Accident Analysis and Prevention 156 (2021) 106086

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Investigating the relation between instantaneous driving decisions and safety critical events in naturalistic driving environment

Zulqarnain H. Khattak [a,*], Michael D. Fontaine [b], Wan Li [a], Asad J. Khattak [c], Thomas Karnowski [a]

[a] *Oak Ridge National Laboratory, Oak Ridge, 37830, TN, United States*
[b] *Virginia Transportation Research Council, Charlottesville, VA, 22903, United States*
[c] *University of Tennessee, Knoxville, TN, United States*

ARTICLE INFO

Keywords:
Safety critical events
Naturalistic driving
Driving volatility
Bollinger bands
Vehicle kinematics
Mixed logit
Driving instability

ABSTRACT

The availability of large-scale naturalistic driving data provides enormous opportunities for studying relationships between instantaneous driving decisions prior to involvement in safety critical events (SCEs). This study investigates the role of driving instability prior to involvement in SCEs. While past research has studied crash types and their contributing factors, the role of pre-crash behavior in such events has not been explored as extensively. The research demonstrates how measures and analysis of driving volatility can be leading indicators of crashes and contribute to enhancing safety. Highly detailed microscopic data from naturalistic driving are used to provide the analytic framework to rigorously analyze the behavioral dimensions and driving instability that can lead to different types of SCEs such as roadway departures, rear end collisions, and sideswipes. Modeling results reveal a positive association between volatility and involvement in SCEs. Specifically, increases in both lateral and longitudinal volatilities represented by Bollinger bands and vehicular jerk lead to higher likelihoods of involvement in SCEs. Further, driver behavior related factors such as aggressive driving and lane changing also increases the likelihood of involvement in SCEs. Driver distraction, as represented by the duration of secondary tasks, also increases the risk of SCEs. Likewise, traffic flow parameters play a critical role in safety risk. The risk of involvement in SCEs decreases under free flow traffic conditions and increases under unstable traffic flow. Further, the model shows prediction accuracy of 88.1 % and 85.7 % for training and validation data. These results have implications for proactive safety and providing in-vehicle warnings and alerts to prevent the occurrence of such SCEs.

1. Introduction

Every year worldwide, 1.2 million people die in crashes while 50 million people sustain injuries (Asirt, 2016). These crashes impose substantial economic and societal costs. Driver behavior and human factors are considered to be one of the major factors influencing these crashes, contributing to 94 % of these events (NHTSA, 2015). Thus, several researchers have investigated the relationship between behavioral attributes and crash risk over the past several decades. One major challenge when attempting to reduce these safety critical events (SCEs) involving crashes and near crashes is proactively identifying conditions leading to SCEs prior to their occurrence and then developing countermeasures to prevent them (Lefèvre et al., 2014). Traditionally, such analysis has relied on police reported crash statistics as a primary data source. This data has severe limitations of underreporting, errors, and lack of information on behavior prior to involvement in a SCEs.

Recent advances in information technology have enabled the collection of countless terabytes of vehicle kinematics and relevant human behavior data. The innovations in sensor technologies such as radar have made such real-world data collection a reality (Campbell, 2012; Henclewood, 2014). The emergence of this large-scale data provides the potential to understand microscopic driving behavior prior to occurrence of SCEs. Deviation from normal driving could be detected before the behavior culminates in SCE. The concept of driving volatility that captures the abrupt variations in acceleration/deceleration regimes and vehicular jerks is relevant in this context and could indicate these SCEs.

With the above background, this paper aims to investigate the role of

* Corresponding author.
E-mail addresses: khattakzh@ornl.gov, zk6cq@virginia.edu (Z.H. Khattak).

https://doi.org/10.1016/j.aap.2021.106086
Received 26 September 2020; Received in revised form 16 December 2020; Accepted 8 March 2021
Available online 18 April 2021

instantaneous driving decisions represented by volatility in SCEs. The SCE types in this study are defined as rear end, roadway departure, sideswipe, and head-on events. An observational study design is utilized to assess the association of driving behavior and volatility with normal driving and SCEs. The Naturalistic Driving Study (NDS) Database consisting of large-scale instantaneous driving data prior to SCEs occurrence (unsafe driving as opposed to normal driving) are analyzed for all events and new driving volatility measures are quantified. Further, the volatility measures are linked to SCEs, behavior and pre-event maneuvers, traffic flow, and roadway related parameters. Advanced statistical techniques were utilized to identify relationships that could be utilized to provide proactive alerts and warnings in a connected vehicle environment and advanced driving assistance systems. The study develops new measures of volatility based on Bollinger bands, which compare the relative moving average of speeds and accelerations to provide an efficient way of comparing volatility to the relative average value over time.

2. Literature review

A significant amount of literature has explored the relation between crash frequency/severity and traffic related factors (Khattak et al., 2017a, 2019a,b), roadway factors (Haghighi et al., 2018; Hu and Donnell, 2010), weather (Gaweesh and Ahmed, 2019; Naik et al., 2016) and driver behavior (Behnood and Mannering, 2017; Hassan et al., 2017; Khattak et al., 2018; Yan et al., 2008). Several studies have recognized driver behavior as a major contributory factors in crashes (Boyle et al., 2008; Hassan et al., 2017; Yan et al., 2008). In some cases, driver characteristics, age, and socioeconomic characteristics are used to identify relationships between driver behavior and crashes (Mitchell et al., 2014; Weng and Meng, 2012). Police reported crashes are primarily used to assess crash outcomes, which may not be truly representative of the circumstances surrounding the crash. Police crash reports don't provide necessary information about kinematics and driver behavior related factors. Some studies (Mannering, 2009; Scott-Parker and Oviedo-Trespalacios, 2017; Smorti and Guarnieri, 2014) have also analyzed the crash and driver behavior relationship using questionnaire surveys.

The emergence of naturalistic driving data has provided an opportunity to utilize terabytes of real-world second-by-second kinematic data available from sensors, radar, and video technologies to study safe and unsafe events. Further, the role of driver behavior prior to involvement in SCEs can also be studied. Several studies have used NDS data to assess the effects of driver distraction (Fitch et al., 2015; Rakauskas et al., 2004; Ye et al., 2017), weather impacts (Ghasemzadeh et al., 2018; Ghasemzadeh and Ahmed, 2018, 2017), and speed behavior (Richard et al., 2020). Further, kinematics have been used to identify crashes and near crashes (Ali et al., 2019; Kluger et al., 2016; Osman et al., 2019; Perez et al., 2017). Ali et al. (2019) studied detection of near crash events as a surrogate measure of crash risk for identification of near crashes in adverse weather conditions. The study utilized a time chunking technique with parametric and non-parametric methods to detect near crashes on freeway. The results revealed a good fit with logistic regression to detect near crashes. However, k-nearest neighbor and artificial neural networks showed higher detection accuracy for near crashes based on time slices. Likewise, another study (Kluger et al., 2016) proposed a technique to detect SCEs using vehicle kinematic data. Discrete Fourier transforms along with k-means clustering was utilized to flag crash or near crash patterns in a vehicle time series. Their algorithm detected around 78 % of crashes and near crashes. Perez et al. (2017) used naturalistic driving data to detect crash events by validating several kinematic thresholds. Their results indicated low sensitivity for their approach, with potential for improvement based on threshold. Osman et al. (2019) utilized vehicle kinematics to detect near crashes using naturalistic driving data. The authors tested several machine learning algorithms and observed Ada Boost to outperform all other algorithms with a recall of 100 % and 98 % precision. While these studies utilized NDS data, their focus was to use machine learning algorithms to detect crash and near crash events without controlling for the association of multiple confounding factors.

Recently, the concept of driving volatility, representing abrupt variations in driving behavior, has also been employed (Arvin et al., 2019; Wali et al., 2019; Wali and Khattak, 2020). Wali and Khattak (2020) used naturalistic driving data to analyze the association between volatility and crashes and near crashes in school zones. They observed that volatility in lateral and longitudinal direction is associated with probability of unsafe outcomes. Wali et al. (2019) utilized naturalistic driving data to study the relation between volatility in jerk in the lateral and longitudinal directions and crash/near crash events. A random parameters logit model revealed that positive vehicular jerk in the longitudinal direction increased crash risk by 15.79 % and 12.52 %. Arvin et al. (2019) used 617 events from naturalistic driving data to analyze the relation between volatility and crash intensity. Results from fixed and random parameter models revealed a higher likelihood of severe crashes with volatile driving.

While these studies have used naturalistic data to examine crash risk, there is a lack of guidance on how event-based volatility and unstable driving contributes to different safety critical event types such as rear end, roadway departure, sideswipe, and head on crash events. This study therefore utilizes vehicle kinematics from a large-scale microscopic naturalistic driving data to calibrate models that explore how event-based volatility and driving instability contributes to these SCEs.

