only used radar tracking 1 and 2 (radar variables). We also removed the binary radar variables.
- Rolling windows were applied on data with a window size of 30 secs and a sliding window of 15 secs.
- Missing values from GPS speed were replaced using the Last Observation Carried Forward imputation method. It was observed that the GPS speed was only recorded when it changed, so any range of NA values were replaced by the value that preceded the NA value.
- All other missing values in other variables were replaced using Kalman Smoothing. In this method, an initial estimate was developed for the state parameters in a forward pass, similarly to the traditional Kalman Filter. Following this step, a backwards pass estimate was conducted, and the error was minimized through Expectation Maximization. It was assumed these values were missing due to the equipment going offline. So, the same replacement of NA values was applied for the steering position variable and pedal gas variable. This was done so the missing values followed the trend of the closest previous observation. This was done using the imputeTS package in R.
- After the rolling window was performed, all rows containing NAs were removed and the data was exported into a CSV format.

Behavioral Indicators and Modeling

Using the SHRP 2 data, two distinct analysis levels were created and analyzed. First, a person-level dataset was established to describe the participants. This dataset included scoring the selected personality and behavioral self-report measures and aggregating trip-level data to the person-level. The second analysis level consisted of trips, subsampled to match the event data. Roughly 41,400 trips were analyzed using behavioral indicators, then further matched to event data, which consisted of 1,836 crashes, 6,881 near-crashes, and 32,581 baseline sampled events. These crash and near-crash (CNC) events provided full context of situational factors and driver behaviors, while the baseline events provided the full context of a nominal, or normal driving extraction. Among the truck database, all trips were analyzed to produce similar datasets.

Results

Calculated Indices

The research team identified and calculated a number of indices and indicators to describe an individual's collection of trips.

Crash, Near-crash

In order to estimate crash-risk, individuals were divided into groupings. These groupings reflect whether the following was experienced during the data collection period:

- Crash (vs. no crash)
- Crash and/or near-crash (vs. neither crash nor near-crash)
- At-fault crash or at-fault near-crash event (vs. not at fault or no event)

Acceleration-based Indicators

Acceleration events were calculated for each trip and were aggregated within trip and within person. These events were based on selected thresholds and represent safety-critical events that are not typically present in normal operations. The acceleration events were calculated per trip, and their associated bins were as follows.

- Longitudinal accelerations (i.e., fast acceleration)
  - Three bins:
    - 0.1 g to 0.2 g
    - 0.2 g to 0.35 g
    - 0.35 g and above
- Longitudinal decelerations (i.e., hard braking)
  - Three bins:
    - -0.1 g to -0.2 g
    - -0.2 g to -0.35 g
    - -0.35 g and below
- Lateral accelerations (i.e., strong turns)
  - Four bins:
    - ±0.1 g to ±0.20 g
    - ±0.2 g to ±0.30 g
    - ±0.3 g to ±0.35 g
    - ±0.35 g and above/below

Lateral accelerations incorporate forces on the inertial measurement unit from both directions on the y-axis, represented as positive and negative g-forces. Positive forces represent lateral movement to the right-side of the vehicle (i.e., above +0.35 represents a strong positive force against the right of the vehicle as it is turning left or negotiating curves to the left).

For CMV naturalistic datasets, acceleration metrics were calculated based on map-matched data. The metrics were binned based on varying thresholds and were calculated as they happened on specific sections of roadways. This provided the ability to evaluate specific roadway characteristics according to the mapped dataset. Figure 2 denotes an illustrative representation of the longitudinal deceleration bins calculated and their respective thresholds.