3. Scope and contributions

This study utilizes a large-scale microscopic data from the SHRP 2 naturalistic driving study to understand the leading indicators of driving volatility and the association between event-based volatility and SCEs. The events involve normal driving baselines, crashes, near crashes, and pre-crash information about conditions that leads to SCEs. The research demonstrates how measures and analysis of driving volatility based on Bollinger Bands can be leading indicators of crashes and contribute to enhancing safety. Bollinger Bands compare the relative moving average of speeds and accelerations to provide an efficient way of comparing volatility to the relative average value over time. Volatility measures are available prior to involvement in SCEs and are leading safety indicators of such events. Highly detailed microscopic data are used to provide the analytic framework to analyze the behavioral dimensions that can lead to crashes and near crashes. While recent studies have analyzed the association between volatility and crashes/near crashes (Wali et al., 2019) and between volatility and crash intensity (Arvin et al., 2019), the association between volatility (event based) and specific crash types is unclear. This study therefore, addresses these gaps in the literature using SHRP2 naturalistic driving data to analyze the relation between speed, driving behavior, and roadway/environmental factors and involvement in specific crash types using new measures of driving volatility. The study accomplishes the following goals:

1 Analyze pre-crash vehicle kinematics data to study how driving behavior changes prior to involvement in specific types of SCEs as opposed to normal driving.
2 Analyze how unstable and volatile driving contributes to specific types of SCEs. The study develops a set of new event-based volatility measures (Bollinger Bands) to study the association of correlates of volatile driving with involvement in SCEs.
3 Analyze the impact of additional correlates related to driver distraction, traffic, and roadway factors on involvement in SCEs.

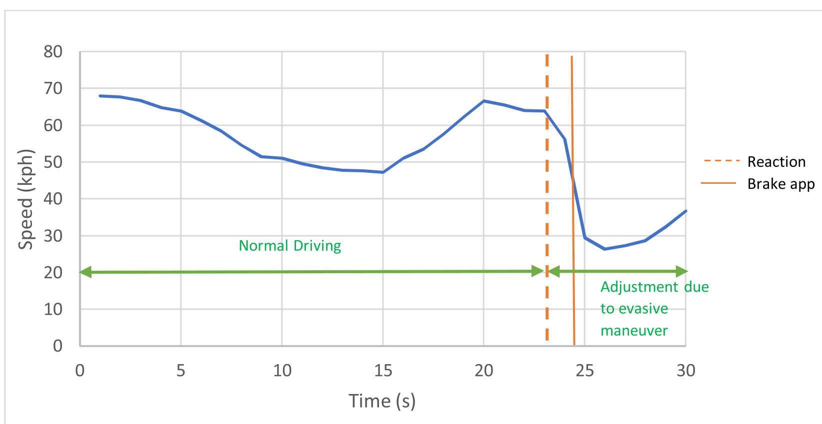
The analysis of such pre-crash driving behavior and stability is critical to devise a framework for providing alerts and messages to connected vehicles via communication medium (vehicle to vehicle and vehicle to infrastructure) or advanced driving assistance (ADAS) systems prior to their involvement in SCEs.

4. Concept of volatility measures

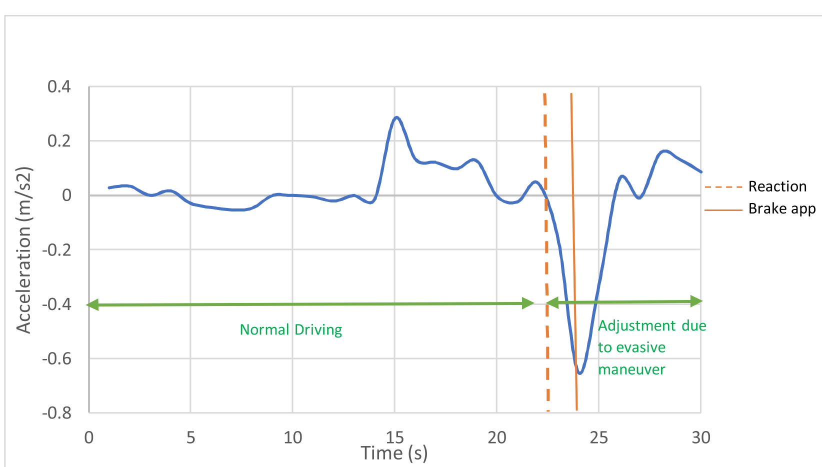
The concept of volatility captures the abrupt variations in instantaneous driving decisions that sheds light on actual maneuvers or behavior such as hard braking or jerky driving, which could lead to occurrence of SCEs. Market volatility has several indicators (Khattak et al., 2019a,b; Khattak et al., 2020a,b; Wang et al., 2015; Wali et al., 2019, 2018) including, coefficient of variation, jerks, mean absolute deviation and Bollinger Bands. The measures utilized in this research are different to standard deviation since these are vehicle kinematics based leading indicators of hazards, which makes the construct useful in predicting hazardous situations. The variations in driving regimes capture different aspects of volatility construct and serve as indicators which are derived from speeds, accelerations and decelerations. The concept of volatility was utilized since it serves as a surrogate for safety risks and lower driver comfort, due to representation of three-dimensional erratic vehicular movements. While coefficient of variation and vehicular jerk-based measures of volatility exist in the literature, this study develops new measures of volatility based on Bollinger bands, which compare the relative moving average of speeds and accelerations to provide an efficient way of comparing volatility to the relative average value over time. These were estimated for individual events (i.e., both crashes and near crashes).

The mathematical description for estimation of volatility measures are provided later in this section. Volatility in both the longitudinal and lateral direction were estimated separately to better quantify the instantaneous driving decisions in those directions. The measures of volatility in longitudinal and lateral direction evaluates different dimensions of vehicle kinematics. For instance, longitudinal acceleration measures the rate of change of velocity in the vehicles direction of travel (longitudinal) while lateral acceleration is the acceleration tangent to the direction of travel and acts when the vehicle corners and is pushed sideways. Lateral acceleration identifies vehicle motion during lane changes. Thus, estimation and analysis of measures of volatility based on both longitudinal and lateral acceleration is necessary to study driver behavior and erratic movements in different dimensions that could lead to occurrence of SCEs. Further, positive and negative acceleration values were separately used to estimate volatility measures.

The available kinematics data in NDS only has 30 s times series for crash/near crash events, which includes pre-event and post-event time. An SCE does not necessarily occur at 30 s of the time series. Example profiles of events shown in Figs. 1 and 2 indicate that an unsafe event occurs at 24 s, with the driver starting to react at around 22 s. Likewise, most of the events were observed to occur within the range of 24−26 s and driver reaction occurred at around 22 s. The reaction point was identified from the reaction time in the event data table and comparison with video clip for the specific event. The two sources showed consistency however, some margin of error may exist. We thus, normalized the markers at 25 s for event occurrence and 22 s for driver reaction. Further, the entire 30 s of timeseries aggregates two portions of volatility

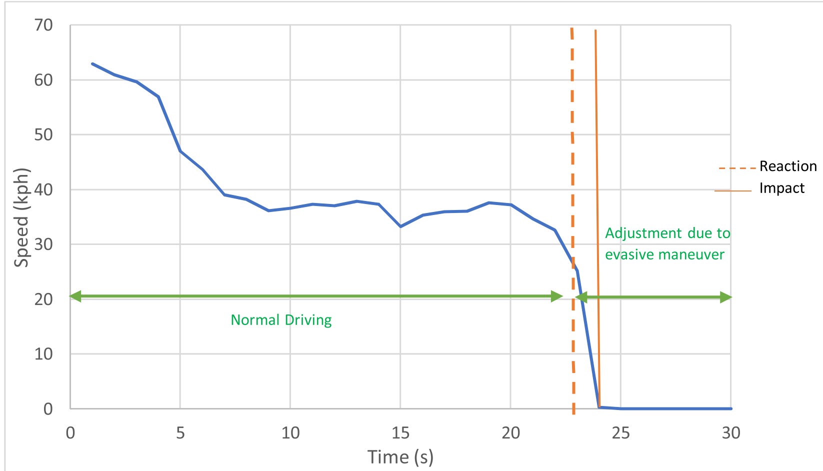


a) Speed Profile for near crash

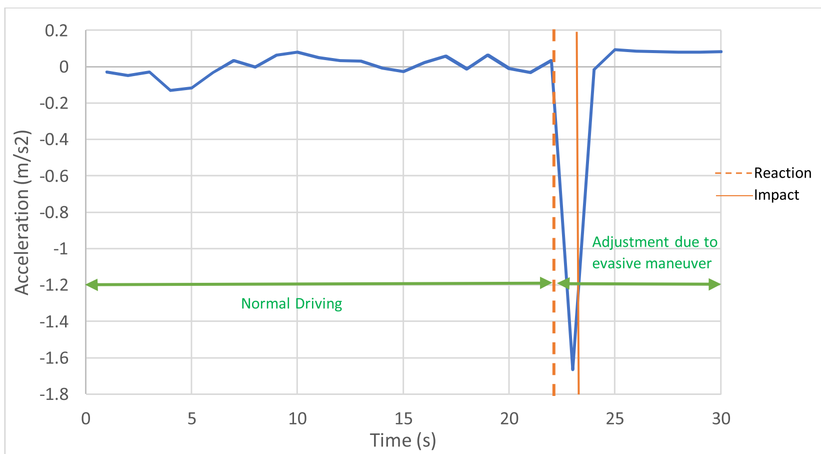


b) Acceleration Profile for near crash

Fig. 1. Kinematics for a near Crash event.