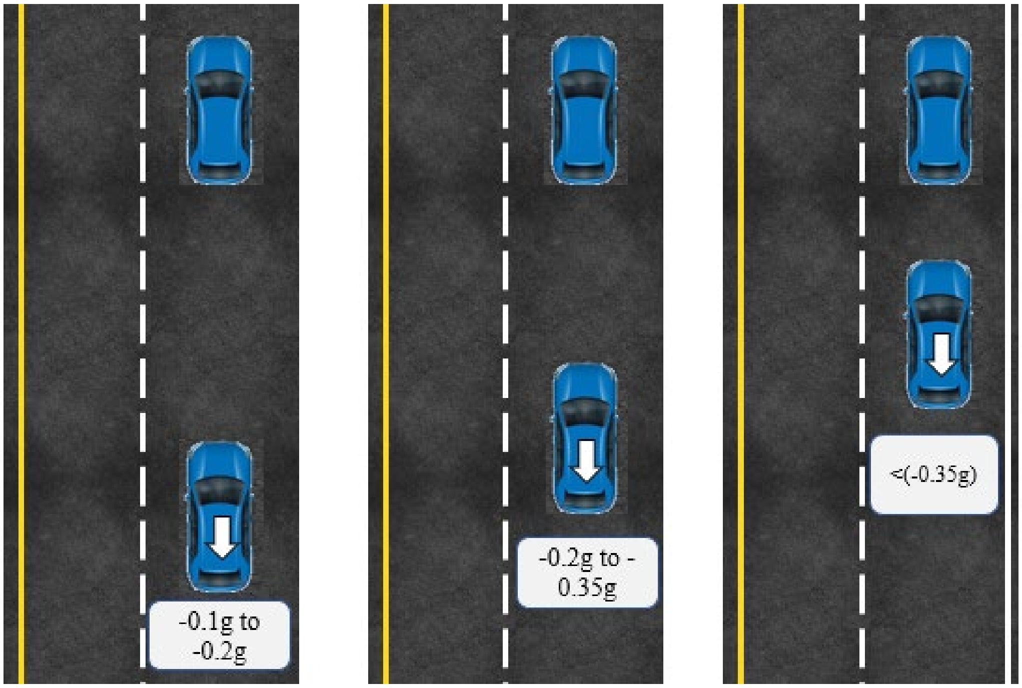

Figure 2. Schematic defining the concept of longitudinal deceleration.

Similarly, Figure 3 illustrates the lateral acceleration bins and thresholds. Negative accelerations dictate a left acceleration (e.g., turning right) and positive accelerations dictate a right acceleration.

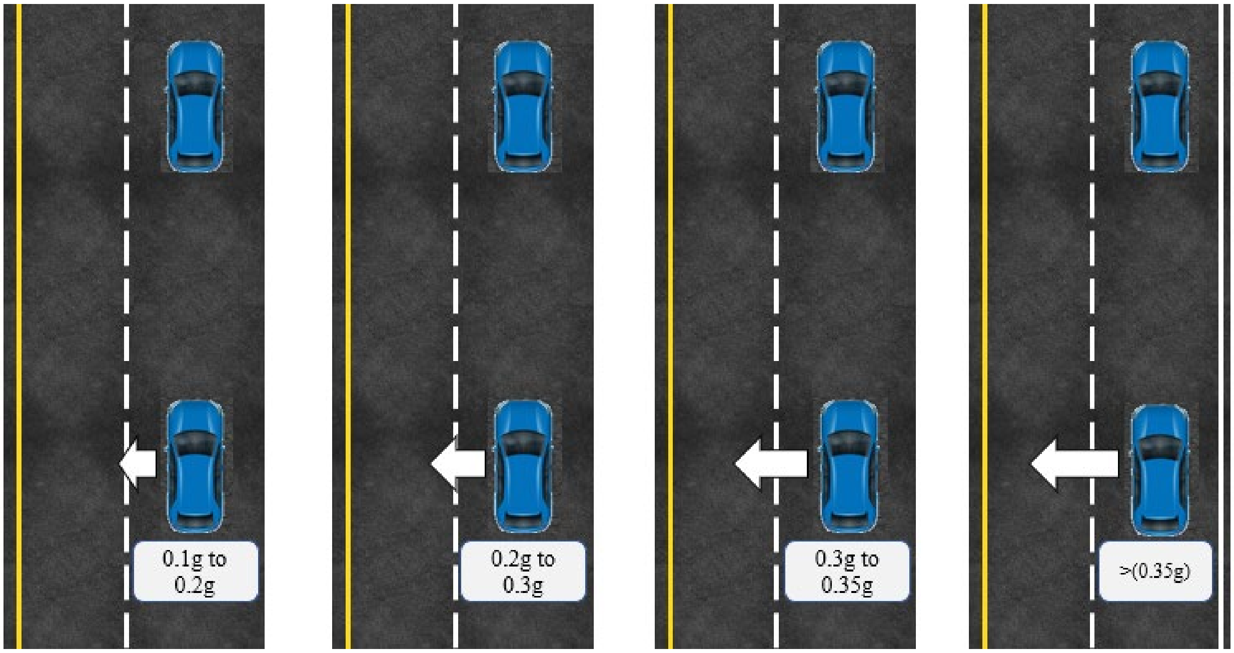

Figure 3. Schematic defining the concept of lateral acceleration.

Headway

Forward headway was calculated as the distance traveled at a categorized time headway when a lead vehicle was present. The categories consisted of eight (8) 0.5 second bins starting at 0.0

seconds and ending at time headway greater than 3.5 seconds. The distance traveled was then normalized by the time spent with a lead vehicle present.

Minimum Headway at Speeds

A minimum time headway was calculated for each trip in which there was a lead vehicle present among categorized speeds of the ego vehicle. The categories of speed consisted of six (6) 10-mph bins starting at 0 mph and ending at ≥ 50 mph. The metric described the lowest elapsed travel time for the following vehicle to reach the space occupied by the leading vehicle.