a) Speed Profile for crash



b) Acceleration Profile for crash

Fig. 2. Kinematics for a crash event.

arising from two different driver behaviors. The first portion represents the true driver behavior (speed choice, acceleration and vehicular jerk) when the driver does not anticipate an unsafe outcome. This portion is known as 'pre-event' volatility. This is the actual aggressive behavior when the driver is in control. The second portion involves the driver adjustment, reaction or evasive maneuver taken by the driver when they anticipate an unsafe event. The second portion is regarded as 'during/post event' volatility. The driver may have lost control during this second portion of volatility. Combining the two sources of volatility may result in bias thus, we censored the data to calculate intentional/pre-event volatility in the time period prior to the unsafe outcome and remove drivers' evasive reactions due to the anticipation of unsafe outcomes. Segmented measures of volatility were calculated for 20 s and 25 s along with the entire 30 s of the times series and used to analyze unsafe outcomes. The description of all measures is provided in later sections.

4.1. Coefficient of variation

The relative dispersion is shown by coefficient of variation, which is a ratio of standard deviation to the mean. The volatility measure is given in Eq. (1).

$$C_v = \frac{S.D}{\bar{x}} * 100\% \tag{1}$$

Where S.D provides standard deviation and speed and acceleration-deceleration mean is given by $\bar{x}$.

4.2. Bollinger bands

Bollinger bands (Lento et al., 2007) are estimated relative to the moving average of a specific measure (such as speeds and accelerations) and provides an efficient way of comparing volatility to the relative average value over time. These are estimated with standard deviations above or below the moving average value to indicate the level of volatility. The width of the band and standard deviation indicates higher volatility. The upper and lower bands were estimated by specifying two standard deviations above and below the moving average value given in Eqs. (2) and (3).

$$BB_u = MA + 2\sqrt{\frac{(x_i - MA)^2}{n}} \tag{2}$$

$$BB_l = MA - 2\sqrt{\frac{(x_i - MA)^2}{n}} \tag{3}$$

Where x_i is the instantaneous speed or acceleration, n represents the total number of observations, and MA is the moving average given by Eq. (4). The moving average (MA) was estimated for each 5 s within the

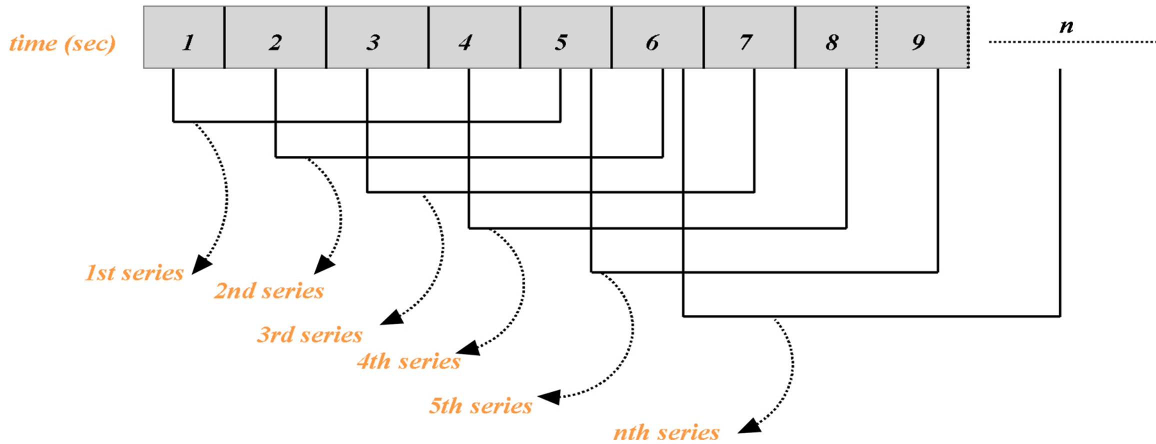


Fig. 3. Use of moving average to estimate Bollinger Bands.

20 s, 25 s and 30 s time series as shown in Fig. 3. For instance, the moving average was estimated for the first 5 s (1 s–5 s) of series, and then the moving average was estimated for the next 5 s (i.e. 2 s–6 s) of the series. This continued until the moving average was estimated for the entire time series of specific clip. This process was repeated for each time series of each event and provided an efficient way of utilizing the time series with changing volatility. These moving averages were used in estimating the upper and lower band (BB_u and BB_l) for Bollinger Bands.

$$MA = \sum_{i=1}^{n} \left(\frac{\sum_{i}^{i+4} x_i}{5} \right) \Big/ n \tag{4}$$

4.3. Vehicular jerk

The change rate of vehicular acceleration with time as a reference is quantified through vehicular jerk. Jerk may better represent volatile behavior in instantaneous driving decisions since it indicates the marginal change rate of acceleration/deceleration. Vehicular jerk is given by Eq. (5).

$$Jerk,\ J = \frac{\partial a}{\partial t} = \frac{\partial^2 v}{\partial t^2} = \frac{\partial^3 s}{\partial t^3} \tag{5}$$

Where $\frac{\partial}{\partial t}$ is the rate of change for the performance measure of acceleration (a), speed (v) and distance (s) with respect to time. ∂t is the small change in time.

5. Data description

This study utilized data from the naturalistic driving study (NDS), which was conducted as a part of Strategic Highway Research Program (SHRP 2) (Hankey et al., 2016). NDS was a comprehensive naturalistic study that included over 3500 driving participants from New York, Florida, Washington, Pennsylvania, Indiana, and North Carolina. About 4300 naturalistic years of driving was collected from 2010 to 2013 (Hankey 2013). The data collection was conducted using data acquisition systems (DAS) installed on vehicles, sensors, accelerometers and cameras to record the driver face/hand and roadway, kinematics (speeds, acceleration, steering position) at 10 Hz, lane offsets, and vehicle control (Hankey et al., 2016).

This study utilized events and continuous data out of multiple categories of data collected in SHRP 2. Detailed information regarding SCEs is provided within the event data. A unique aspect of the NDS is the availability of near crashes or misses along with crashes and baseline events. This provides critical information about SCEs including crashes and near crashes to identify unsafe driving relationships with different driver behaviors, environmental and roadway factors. The baseline events in these cases are required for comparisons with normal driving and are collected using a stratified sampling approach by stratifying participant and vehicle driving time proportions (TRB, 2013). The baseline events were selected by VTTI (Hankey et al., 2016) using a case cohort sampling approach by stratifying the participant and portion of time driven. The time driving was defined to include vehicle speeds above 5 mph in order to eliminate the effects of long stopping time and on periods of time when the vehicle was at risk of crash and near crash. The sample included all participants irrespective of their involvement in a crash or near crash. A minimum of one baseline was included for each driver in the study. Further description for baseline events is available in (Hankey et al., 2016). The SHRP2 NDS uses multiple approaches for identifying SCEs, but the triggering approach is the most systematic. This method used custom algorithms known as “triggers” to post process the incoming data (Hankey et al., 2016). Different thresholds were utilized for triggers, but longitudinal and lateral acceleration triggers are the most common. The trigger thresholds were selected based on acceleration in lateral direction above 0.75 g or lower/equal to −0.75 g with a duration exceeding 0.2 s were selected as trigger thresholds. Video verification was also conducted after trigger execution, which provides confidence in using this data. Further details are available in (Hankey et al., 2016; Jovanis et al., 2011).

The SHRP2 data used in this study included 9553 SCEs events after the data were cleaned and processed. These included 774 rear end events, 207 roadway departure events, 487 sideswipe events, and 525 head-on events, giving a total of 1993 SCE types, and 7565 baseline events. SCEs in this study refers to crashes and near crashes while SCE types refer to rear end, roadway departure, sideswipe and head on events. The data were selected for a set of events from Insight for similar roadways (including freeways, arterials, parking lots and driveways) and drivers and were provided by VTTI. These events do not indicate a complete trip, rather multiple events may be involved in a single trip. These events include around 1600 unique participants, with some participants appearing multiple times in the SCEs set. A 30 s continuous time series of data on vehicle kinematics collected at 10 Hz is available for the SCEs while a 20 s sec time series data collected at 10 Hz frequency is available for the baseline events. These data include acceleration in longitudinal direction and lateral direction, speed, gas pedal position, position of steering wheel, and status of wiper activation. A total of over 2 million temporal samples were utilized to estimate driving volatility measures.

Additionally, detailed information on driving behavior, maneuvers prior to incidents, event legality, secondary task times, traffic, and roadway parameters were collected. The detailed data (events) are linked together with the estimated volatility measures. A total of ten volatility measures were estimated using the vehicle kinematics data. The ten measures of volatility tested included coefficient of variation in longitudinal acceleration, coefficient of variation in lateral acceleration, upper Bollinger bands in longitudinal acceleration, lower Bollinger bands in longitudinal acceleration, upper Bollinger bands in lateral acceleration, lower Bollinger Bands in lateral acceleration, vehicular jerk in longitudinal acceleration, vehicular jerk in longitudinal deceleration, vehicular jerk in lateral acceleration, and vehicular jerk in lateral deceleration. Each of these measures were estimated using Eqs. (1)–(5) for the 20 s, 25 s and 30 s time series data. However, the estimated measures for only 20 s and 30 s time series are provided for the sake of brevity. Each of these measures analyzes different dimensions of vehicle

Table 1
Description of Volatility Measures, SCEs = Crash/Near crash.

Variable	Description	Mean Base/ SCEs	S.D Base/ SCEs	Min Base/ SCEs	Max Base/ SCEs	Sample Base/ SCEs
Volatility measures (based on 20 s data)						
Coefficient of variation-longitudinal acceleration	Measure of longitudinal volatility based on coefficient of variation	0.7745/ 0.8138/ 0.7782	0.3659/ 0.3784/ 0.3167	0.049/ 0.0430 0.0153	2.8284/ 3.1878/ 2.5681	7565/ 674/ 1319
Coefficient of variation-lateral acceleration	Measure of lateral volatility based on coefficient of variation	0.9297/ 0.8916/ 0.9235	0.3599/ 0.4009/ 0.4153	0.0428/ 0.0246/ 0.0159	2.8284/ 3.2943/ 4.4721	7565/ 674/ 1319
Longitudinal Vol Pos- Upper Bollinger Band	Measure of longitudinal volatility based on positive portion of Bollinger bands	0.0861/ 0.1252/ 0.1323	0.0682/ 0.0930/ 0.0916	0.0131/ 0.0134/ 0.0144	0.4307/ 0.6393/ 0.6451	7565/ 674/ 1319
Longitudinal Vol Neg- Lower Bollinger Band	Measure of longitudinal volatility based on negative portion of Bollinger bands	−0.0123/ −0.0227/ −0.0225	0.0216/ 0.0347/ 0.0348	−0.1795/ −0.1673/ −0.2010	0.2088/ 0.1997/ 0.1445	7565/ 674/ 1319
Lateral Vol Pos- Upper Bollinger Band	Measure of lateral volatility based on positive portion of Bollinger bands	0.0836/ 0.1293/ 0.0953	0.0961/ 0.1432/ 0.1151	0.0131/ 0.0134/ 0.0144	0.6193/ 0.9418/ 1.019	7565/ 674/ 1319
Lateral Vol Neg- Lower Bollinger Band	Measure of lateral volatility based on negative portion of Bollinger bands	−0.0209/ −0.0343/ −0.0248	0.0341/ 0.0585/ 0.0456	−0.2828/ −0.5717/ −0.5499	0.2574/ 0.0629/ 0.1232	7565/ 674/ 1319
Longitudinal Vol Pos-Vehicular Jerk	Measure of longitudinal volatility based on positive portion of vehicular jerk	0.0283/ 0.0470/ 0.0450	0.0198/ 0.0371/ 0.0331	0.0102/ 0.0115/ 0.0135	0.1819/ 0.3712/ 0.3122	7565/ 674/ 1319
Longitudinal Vol Neg-Vehicular Jerk	Measure of longitudinal volatility based on negative portion of vehicular jerk	−0.0295/ −0.0466/ −0.0472	0.0154/ 0.0320/ 0.0318	−0.1386/ −0.4995/ −0.3820	−0.0310/ −0.0029/ −0.0025	7565/ 674/ 1319
Lateral Vol Pos-Vehicular Jerk	Measure of lateral volatility based on positive portion of vehicular jerk	0.0286/ 0.0390/ 0.0304	0.0178/ 0.0335/ 0.0231	0.0131/ 0.0137/ 0.0142	0.2039/ 0.3518/ 0.2671	7565/ 674/ 1319
Lateral Vol Neg –Vehicular Jerk	Measure of lateral volatility based on negative portion of vehicular jerk	−0.0307/ −0.0433/ −0.0336	0.0182/ 0.0401/ 0.0251	−0.1789/ −0.4509/ −0.3161	−0.0114/ −0.0110/ −0.0103	7565/ 674/ 1319
Volatility measures (based on 30 s data)						
Coefficient of variation-longitudinal acceleration	Measure of longitudinal volatility based on coefficient of variation	0.7788/ 0.9066/ 0.8679	0.3138/ 0.36684/ 0.3064	0.0179/ 0.2213/ 0.1714	2.8284/ 3.3834/ 2.5559	7565/ 674/ 1319
Coefficient of variation-lateral acceleration	Measure of lateral volatility based on coefficient of variation	0.9360/ 1.1646/ 1.0696	0.3592/ 0.5152/ 0.4568	0.0346/ 0.1344/ 0.1332	2.8284/ 4.0272/ 4.7958	7565/ 674/ 1319
Longitudinal Vol Pos- Upper Bollinger Band	Measure of longitudinal volatility based on positive portion of Bollinger bands	0.0863/ 0.1688/ 0.1833	0.0680/ 0.1178/ 0.0907	0.0141/ 0.0246/ 0.0353	0.4307/ 1.948/ 0.7050	7565/ 674/ 1319
Longitudinal Vol Neg- Lower Bollinger Band	Measure of longitudinal volatility based on negative portion of Bollinger bands	−0.0127/ −0.0431/ −0.0434	0.0212/ 0.0657/ 0.0388	−0.1743/ −1.338/ −0.2435	0.2095/ 0.1082/ 0.1556	7565/ 674/ 1319
Lateral Vol Pos- Upper Bollinger Band	Measure of lateral volatility based on positive portion of Bollinger bands	0.0846/ 0.2178/ 0.1297	0.0960/ 0.1709/ 0.1309	0.0263/ 0.0274/ 0.0298	0.6193/ 1.105/ 1.067	7565/ 674/ 1319
Lateral Vol Neg- Lower Bollinger Band	Measure of lateral volatility based on negative portion of Bollinger bands	−0.0215/ −0.0848/ −0.0453	0.0346/ 0.0906/ 0.0621	−0.2828/ −0.6654/ −0.5944	0.2574/ 0.0884/ 0.0408	7565/ 674/ 1319
Longitudinal Vol Pos- Vehicular Jerk	Measure of longitudinal volatility based on positive portion of vehicular jerk	0.0283/ 0.0709/ 0.0833	0.0196/ 0.0396/ 0.0334	0.0147/ 0.0189/ 0.0231	0.1819/ 0.3378/ 0.3122	7565/ 674/ 1319
Longitudinal Vol Neg-Vehicular Jerk	Measure of longitudinal volatility based on negative portion of vehicular jerk	−0.0295/ −0.0695/ −0.0834	0.0153/ 0.0375/ 0.0321	−0.1434/ −0.4556/ −0.3035	−0.0105/ −0.0130/ −0.1134	7565/ 674/ 1319
Lateral Vol Pos-Vehicular Jerk	Measure of lateral volatility based on positive portion of vehicular jerk	0.0286/ 0.0587/ 0.0382	0.0176/ 0.0411/ 0.0276	0.0103/ 0.0135/ 0.0142	0.2039/ 0.4951/ 0.2230	7565/ 674/ 1319
Lateral Vol Neg -Vehicular Jerk	Measure of lateral volatility based on negative portion of vehicular jerk	−0.0306/ −0.0664/ −0.0416	0.0180/ 0.0461/ 0.0283	−0.1916/ −0.4509/ −0.2561	−0.0306/ −0.0291/ −0.0287	7565/ 674/ 1319

kinematics and identifies erratic movements that could lead to occurrence of SCEs. These measures were tested individually within the model using an iterative process to identify significant measures with predictive power and remove non-significant measures. This resulted in multiple models, however only the best fit models are presented in the paper.

The event data were manually checked with set of randomly selected video events. This verification and reasonableness of descriptive statistics from Tables 1–3 justified validity of the data. Table 1 provides the descriptive statistics for estimated measures of volatility. The descriptive statistics for driver behavior and lateral maneuver related variables are provided in Table 2 while Table 3 provides descriptive statistics for secondary task, traffic and roadway related variables. Each of these variables were tested for their significance and correlations and variables with better fit and lower correlation were selected.

6. Methodology

The association between volatility and SCE types is explored after estimating multiple volatility indices. The framework for the analysis is provided in Fig. 4. The outcome involves four SCE types including rear end, roadway departure, sideswipe, head on, and baseline, which serves as the dependent variable.

Considering the categorical nature of dependent variables, categorical probability models such as mixed logit can be applied as widely done in the literature (Khattak and Fontaine, 2020). The independent variables included volatility indices, driver behavior, secondary tasks, traffic density, roadway factors and environmental conditions. For the modeling framework, a crash function is developed to determine the proportions of the four SCE types and baseline outcomes using Eq. (6) as specified in (Milton et al., 2008).