Minimum Time-to-collision at Speeds

A minimum time-to-collision was calculated among categorized speeds of the ego vehicle for each trip in which there was a lead vehicle present. The categories of speed consisted of six (6) 10-mph bins starting at 0 mph and ending at ≥ 50 mph. This measure was similar to minimum headway, though time-to-collision was calculated as the elapsed time before a collision between two vehicles if both vehicles maintained their current velocities.

Speed Behaviors

Speed behaviors were analyzed for CMVs as the distance traveled at certain speeds relative to the posted speed limit. Specifically, the metric comprised the difference between average speed and the posted speed limit, which was then calculated in miles driven over road segments. An illustration of speed behavior is presented in Figure 4.

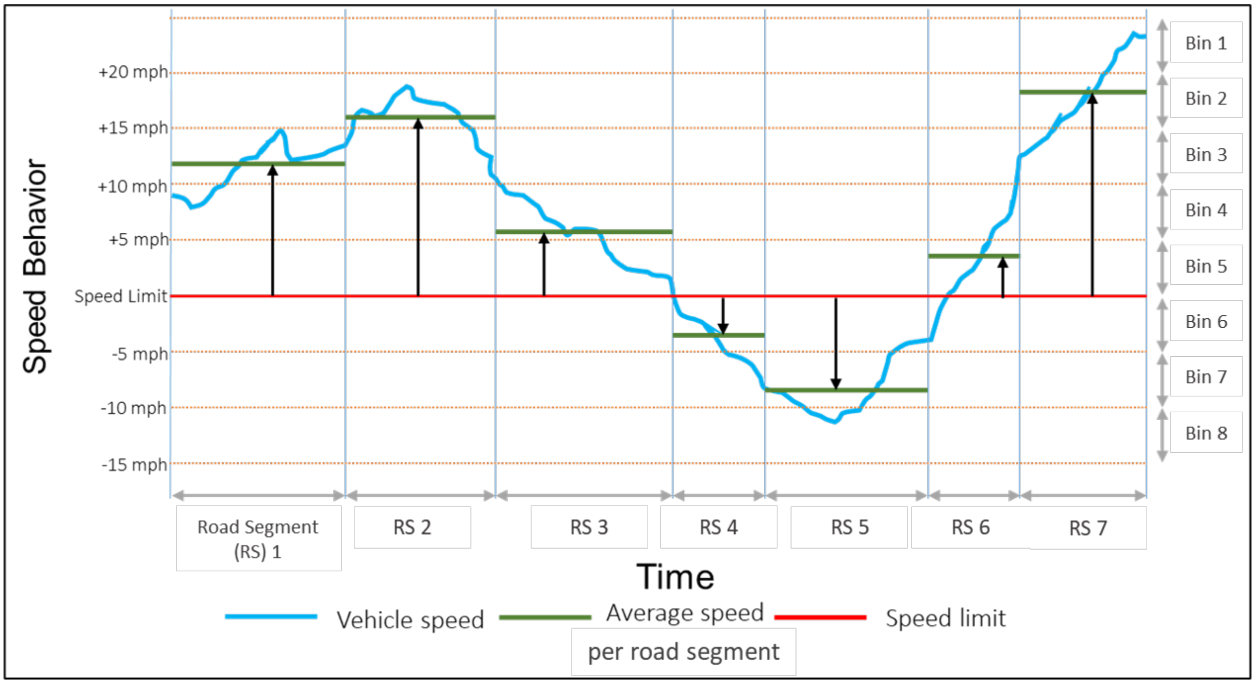

Figure 4. Schematic defining the concept of speed behavior.

Lane Deviations

For CMVs, lane deviations were measured by the lateral distance between the centerline of the lane and the middle point of the ego vehicle, as calculated by computer vision detection of lane lines and extrinsic positioning of the forward-facing camera. Lane deviations (illustrated in Figure 5) were calculated as the miles traveled in each bin, with the first bin representing no lane deviation, and subsequent bins representing up to an additional 12 inches outside of the lane. Specifically:

- Bin 1: Deviation of 0–21 inches, or tires on both sides are inside the lane.
- Bin 2: Deviation of 21–33 inches, or one tire is 0–12 inches outside the lane.
- Bin 3: Deviation of 33–45 inches, or one tire is 13–24 inches outside the lane.
- Bin 4: Deviation of > 45 inches, or one tire is > 24 inches outside the lane.

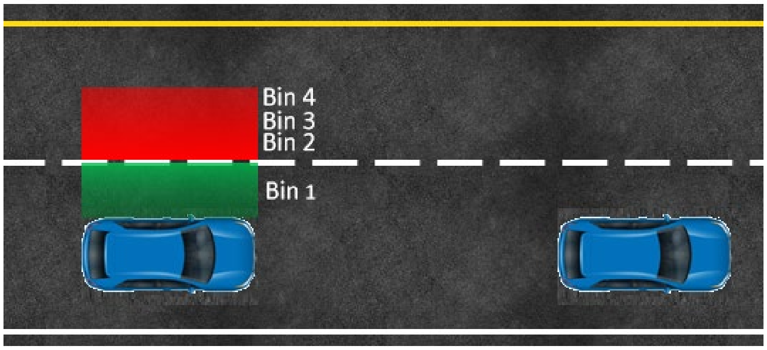

Figure 5. Schematic defining the concept of lane deviation.