Table 2
Data Description for Driver Behavior and Lateral Maneuver related Variables, *SCEs = Crash/Near crash.*

Variable	Description	Mean Base/ SCEs	S.D Base/ SCEs	Min Base/ SCEs	Max Base/ SCEs	Sample Base/ SCEs
Driver Behavior						
Aggressive	An indicator variable representing driver aggressiveness while maneuvering the vehicle (1=aggressive).	0.0026/ 0.0089/ 0.0318	0.0162/ 0.0939/ 0.1756	0/ 0/ 0	1/ 1/ 1	7565/ 674/ 1319
Distracted	An indicator variable representing driver distraction represented by involvement in unnecessary activities while maneuvering the vehicle (1=distracted)	0.2984/ 0.3601/ 0.3320	0.4216/ 0.4805/ 0.4711	0/ 0/ 0	1/ 1/ 1	7565/ 674/ 1319
Angry	An indicator variable representing driver behavior in terms of their level of anger due to personal or roadway characteristics while maneuvering the vehicle. This was determined from driver's expression of their feelings and their response to roadway conditions observed from video repositories. (1=Anger)	0.0027/ 0.0044/ 0.0182	0.0526/ 0.0666/ 0.1337	0/ 0/ 0	1/ 1/ 1	7565/ 674/ 1319
Inexperienced	An indicator variable representing drivers that are inexperienced in performing certain driving tasks or their general level of driving experience aggressiveness. This was determined from their level of experience and skills in maneuvering driving tasks.	0.0010/ 0.0459/ 0.0128	0.0325/ 0.2096/ 0.1128	0/ 0/ 0	1/ 1/ 1	7565/ 674/ 1319
Performing Safe Maneuver	An indicator variable representing whether the driver is performing a safe maneuver that may not cause harm to the driver or roadway users during the driving tasks. This was determined by ranking the maneuvers safe and unsafe from video repositories.	0.9323/ 0.7195/ 0.7581	0.2512/ 0.4495/ 0.4283	0/ 0/ 0	1/ 1/ 1	7565/ 674/ 1319
Performing Illegal Maneuver	An indicator variable representing whether a driver is performing a maneuver that is illegal and may cause harm to the driver or roadway users. (1=illegal maneuver)	0.0227/ 0.0252/ 0.0604	0.1491/ 0.1569/ 0.0776	0/ 0/ 0	1/ 1/ 1	7565/ 674/ 1319
Lateral Maneuvers						
Lane Changing	An indicator variable representing whether a driver is performing a lane changing maneuver. (1 = lane change)	0.0332/ 0.0326/ 0.0667	0.1791/ 0.1778/ 0.2496	0/ 0/ 0	1/ 1/ 1	7565/ 674/ 1319
Merging	An indicator variable representing whether a driver is merging onto mainline traffic from a ramp or merging between vehicles within mainline traffic. (1=merging)	0.0019/ 0.0074/ 0.0098	0.0444/ 0.0858/ 0.0988	0/ 0/ 0	1/ 1/ 1	7565/ 674/ 1319
Overtaking	An indicator variable representing whether a driver is performing an overtaking maneuver. (1=overtaking)	0.0444/ 0.0483/ 0.0985	0.0668/ 0.0385/ 0.0988	0/ 0/ 0	1/ 1/ 1	7565/ 674/ 1319
Negotiating a Curve	An indicator variable representing whether a driver is negotiating a curve. (1= in a curve)	0.0947/ 0.0727/ 0.0501	0.2929/ 0.2598/ 0.2181	0/ 0/ 0	1/ 1/ 1	7565/ 674/ 1319

Table 3
Data Description for secondary tasks, traffic and roadway related Variables, *SCEs = Crash/Near crash.*

Variable	Description	Mean Base/ SCEs	S.D Base/ SCEs	Min Base/ SCEs	Max Base/ SCEs	Sample Base/ SCEs
Secondary Tasks						
Duration of Secondary Task 1	Variable representing the amount of time the driver is involved in a secondary task while maneuvering the vehicle. Secondary tasks are classified into two categories and this represents the first category.	1.751/ 3.580/ 3.283	2.158/ 4.215/ 3.822	0/ 0/ 0	20.14/ 24.19/ 15.68	7565/ 674/ 1319
Duration of Secondary Task 2	Variable representing the amount of time the driver is involved in a secondary task while maneuvering the vehicle. This represents the second category of secondary tasks.	0.2472/ 0.7525/ 0.7634	0.9122/ 2.001/ 2.054	0/ 0/ 0	5.74/ 14.22/ 13.70	7565/ 674/ 1319
Texting	Indicator variable representing the whether a driver is texting. (1=texting)	0.0152/ 0.0401/ 0.0576	0.1223/ 0.1962/ 0.2331	0/ 0/ 0	1/ 1/ 1	7565/ 674/ 1319
Cell Phone Use	Indicator variable representing whether a driver is using cell phone. (1= cell phone use)					

(continued on next page)

Table 3 (*continued*)

Variable	Description	Mean Base/ SCEs	S.D Base/ SCEs	Min Base/ SCEs	Max Base/ SCEs	Sample Base/ SCEs
		0.0639/ 0.0875/ 0.1046	0.2447/ 0.2828/ 0.3061	0/ 0/ 0	1/ 1/ 1	7565/ 674/ 1319
Traffic Density						
Free Flow (LOS A)	Indicator variable representing free flow traffic on the roadway represented by level of service A. (1=free flow)	0.7504/ 0.7329/ 0.4366	0.4327/ 0.4427/ 0.4961	0/ 0/ 0	1/ 1/ 1	7565/ 674/ 1319
Stable (LOS B)	Indicator variable representing stable flow of traffic represented by level of service B. (1=stable flow)	0.2286/ 0.2373/ 0.4579	0.4201/ 0.4257/ 0.4984	0/ 0/ 0	1/ 1/ 1	7565/ 674/ 1319
Unstable (LOS F)	Indicator variable representing unstable flow or breakdown conditions represented by level of service F. (1=congested)	0.0194/ 0.0237/ 0.0985	0.1380/ 0.1523/ 0.2981	0/ 0/ 0	1/ 1/ 1	7565/ 674/ 1319
Roadway Features						
Divided Roadway	Indicator variable representing whether the roadway is divided. (1=divided)	0.4159/ 0.2210/ 0.3995	0.4929/ 0.4152/ 0.4899	0/ 0/ 0	1/ 1/ 1	7565/ 674/ 1319
Traffic Signals	Indicator variable representing whether traffic signals are presents (1=signal present)	0.0834/ 0.1647/ 0.1948	0.2765/ 0.3712/ 0.3962	0/ 0/ 0	1/ 1/ 1	7565/ 674/ 1319
Construction	Indicator variable representing whether the roadway has any construction activity. (1= under construction)	0.0179/ 0.0222/ 0.0371	0.1328/ 0.1476/ 0.1892	0/ 0/ 0	1/ 1/ 1	7565/ 674/ 1319
Stop Sign	Indicator variable representing whether the roadway has any stop signs (1= stop sign present)	0.0288/ 0.0697/ 0.0333	0.1673/ 0.2548/ 0.1796	0/ 0/ 0	1/ 1/ 1	7565/ 674/ 1319
Driveway	Indicator variable representing whether the roadway has driveways. (1=driveway present)	0.0698/ 0.0608/ 0.0409	0.2548/ 0.2392/ 0.1982	0/ 0/ 0	1/ 1/ 1	7565/ 674/ 1319
Interchange	Indicator variable representing whether the roadway has interchange. (1=interchange present)	0.0655/ 0.0326/ 0.0826	0.2475/ 0.1778/ 0.2754	0/ 0/ 0	1/ 1/ 1	7565/ 674/ 1319
Weather and Lighting						
Adverse Weather	Indicator variable representing whether any adverse weather conditions are present. (1=adverse weather present)	0.0970/ 0.1394/ 0.1167	0.2960/ 0.3467/ 0.3212	0/ 0/ 0	1/ 1/ 1	7565/ 674/ 1319
Dark	Indicator variable for whether the conditions are dark. (1= dark)	0.2384/ 0.2848/ 0.2229	0.4261/ 0.4517/ 0.4163	0/ 0/ 0	1/ 1/ 1	7565/ 674/ 1319