Exposure Metrics

In order to normalize driving behavior, a series of exposure-based metrics were created to enhance other calculations. Exposure, representing the data of a complete trip, was critical for making comparisons between and within individuals.

Speed-based Time and Distance

Time and distance bins were calculated at various vehicle speeds to determine exposure rates of events or other relevant calculations relative to speed. These bins are represented by increments of 10 mph starting at 0–10 mph and ending at above 80 mph. Nine (9) speed bins were calculated for time and distance. As an example of a distance calculation, a trip may consist of 14 miles traveled between 21–30 mph. These metrics were calculated and reported per trip.

Posted Speed Limits

We calculated the distance traveled while under various posted speed limits. Speed limits were provided at limits of 35 mph, 50 mph, 65 mph, and over 65 mph, along with information on how much of the posted speed limit data was not available.

Roadway Type

Various roadway types, generally classifying roads as urban or rural, had calculations produced for CMV (only local and highway driving) and SHRP 2 (full highway safety information system breakdown) datasets. The road type was calculated as a percentage of time spent across each category.

Following Distance Metrics

To normalize headway and time-to-collision metrics, radar-based exposure metrics were produced. We calculated the number of targets that the radar classified, as well as the time spent with a lead vehicle present.

Between-subjects Behaviors

The first step in using the behavioral indicators was to produce a correlation matrix of CNCs with select behaviors (Table 2). The behaviors included acceleration-based metrics and forward headway to the lead vehicle. Many of the relationships were significant and were explored further.

Further investigation into these between-subjects metrics revealed that individuals who were involved in a CNC were more likely to perform closer following and strong accelerations (Table 1).

Table 1. Percent of Time Drivers Operated across Headways Based on CNC-involvement

	HW 0–0.5 s	HW 0.5–1.0 s	HW 1.0–1.5 s	HW 1.5–2.0 s	HW 2.0–2.5 s	HW 2.5–3.0 s	HW 3.0–3.5 s	HW > 3.5 s
No CNCs (n = 1,143)	0.89%	6.55%	13.32%	14.22%	12.25%	9.82%	7.8%	35.14%
CNCs (n = 2,031)	1.4%	9.96%	15.9%	15.18%	12.11%	9.17%	7.15%	29.13%
Percent Change	+57%	+52%	+19%	+7%	-1%	-7%	-8%	-17%

Within-subjects Behaviors

Defining a real-time crash risk involved determining the behavioral attributes of drivers across various circumstances and under a variety of conditions. Current investigations defined high-threshold longitudinal acceleration and deceleration events, along with high threshold lateral acceleration events (positive and negative), and the average headway when a lead vehicle was present, and the ego vehicle was below a 3.5-second headway. This selection of headway removed times larger than 3.5 seconds, as this removed confounding variables related to the binned times of headway that did not account for times greater than 3.5 seconds. The radar-sensing technology could detect much larger gaps, which were factored into the average. Indicators were created across the entirety of a trip and normalized as appropriate.

Table 2. Correlation Matrix of Crashes and Near-crashes with Acceleration and Headway Metrics

Variable	CNC	Long. Accel	Long. Decel	Lat. Accel	Headway 0–0.5 s	Headway 0.5–1.0 s	Headway 1.5–2.0 s	Headway 2.0–2.5 s	Headway 2.5–3.0 s	Headway 3.0–3.5 s	Headway > 3.5 s	
Crash & Near Crash	—											
Long. Accel	0.407***	—										
Long. Decel	0.787***	0.56***	—									
Lat. Accel	0.508***	0.419***	0.598***	—								
Headway 0–0.5 s	0.205***	0.107***	0.199***	0.186***	—							
Headway 0.5–1.0 s	0.376***	0.178***	0.378***	0.303***	0.435***	—						
Headway 1.0–1.5 s	0.336***	0.151***	0.325***	0.272***	0.144***	0.631***	—					
Headway 1.5–2.0 s	0.274***	0.103***	0.23***	0.165***	0.014	0.258***	0.654***	—				
Headway 2.0–2.5 s	0.219***	0.065***	0.148***	0.093***	-0.058***	0.041*	0.301***	0.651***	—			
Headway 2.5–3.0 s	0.165***	0.03	0.082***	0.02	-0.091***	-0.078***	0.090***	0.379***	0.67***	—		
Headway 3.0–3.5 s	0.123***	0.01	0.044*	-0.011	-0.103***	-0.125***	-0.01	0.229***	0.44***	0.659***	—	
Headway > 3.5 s	0.07***	-0.017	-0.021	-0.063***	-0.129***	-0.281***	-0.281***	-0.092***	0.124***	0.29***	0.44***	—