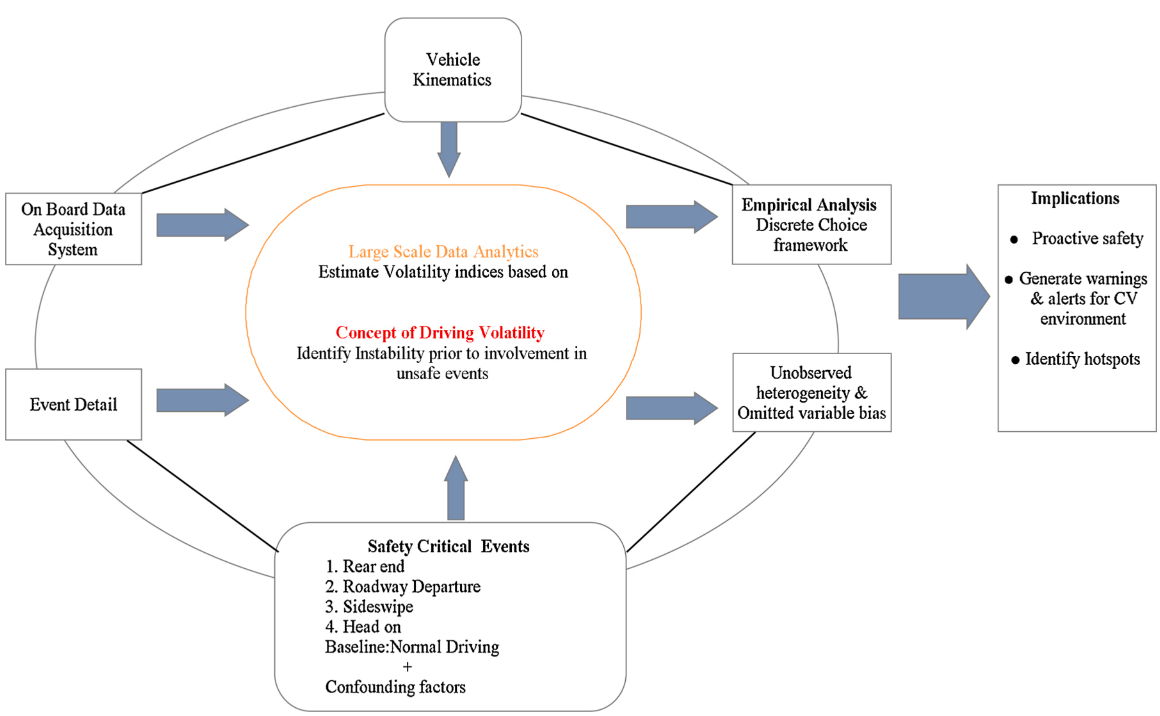


Fig. 4. Framework for Analysis of Event Types and Volatility.

$$Y_{ij} = \beta_i X_{ij} + \varepsilon_{ij} \quad (6)$$

where Y_{ij} is the function for determining SCE types and baseline events for crash j with all possible event categories i (such as rear-end, roadway departure, sideswipe, and head-on) (Ben-Akiva and Lerman, 1985; Washington et al., 2011). X_{ij} is the explanatory variable including volatility, driver behavior, or traffic density related variables, β_i is the variable coefficient and the unobserved effects arising due to driver and SCE characteristics are accounted using error term ε_{ij}. The multinomial logit model is given by Eq. (7) assuming generalized extreme value for ε_{ij} (McFadden, 1981):

$$P_j(i) = \frac{\exp(\beta_i X_{ij})}{\sum_I \exp(\beta_i X_{ij})} \quad (7)$$

Where the SCE outcome probability $P_j(i)$ for a crash (j) with all possible event categories i (such as rear-end, roadway departure, sideswipe, and head-on) resulting in SCE outcomes. The unobserved heterogeneity is account by allowing parameters variation across observations (variations in β). This is achieved using a mixing distribution, which provides the SCE estimates shown in Eq. (8) as follow.

$$P_j(i) = \frac{\exp(\beta_i X_{ij})}{\sum_I \exp(\beta_i X_{ij})} . f(\beta|\varnothing) d\beta \quad (8)$$

Where $\varnothing$ refers to vectors of parameters of density function including mean and variance and other term terms in Eq. (8), and the $f(\beta|\varnothing)$ estimates the joint density function of β. The density function $f(\beta|\varnothing)$ is used to estimate β, while accounting for unobserved heterogeneity (Milton et al., 2008). The values of β are drawn from density function $f(\beta|\varnothing)$ for given values of $\varnothing$ to estimate probabilities. The normal distribution $\beta_{ij \sim Normal(\beta_{j}, \sigma^2)}$ provided the best fit for random parameters (Ahmed et al., 2018; Xie et al., 2018) after considering multiple distributions. Due to the complexity of maximum likelihood estimation, simulation based maximum likelihood was utilized with 200 Halton draws (Halton, 1960; Train, 2003). The results were further interpreted using marginal elasticities, which provides interpretation for change in propensity of intermediate SCEs. The change in dummy variables are estimated using a unit difference of 0−1. The continuous variables are expressed as a change of one standard deviation from the mean. R Studio was used for modeling. Highly correlated variables having correlation coefficient (CC) above 0.90 were discarded (Khattak et al., 2017b; Washington et al., 2011). The selected variables had CC values lower than 0.80, which is consistent with the literature (Khattak et al., 2017b; Washington et al., 2011). The selected variables also had a variance inflation factor (VIF) less than 2, showing no multicollinearity issues. This is consistent with the literature studies (Chatterjee and Hadi, 2015; Greene, 2008) as VIF below 5 is utilized as a variable selection threshold.

7. Results and discussion

This section shows the modeling results for the association of volatility with involvement in SCE types. The SCE types within the dependent variable include rear end events, roadway departure events, sideswipe events and head-on events. The results are presented for the best fit models using both the volatility in the 20 s prior to involvement in SCE types (Table 4) and the 30 s prior to involvement in SCE types (Table 5). However, the discussion will mainly focus on volatility in driving regimes 20 s prior to involvement in SCEs since that represents the driver behavior corresponding to pre-event volatility prior to involvement in a SCEs. Further, the overall focus is to analyze the relation of pre-trip behavior and variables such as volatility, distraction, anger state with SCEs while other roadway features related factors serve as control.

Table 4
Modeling results for volatility indices with 20 s data, [RE]=Rear end, [RD]=Roadway departure, [SS]=Sideswipe, [HO]=Head-on.

Variable	Fixed Parameters Logit		Mixed Logit		Marginal Effects for Mixed Logit				
	Coef	p-value	Coef	p-value	Rear end	Roadway departure	Sideswipe	Head-on	Baseline
Volatility measures (based on 20 s data)									
Longitudinal Vol pos- Bollinger band [RE]	1.63	0.063	1.084	<0.01	0.0159	0.051	0.093	0.0237	−0.1836
Longitudinal Vol neg- Bollinger band [RE]	−2.34	0.055	−0.366	<0.01	−0.060	−0.097	−0.0262	−0.0174	−0.200
Lateral Vol pos-vehicular jerk [RD]	−0.79	0.061	0.644	0.06	0.063	0.0106	0.0342	0.0301	−0.137
Lateral Vol neg-vehicular jerk [RD]	1.63	0.11	0.402	0.07	0.0116	−0.0171	−0.045	−0.025	0.0757
Driver Behavior									
Subject at fault [RE]	1.18	0.08	1.64	<0.05	0.132	0.053	0.061	0.110	−0.356
Aggressive [SS]	1.21	0.18	0.626	<0.05	0.144	0.048	0.071	0.083	−0.346
Lane changing [RD]	0.17	0.072	1.024	<0.05	0.032	0.015	0.020	0.005	−0.072
Inexperienced [RD]	2.33	0.16	0.634	0.070	0.077	0.033	0.107	0.104	−0.321
Performing safe maneuver [HO]	−1.11	0.076	−1.333	0.054	−0.044	−0.017	−0.051	−0.051	0.163
Secondary Task (sec)									
Duration of secondary task 1 [HO]	0.104	<0.05	0.135	<0.05	0.008	0.002	0.006	0.005	−0.020
Duration of secondary task 2 [RE]	0.142	0.075	0.226	0.075	0.007	0.001	0.004	0.006	−0.019
Traffic Density									
Free flow (LOS A) [RE]	−0.13	0.141	−2.451	<0.01	−0.084	−0.024	−0.019	−0.009	0.136
Unstable (LOS F [SS]	–	–	−1.280	<0.01	0.039	0.009	−0.041	0.006	−0.013
Roadway Features									
Divided roadway [HO]	−0.734	0.084	−0.907	0.064	−0.001	0.004	−0.018	−0.037	0.053
Traffic signals [RE]	–	–	0.712	0.071	0.022	0.009	0.027	0.037	−0.094
Weather and Lighting									
Adverse weather [RD]	–	–	0.214	0.068	0.009	0.003	0.017	0.006	−0.035
Dark [HO]	0.414	0.084	0.212	0.077	−0.002	−0.006	0.006	0.009	−0.008
Unobserved effects (Standard deviation)									
Vol pos- Bollinger band	–	–	0.843	<0.01	–	–	–		
Vol pos- vehicular jerk			0.304	0.043					
Aggressive	–	–	0.574	<0.01	–	–	–		
Summary Statistics									
Log likelihood at Null	−7228.71		−7228.71		N//A	NA	N/A	N/A	N/A
Log likelihood at Convergence	−5693.4		−5482.3		N/A	N/A	N/A	N/A	N/A
AIC	2214		2143		N/A	N/A	N/A	N/A	N/A

Table 5
Modeling results for volatility indices with entire 30 s data, [RE]=Rear end, [RD]=Roadway departure, [SS]=Sideswipe, [HO]=Head-on.

Variable	Fixed Parameters Logit		Mixed Logit		Marginal Effects for Mixed Logit				
	Coef	p-value	Coef	p-value	Rear end	Roadway departure	Sideswipe	Head-on	Baseline
Volatility measures (based on 30 s data)									
Longitudinal Vol pos- Bollinger band [RE]	1.84	<0.01	1.63	<0.01	0.0702	0.0967	0.2761	−0.6469	0.0702
Longitudinal Vol neg- Bollinger band [RE]	−2.71	0.064	−2.34	<0.01	−0.0202	−0.0401	−0.0601	0.1204	−0.0202
Lateral Vol neg-vehicular jerk [RD]	−0.16	0.04	−0.79	<0.01	−0.0320	−0.0800	−0.0726	0.1846	−0.0320
Driver Behavior									
Aggressive [NI]	1.36	<0.05	1.21	<0.05	0.0276	0.0192	0.0434	−0.2037	0.0276
Lane changing [SS]	0.17	<0.05	0.17	<0.05	0.0101	−0.0078	0.0056	−0.0085	0.0101
Inexperienced [RD]	1.45	0.076	2.33	0.070	0.0223	0.0619	0.0769	−0.2240	0.0223
Performing safe maneuver [HO]	–	–	−1.11	0.054	−0.0106	−0.0251	−0.0339	0.1057	−0.0106
Secondary Task (sec)									
Duration of secondary task 1 [HO]	0.54	0.073	0.14	<0.05	0.0017	0.0035	0.0044	−0.0166	0.0017
Duration of secondary task 2 [RE]	–	–	0.12	0.075	0.0011	0.0026	0.0051	−0.0151	0.0011
Traffic Density									
Free flow (LOS A) [RE]	−1.02	0.14	−1.89	<0.01	−0.0192	0.0053	0.0031	0.0823	−0.0192
Unstable (LOS F [SS]	0.66	0.04	0.71	<0.01	0.0074	−0.0221	0.0064	−0.0216	0.0074
Roadway Features									
Divided roadway [HO]	0.41	0.097	−0.30	0.064	0.0056	−0.0056	−0.0200	0.0125	0.0056
Traffic signals [RE]	0.66	0.071	0.49	0.071	0.0048	0.0149	0.0230	−0.0558	0.0048
Weather and Lighting									
Adverse weather [RD]	0.55	0.11	0.36	0.068	0.0019	0.0095	0.0059	−0.0272	0.0019
Dark [HO]	0.47	0.092	0.33	0.077	−0.0032	0.0068	0.0106	−0.0155	−0.0032
Unobserved effects (Standard deviation)									
Vol pos- Bollinger band	–	–	0.48	<0.01	–	–	–		
Vol neg- Bollinger band	–	–	0.54	<0.01	–	–	–		
Summary Statistics									
Likelihood at Null	−7547.1		−7547.1		N/A	N/A	N/A	N/A	N/A
Likelihood at Convergence	−5970.3		−5834.9		N/A	N/A	N/A	N/A	N/A
AIC	2547		2466		N/A	N/A	N/A	N/A	N/A

7.1. Rear end events

Since volatilities are leading indicators of crashes that are available prior to involvement in these SCEs, they were a primary focus in this study. The involvement in SCEs shows a positive association with volatility in driving regimes represented by Bollinger Bands and vehicular jerks. Table 4 reveals that rear end events are observed to increase with increases in positive Bollinger Bands (upper bands) in the acceleration in longitudinal direction as opposed to the baseline events. The marginal effects show an increase of 1.59 % in rear end events with a unit increase in positive Bollinger Bands in acceleration in longitudinal direction. This is intuitive since deviation from normal driving behavior causes instability that increases the risk of occurrence of these SCEs. Similarly, rear end events are negatively affected with increases in negative Bollinger Bands in longitudinal acceleration. A decrease of 6% in rear end events is observed with a unit increase in Bollinger Bands in longitudinal acceleration. Likewise, the likelihood of rear end events increases with vehicular jerks in the lateral acceleration. Marginal effects show an increase of 6.30 % in rear end events with a unit increase in vehicular jerks in lateral acceleration. The variable shows variation across different events based on its random effects with a mean of 0.304 and standard deviation of 0.043. Rear end events increase significantly with driver fault, and an increase of 13.2 % in rear end events is observed with a unit increase in driver fault. Rear end events also increase with aggressive driving and lane changes. A unit increase in aggressive driving increases the likelihood of rear end event by 14.4 %. Likewise, rear end events increase by 3.20 % as a result of unit increase in lane changes. This is expected since lane changes are safety critical and such interactions creates disturbance in the traffic stream leading to increased collision risk with the following vehicles. Further, rear end event likelihood increases with increase in driver inexperience as opposed to baseline events. Marginal effects reveal a 7.7 % increase in risk of rear end events.

Rear end events also increase with driver involvement in secondary tasks. Secondary tasks here refer to non-driving related distraction and glances away from the direction of travel including involvement with the use of cell phone, texting, vehicular infotainments, and looking at pedestrians or roadway scenery (Klauer et al., 2006). An increase of 0.8 % in likelihood of rear end events is observed with a unit increase in duration of first secondary task while an increase of 0.71 % in likelihood of rear end events is observed with a unit increase in duration of second secondary task. These are likely intuitive findings since increased exposure to distraction related occurrences is expected to lead to higher propensity of SCEs. Flow of traffic also affects the occurrence of rear end events. The likelihood of rear end events increases by 8.4 % with a unit increase in free flow conditions corresponding to LOS A1 with no lead traffic in the stream and LOS A2 with minimal lead traffic. Unstable flow on the other hand increases the risk of rear end events by 3.9 % as opposed to baseline events. This is intuitive due to potential of conflicts resulting from such congested traffic stream. The likelihood of rear end events increases by 0.4 % with a unit increase in driving over divided highways. This may be attributed to lower interactions with traffic coming from opposite direction. Further, the propensity of rear end events increases by 2.2 % with increase in complex driving environment created by intersections due to the interaction of multiple vehicles.

7.2. Roadway departure events

The likelihood of roadway departure events is also observed to increase with increases in variations represented by Bollinger Bands and vehicular jerk. An increase of 4.10 % in likelihood of roadway departure events is observed with a unit increase in volatility represented by Bollinger bands in the longitudinal acceleration. Likewise, an increase of 1.06 % in roadway departure events is observed with a unit increase in positive vehicular jerk in lateral acceleration. These are intuitive findings since instability caused by deviation from normal driving behavior increases the risk of occurrence of roadway departures. The risk of roadway departure events also increases with aggressive driving and lane changes. An average increase of 4.80 % and 1.50 % in roadway departure events is observed with a unit change in aggressive driving and lane changes. Aggressive driving and lane changes causes instability in traffic stream leading to increased risk of roadway departures. Further, the risk of roadway departure events also increases with driver involvement in secondary tasks. An average increase of 0.1 % and 0.2 % in roadway departure events is observed with a unit change in involvement in first and second secondary task. These are intuitive findings since distracted drivers are expected to lead to unsafe events causing roadway departures.

Likewise, the risk for roadway departure events are expected to be influenced by roadway congestion. Marginal effects reveal a 2.4 % decrease in roadway departure events with a unit increase in free flow traffic conditions represented by LOSA1 with no lead traffic and LOSA2 with minimal lead traffic. Further, roadway departure events increase by 0.9 % with a unit increase in unstable flow conditions represented by LOS F. This is expected due to potential conflicts resulting from congested flow conditions. The propensity of roadway departure events increases by 0.4 % with a unit increase in driving over divided highways as opposed to baseline case. This finding however, appears to be confounded with volume. The propensity of roadway departure events increases by an average of 0.9 % with a unit increase in complex interaction created by presence of intersections as opposed to base line cases. Further, the risk of roadway departure events increases by 0.30 with a unit increase in adverse conditions and decreases by 0.6 % with a unit increase in darkness.

7.3. Sideswipe and Head-on events

Similarly, the risk of sideswipe and head on events increases by 9.3 % and 2.37 % with a unit increase in Bollinger bands in the longitudinal acceleration. Likewise, the sideswipe and head on events are negatively affected by vehicular jerks in lateral acceleration. These SCEs increase by 3.42 % and 3.01 % with a unit increase in vehicular jerk in the lateral acceleration. Both sideswipe and head on events are also affected by driver behavior related factors. Specifically, the propensity of these events increases with increase in aggressive driving, lane changing, and driving inexperience while the propensity of these events decreases with increase in safe driving maneuvers. Head on events are also affected significantly by driver fault representing an increase of 11 %. The propensity of sideswipe and head on events is also affected by distracted drivers and the propensity of these events increases by an average of 0.6 % and 0.5 % as opposed to baseline driving. The propensity of sideswipe and head on events decreases by an average 1.9 % and 0.9 % with increase in free-flow traffic condition with minimal or no leading traffic while sideswipe events reduce and head on events increases with a unit increase in unstable flow conditions represented by LOS F. Likewise, the risk of sideswipe and head on events is affected significantly by roadway features and weather and lightening conditions.