* p < .05, ** p < .01, *** p < .001

Table 3. Correlation Matrix of Crashes and Near-crashes with Acceleration and Headway Metrics

Variable	Average Trip Headway	Average Trip Long. Accel	Average Trip Long. Decel	Average Trip Lat. Accel
Average Trip Headway	—			
Average Trip Long. Accel	-0.074	—		
Average Trip Long. Decel	-0.332***	0.317***	—	
Average Trip Lat. Accel	-0.318***	0.279***	0.349***	—

Table 3 provides a series of paired-sample t-tests evaluating the trips of 749 drivers who experienced CNC events, as well as those with recorded baseline events. The full trips for these drivers were analyzed for acceleration and headway indicators and split between those trips that contained CNCs and those that did not.

Table 3. Behavioral Indicator across CNC-involved or Baseline Trips

	Longitudinal Acceleration	Longitudinal Deceleration	Lateral Acceleration	Average Headway
Non-CNC trip	0.039	0.163	0.247	2.226
CNC trip	0.063	0.566	0.299	2.154
Difference	-0.024	0.403	-0.052	0.072
Effect size	0.038	0.052	0.034	0.034
t-score	-3.532	-26.545	-2.969	5.457
p-value	< 0.001	< 0.001	0.003	< 0.001

Results indicated that drivers experienced more longitudinal accelerations/decelerations, more lateral accelerations, and shorter headways when their trip involved a CNC event compared to no CNC.

Discussion

The use of real time crash prediction models has largely focused on indicators related to infrastructure (e.g., Yang et al., 2018), traffic (e.g., traffic flow; Xu et al., 2015) or environmental conditions (e.g., weather; Abdel-Aty et al., 2012). The foundation of these models includes the segmentation and categorization of roadways based on common infrastructure elements, and then further modeling of these segments under various strains of traffic patterns and/or environmental conditions. A common output of these models is the identification of higher crash likelihoods among certain road attributes when crossed with traffic or environmental conditions, indicating the interaction-level effects of crashes (e.g., Sun et al., 2020).

Despite the collective efforts of researchers into predicting real-time crash risk, very few models have incorporated any driving behaviors, and, if these behaviors were represented, it was usually a cross-sectional documentation of a single behavior. The primary example of this inclusion would be vehicle speed, commonly identified through the same detection methods as those used to collect traffic flow (as presented in Dutta & Fontaine, 2020).

The marrying of infrastructure- and traffic-based data with the individual-level variables of driving styles, personality-based metrics, and other behavioral indicators is so uncommon due to the natural mismatch in data collection methods and presentation of models. Typically, these individual difference-centric variables are gathered within laboratory settings or via naturalistic data, where a large set of behaviors can be operationalized and collected. However, the nature of collecting these data leads to large sets of behaviors captured across a smaller set of individuals, and often there is little meaningfully captured data that relates to infrastructure or traffic. Similarly, the epidemiologic collection of crash data and related roadway attribute and environmental data lacks the capabilities to connect such individual-centric information as is necessary to begin collating the two streams of micro (i.e., individual) and macro (i.e., crash-oriented) data. Despite this mismatch, recent improvements in naturalistic driving data collections have allowed researchers to begin making these connections, though a number of concessions exist for generalizations and modeling crash risk.

Each step undertaken by researchers and the industry reveals additional insight into driving behaviors and their relation to crash risk. The ability to classify drivers according to their driving style, as defined by actual behaviors on road, leads to increasingly accurate identification of risky drivers and risky behaviors. Exploration of new behavioral indicators and the interactions of these indicators across varying conditions provides additional insight into the relationship between nominal driving and crash likelihood, while also providing limited insights into crash avoidance techniques or skills as they relate to individual-level variables.

The results of this study investigated individuals' driving styles, traditionally operationalized by some combination of acceleration and headway behaviors, and tailor these driving styles for comparison to outcome-oriented behaviors related to crashes, near-crashes, or some other driving-related event (e.g., following traffic laws). The creation and use of specific behavioral indicators was meant to further explicate the relationship to crash risk by comparing to CNC events.

Results also indicated that various driving styles may be adopted by the same individual. The relationship between driving style and situational context will provide additional insight into how various outcomes occur, and the relationship of those interaction-effects on crash likelihood. At-risk in situ behaviors may be performed by an individual; these will then impact the representation of an individual's driving style through the existing operationalization of driving styles. It is currently unknown under what conditions these less safe behaviors occur.

Several limitations apply across the dataset, as the typical difficulties of naturalistic data include describing crash risk and generalizing to a larger population of individuals. The smaller number of crashes and use of near-crashes that act as surrogate to crashes limit the capacity to properly identify individual-centric contributions specifically related to crashes. Further, the varying levels of severity in crashes is usually not represented in naturalistic data compared to epidemiologic datasets. In addition, defining at-fault, shared-fault, and partially at-fault contributions towards crash events is difficult without fully contextualizing the situation by making assumptions during data reduction.

Specific to the analyses, when performing within-subjects designs, capturing data not related to the CNC event would reduce the multicollinearity of comparing driving styles to CNC metrics, where the CNC serves as predictor and outcome. This effort may reduce some significance of results if the CNC was the result of a specific driving style element; however, not including these behaviors may lead to improper classifications. Ultimately, operationalizing a near-crash event by more strictly confining it to outcome and not behaviors (e.g., hard brake beyond threshold alone) would alleviate the described multicollinearity.

Impairment within naturalistic data is currently under-collected and methods of detecting the varying aspects of impairment are under-implemented. Distraction and fatigue are often limited to classification during reduction of event-based or baseline data, though computer vision methods for detecting these impairments is becoming an increasing reality among naturalistic driving data collections as well as in driver monitoring system implementations. These data could be used to further describe driving styles or other systematic behaviors on-road. Impairment related to alcohol or drug usage is also difficult to capture in naturalistic data, as is epidemiologic data, which makes attributions of driving style to crash likelihood more difficult. However, future efforts may evaluate the ability to detect alcohol- or drug-impaired driving through measuring variability in driving styles.

Conclusions and Recommendations

There have been many efforts to classify and relate a particular driving element—whether it be individual, infrastructure or environment-based—to crash risk. The implications of understanding crash risk are critical within the transportation industry, but also indirectly impact the daily lives the many individuals using the built roadway environment. Understanding how individuals make driving choices, such as the way they drive their vehicle across the various conditions they operate in, the errors of judgement made when driving impaired, or choosing what information is enough to make some roadway maneuver, could help researchers and industry make the roads safer. More information about driver behavior characteristics would enhance the ability to create training tools, performance metrics, infrastructure vehicle failsafe redundancies, or establish some other barrier to prevent or alleviate those identified poor driving habits and behaviors. While finding a solution seems unwieldy and untenable, exploring the variance within and between individuals' driving styles will be key in understanding how drivers deviate from nominal, non-crashworthy behaviors.

Toward a Real-time Evaluation of Human Performance

Extensive work is necessary to define and model crash risk for the in-situ vehicle operator. Existing models for real-time crash risk typically use traffic flows and infrastructure variables, and provide insight on when conflict will exist for that road segment. Adopting the estimates of crash risk from those models into an individual-centric model would allow for the crash risk to be calculated per road segment (and per other factors, such as traffic flow or speed) within the collective trip. These estimates would then be modified by adding additional layers of information into the model, including enduring person-focused variables (e.g., driving style, personality, risk-taking behaviors), temporal person-focused variables (e.g., mood, anxiety, fatigue), and situational variables (e.g., environment, trip type, traffic).

Ultimately, a cumulative risk score may be modeled to describe the main effect and interactive effects differing parameters have on an individual's crash risk. Risk can be described as time-, road-, or trip-based, as determined by the specificity of the data collected.

Recommendations

Several recommendations can be made based on the exploration of topics discussed in this study. These include recommendations regarding behavior-based predictive safety analytics, driving styles, environmental and traffic-based conditions, and crash risk. These recommendations exist for both research and industry and cover future data collection efforts, analyses, and data representations.

- Break down known crash precursors and correlate them to driving styles. Typical driving styles have been correlated to road safety elements, but little information exists as to how they relate to crash precursors. Speeding behaviors are typically operationalized within driving styles, which serves as a precursor. However, little is known about the propensity towards distraction, fatigue, or other impairment as a driving style, or how drivers operating a vehicle under those conditions alter their behaviors from their typical driving habits.

- Better define driving style-based behavioral indicators and associated metrics. A continual effort within naturalistic data use is to further create and refine useful metrics to evaluate safe and effective performance. The ongoing efforts to define behavioral indicators will produce additional content on the interaction between person-factors and situational contexts. Examples include when a person's typical driving style is characterized as safe and patient, but that person is in a situation that alters their driving behaviors, such as driving in an unfamiliar area of a city in which other drivers are aggressive. This situational combination of stressors may decrease the driver's safe behaviors in a way that would affect their category of driving styles more than others would be affected. This effect could manifest as increased distraction, decreased situational awareness, poorer driving performance, or any number of other different ways.

- Capture trip-level data across the industry. Current epidemiologic data excels at capturing information across local, state, and national levels. Crash, vehicle, and violation data are heavily represented through state and federal data collection efforts. Similarly, miles traveled, traffic flows, and infrastructure capacities are largely understood at a national level. Further, crash event data is becoming increasingly available through police accident reports, state tracking, or onboard monitoring systems implemented within CMV fleets. However, a shift toward collecting trip-level data, in conjunction with other data streams, would provide much more information that could be utilized effectively and efficiently to identify other systematic elements of roadway usage.

Additional Products

Education and Workforce Development Products

This project involved three graduate students. Dr. McDonald mentored two master's students on this project. Mr. Miller and Dr. Sarkar mentored one Virginia Tech master's student who worked closely with VTTI.

Technology Transfer Products

A number of planned publications are to be included as technology transfer products. These may include any of the following expected publication topics:

1. Calculating behavioral indicators of unsafe driving based on naturalistic driving data (focus on computational methods).
2. How do enduring personal characteristics cause crashes?
3. Predicting crash involvement based on enduring personal characteristics and observable driver behavior.
4. A prototype analytics tool for identifying unsafe drivers.
5. The role of different naturalistic driving datasets in developing crash risk indicators.
6. Comparing between dynamic Bayesian networks, support vector models, and baseline risk rates.
7. Illustrating how naturalistic driving data can be combined with public data to identify risk.

Data Products

This effort produced three datasets.

1. SHRP 2 sampled trip-level data. This dataset contains roughly 41,500 observations and calculated indices.
2. SHRP 2 person-level data. Variables include trip data aggregated at the person-level.
3. CMV-based person-level data. Minimal survey data are included in this dataset.

References

Abdel-Aty, M. A., Hassan, H. M., Ahmed, M., & Al-Ghamdi, A. S. (2012). Real-time prediction of visibility related crashes. *Transportation research part C: emerging technologies, 24*, 288-298.

Bharadwaj, N., Edara, P., & Sun, C. (2021). Sleep disorders and risk of traffic crashes: A naturalistic driving study analysis. *Safety Science, 140*. https://doi.org/10.1016/j.ssci.2021.105295

Braitman, K. A., & Braitman, A. L. (2017). Patterns of distracted driving behaviors among young adult drivers: Exploring relationships with personality variables. *Transportation research part F: traffic psychology and behaviour, 46*, 169-176.

de Winter, J. C., & Dodou, D. (2010). The Driver Behaviour Questionnaire as a predictor of accidents: A meta-analysis. *Journal of safety research, 41*(6), 463-470.

Demir, B., Demir, S., & Özkan, T. (2016). A contextual model of driving anger: A meta-analysis. *Transportation Research Part F: Traffic Psychology and Behaviour, 42*, 332-349.

Devlin, A., & McGillivray, J. (2016). Self-regulatory driving behaviours amongst older drivers according to cognitive status. *Transportation research part F: traffic psychology and behaviour, 39*, 1-9.

Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., & Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proceedings of the National Academy of Sciences, 113*(10), 2636-2641.

Dingus, T.A., Hankey, J.M., Antin, J.F., Lee, S.E., Eichelberger, L., Stulce, K.E., McGraw, D., Perez, M., Stowe, L. (2015). *Naturalistic driving study: Technical coordination and quality contro*l (No. SHRP 2 Report S2-S06-RW-1).

Dutta, N., & Fontaine, M. D. (2020). Assessment of the effects of volume completeness and spatial and temporal correlation on hourly freeway crash prediction models. *Transportation research record, 2674*(9), 1097-1109.

Elander, J., West, R., & French, D. (1993). Behavioral correlates of individual differences in road traffic crash risk: An examination of methods and findings. *Psychological Bulletin, 113*, 279–294.

Engström, J., Bärgman, J., Nilsson, D., Seppelt, B., Markkula, G., Piccinini, G. B., & Victor, T. (2018). Great expectations: a predictive processing account of automobile driving. *Theoretical issues in ergonomics science, 19*(2), 156-194.

Krum, A., Bowman, D.S., Soccolich, S., Deal, V., Golusky, M., Joslin, S., Miller, A. & Hanowski, R.J. (2016). *Federal Motor Carrier Safety Administration's Advanced System Testing Utilizing a Data Acquisition System on the Highways (FAST DASH) Safety Technology Evaluation Project# 2: Driver Monitoring, Final Report* (No. FMCSA-RRR-16-002). Federal Motor Carrier Safety Administration. Office of Analysis, Research, and Technology.

National Highway Traffic Safety Administration. Fatality Analysis Reporting System. https://www-fars.nhtsa.dot.gov/Main/index.aspx

Filtness, A. J., Hickman, J. S., Mabry, J. E., Glenn, L., Mao, H., Camden, M., & Hanowski, R. J. (2020). Associations between high caffeine consumption, driving safety indicators, sleep and health behaviours in truck drivers. *Safety science, 126*, 104664.

Fitch, G. M., Soccolich, S. A., Guo, F., McClafferty, J., Fang, Y., Olson, R. L., ... & Dingus, T. A. (2013). *The impact of hand-held and hands-free cell phone use on driving performance and safety-critical event risk* (No. DOT HS 811 757).

Guo, F., Fang, Y., 2013. Individual driver risk assessment using naturalistic driving data. *Accident Analysis and Prevention, 61*, 3-9. doi:10.1016/j.aap.2012.06.014

Guo, F., S. G. Klauer, J. M. Hankey, and T. A. Dingus. 2010. Near Crashes as Crash Surrogate for Naturalistic Driving Studies. *Transportation Research Record, 2147*, 66-74.

Hakkert, A. S., Gitelman, V. (2014). Thinking about the history of road safety research: Past achievements and future challenges. *Transportation Research Part F, 25*, 137-149.

Hammond, R.L.; Soccolich, S.A.; Han, S.; Guo, F.; Glenn, T.L.; and Hanowski, R.J. (August 2021). Analysis of naturalistic driving data to assess distraction and drowsiness in drivers of commercial motor vehicles. Report No. FMCSA-RRR-20-003. Washington, DC: Federal Motor Carrier and Safety Administration, USDOT.

Horswill, M. S., Hill, A., & Jackson, T. (2020). Scores on a new hazard prediction test are associated with both driver experience and crash involvement. *Transportation research part F: traffic psychology and behaviour, 7*1, 98-109.

Hu, L., Bao, X., Wu, H., & Wu, W. (2020). A Study on Correlation of Traffic Accident Tendency with Driver Characters Using In-Depth Traffic Accident Data. *Journal of Advanced Transportation, 2020*. https://doi.org/10.1155/2020/9084245

Knipling, R.R., (2004). *Individual differences and the" high-risk" commercial driver*. Transportation Research Board.

Knipling, R.R., (2009). *Safety for the long haul: Large truck crash risk, causation, & prevention.* American Trucking Association.

McClintock, M. (1936). Unfit for Modern Motor Traffic. *Fortune Magazine* (Aug. 1936).

McKenna, F. P. (1983). Accident proneness: A conceptual analysis. *Accident Analysis and Prevention, 15*, 65–71.

P. Cullen, H. Möller, M. Woodward, T. Senserrick, S. Boufous, K. Rogers, J. Brown, R. Ivers. (2021). Are there sex differences in crash and crash-related injury between men and women? A 13-year cohort study of young drivers in Australia. SSM – *Popular Health, 14* (2021), p. 100816

Ryan, A. H., & Warner, M. (1936). The effect of automobile driving on the reactions of the driver. *The American Journal of Psychology, 48*(3), 403-421.

Sagberg, F., Selpi, Piccinini, G. F. & Engström, J. 2015. A Review of Research on Driving Styles and Road Safety. *Human Factors, 57*(7), 1248–75.

Shinar, D. (Ed.). (2017). Traffic safety and human behavior. Emerald Group Publishing.

Simons-Morton, B.G., Zhang, Z., Jackson, J.C., Albert, P.S., 2012. Do Elevated Gravitational-Force Events While Driving Predict Crashes and Near Crashes? *American Journal of Epidemiology, 175*(10) , 1075–1079. doi:10.1093/aje/kwr440

Siren, A., & Meng, A. (2013). Older drivers' self-assessed driving skills, driving-related stress and self-regulation in traffic. *Transportation research part F: traffic psychology and behaviour, 17*, 88-97.

Stewart, T. (2022, March). *Overview of motor vehicle crashes in 2020* (Report No. DOT HS 813 266). National Highway Traffic Safety Administration.

Sun, D., Ai, Y., Sun, Y., & Zhao, L. (2020). A highway crash risk assessment method based on traffic safety state division. *PLoS one, 15*(1), e0227609.

Transportation Research Board. 2014. *Naturalistic Driving Study: Technical Coordination and Quality Control*. The National Academies Press. https://doi.org/10.17226/22362.

Velazquez, E. M. (2020). *Understanding Aggressive Driving Behavior: The Role of Personality and Individual Differences* [Bachelor's honors thesis, University of Central Florida].

Wong, I. Y., Smith, S. S., & Sullivan, K. A. (2015). The development, factor structure and psychometric properties of driving self-regulation scales for older adults: has self-regulation evolved in the last 15 years?. *Accident Analysis & Prevention, 80*, 1-6.

World Health Organization. (2018). *Global Status Report on Road Safety 2018.*

Xu, C., Wang, W., Liu, P., & Zhang, F. (2015). Development of a real-time crash risk prediction model incorporating the various crash mechanisms across different traffic states. *Traffic injury prevention, 16*(1), 28-35.

Yang, K., Wang, X., & Yu, R. (2018). A Bayesian dynamic updating approach for urban expressway real-time crash risk evaluation. *Transportation research part C: emerging technologies, 96*, 192-207.

Zhang, X., Zhao, X., Du, H., & Rong, J. (2014). A study on the effects of fatigue driving and drunk driving on drivers' physical characteristics. *Traffic injury prevention, 15*(8), 801–808.

Zuckerman, M., 1994. *Behavioral expressions and biosocial bases of sensation seeking.* Cambridge university press.