8. Model validation

The classification accuracy of the model was also tested to analyze its practical value. Forecast accuracy is mostly tested using the same data used to calibrate the models (training data). However, it is imperative to see how a model performs using a separate data that is not used to train the model (validation data). Thus, the model accuracy was also tested using a 20 % validation holdback data set. The model forecast accuracy was estimated based on confusion matrix in Table 6. The confusion matrix provides the number of observed and predicted outcomes for each SCE type (including rear end, road departure, sideswipe, and head on along with baseline driving) for the training and validation observations. The total number of observed and correct predictions are also provided for each category.

The forecast accuracy based on correct prediction is provided as function of correct predictions over all observations in specific categories. The results show an overall good prediction power of 88.1 % and 85.8 % over both training and validation data. However, the prediction accuracy is slightly lower for validation data as opposed to the training data. This is expected due to the rare nature of crashes and lower sample of such events in the validation data. However, the model predictive power is higher than those reported elsewhere (Khattak et al., 2019a,b; Wali et al., 2019; Zhao and Khattak, 2015). It should be noted that the simulation based discrete choice methods can be used for prediction, however they may not perform well compared to machine learning methods due to their inferential nature of capturing the complex dependencies in the data.

9. Limitations and future work

While NDS represents a rich set of sensor data with vehicle kinematics, the data may not be representative of all drivers since driver behavior varies from person to person. Further, the participants were compensated for their participation. The crashes utilized are also representative of some of the minor events that are not police reportable. Further, the vehicles utilized were not representative of all the driving vehicles on the road and may be in a better condition since they were accepted for the study. The data utilized also lacks sociodemographic variables such as age, gender, sex etc. for the driver participants. Although the current study accounts for unobserved heterogeneity, future research could incorporate these factors into the analysis. The event time for a subsample was manually checked with the video files, however, it was impractical to test all event times. Testing all event times may have resulted in different event times. Further, future research could also focus on different roadway types for SCEs and develop kinematic thresholds for SCEs.

10. Conclusions

The availability of large-scale sensor data provides enormous opportunities for studying the association of driver behavior prior to

Table 6
Confusion Matrix for Observed and Predicted Outcomes (20 s timeseries).

Observed Outcomes (Calibration)	Model Predictions (Calibration)					Total	Correct Predictions (%)
	Rear end	Road Departure	Sideswipe	Heads on	Baseline		
Rear end	422	45	67	59	26	619	68.1%
Road Departure	30	86	24	5	20	165	52.1%
Sideswipe	33	30	224	34	68	389	57.5%
Head on	30	17	24	252	97	420	60.0%
Baseline	104	66	62	70	5745	6047	95.0%
Total	619	244	401	420	5956	7640	88.1%
Observed Outcomes (Validation)	**Model Predictions (Validation)**					**Total**	**Correct Predictions (%)**
	Rear end	Road Departure	Sideswipe	Heads on	Baseline		
Rear end	101	13	22	9	10	155	65.1%
Road Departure	2	21	4	2	13	42	49.7%
Sideswipe	7	12	53	6	20	98	54.1%
Head on	6	3	5	61	30	105	58.3%
Baseline	7	15	47	37	1407	1513	93.0%
Total	123	64	131	115	1480	1913	85.8%

involvement in crashes with SCEs. The study harnessed large-scale (big) data from multiple sources; specifically, second by second kinematics data and event data into a manageable format for the analysis of SCEs. The events involve crashes, near crashes, and baselines and pre-crash information about conditions that lead to these events. The data is unique in the sense that such pre-crash information is not available in traditional police reported crashes. The research demonstrates how measures and analysis of driving volatility can be leading indicators of crashes and contribute to enhancing safety. Highly detailed microscopic data are used to provide the analytic framework to rigorously analyze the behavioral dimensions that can lead to different type of SCEs. The analysis provides a fresh and new look at determining the role of driver behavior in naturalistic and diverse settings. Specifically, this study investigates the role of driving instability in SCEs prior to involvement in such events. While past research has studied crash types and their factors, the role of pre-crash behavior in such events has not been explore as extensively. This study contributes by utilizing the concept of driving volatility to study the role of driving instability in specific types of SCEs.

The study utilized SHRP 2 naturalistic driving data to analyze pre-crash driving behavior. Modeling results reveal a positive association between volatility and involvement in SCE types such as roadway departures, rear end collision, head-on and sideswipes. Specifically, the propensity of all four types of events increases with increase in both lateral and longitudinal volatilities represented by Bollinger bands and vehicular jerk. Further, the propensity of rear end and roadway departure events is affected most and increases by 14.4 %, 4.8 % and 3.2 and 1.5 % in response to increase in driver behavior related factors such as aggressive driving and lane changing. Likewise, the risk of involvement in rear end event decreases under free flow traffic conditions while increases under unstable traffic flow. Further, the results showed an overall average prediction accuracy of 88.1 % and 85.7 % for training and test data.

The results have implications for proactive safety and impact transportation practitioners and developers of new technologies. Transportation practitioners can use the research to target roadway countermeasures at locations to potentially reduce volatility that comes from roadway geometry. Technology developers can provide in-vehicle warnings and alerts to prevent the occurrence of such SCEs well before the driver anticipates them. By monitoring instantaneous driving, alerts can be provided based on highly volatile driving. Such monitoring and alerts could potentially improve the performance of connected and automated vehicles (Khattak et al., 2020a,b) in the future and provide potential warnings to surrounding connected vehicles. However, further research is needed on this topic. The study can be used to cross-fertilize the volatility idea in other safety fields, e.g. occupational safety of commercial drivers and construction safety of workers, to obtain new insights into the mechanisms that underlie incidents or crashes and explore countermeasures.

Author contributions

The authors confirm contribution to the paper as follows: study conception and design: Z. Khattak, ; data collection: Z. Khattak, T. Karnowski; analysis and interpretation of results: Z. Khattak ; draft manuscript preparation and revision: Z. Khattak, M. Fontaine, W. Li, A. Khattak. All authors reviewed the results and approved the final version of the manuscript.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgements

The study was supported by Oak Ridge National Laboratory through a grant with Federal Highway Administration. The authors thank the supporting agencies. The authors also thank the two anonymous reviewers for their valuable comments. This manuscript has been authored by UT-Battelle, LLC, under Contract No. DE-AC05-00OR22725 with the US Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a nonexclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

References

Ahmed, M.M., Franke, R., Ksaibati, K., Shinstine, D.S., 2018. Effects of truck traffic on crash injury severity on rural highways in Wyoming using Bayesian binary logit models. Accid. Anal. Prev. 117, 106–113.
Ali, E.M., Ahmed, M.M., Wulff, S.S., 2019. Detection of critical safety events on freeways in clear and rainy weather using SHRP2 naturalistic driving data: parametric and non-parametric techniques. Saf. Sci. 119, 141–149. https://doi.org/10.1016/j.ssci.2019.01.007. November 2018.
Arvin, R., Kamrani, M., Khattak, A.J., 2019. The role of pre-crash driving instability in contributing to crash intensity using naturalistic driving data. Accid. Anal. Prev. 132 (April), 105226 https://doi.org/10.1016/j.aap.2019.07.002.
Asirt, 2016. Annual Global Road Crash Statistics. Association for Safe International Road Travel.
Behnood, A., Mannering, F.L., 2017. The eff ;ects of drug and alcohol consumption on driver injury severities in single-vehicle crashes. Traffic Inj. Prev. 18.
Ben-Akiva, M., Lerman, S.R., 1985. Discrete Choice Analysis. MIT Press.
Boyle, L.N., Tippin, J., Paul, A., Rizzo, M., 2008. Driver performance in the moments surrounding a microsleep. Transp. Res. Part F 11.
Campbell, K.L., 2012. The SHRP 2 naturalistic driving study: addressing driver performance and behavior in trafc safety. TR News 282.
Chatterjee, S., Hadi, A.S., 2015. Regression Analysis by Example, 5th editio.
Fitch, G.M., Bartholomew, P.R., Hanowski, R.J., Perez, M.A., 2015. Drivers' visual behavior when using handheld and hands-free cell phones. J. Safety Res. 105.
Gaweesh, S.M., Ahmed, M.M., 2019. Evaluating the safety effectiveness of a weather-based variable speed limit for a rural mountainous freeway in Wyoming variable speed limit for a rural mountainous freeway. J. Transp. Saf. Secur. 0 (0), 1–26. https://doi.org/10.1080/19439962.2019.1583707.
Ghasemzadeh, A., Ahmed, M.M., 2017. Drivers' lane-keeping ability in heavy rain: preliminary investigation using SHRP 2 naturalistic driving study data. Transp. Res. Rec. 2663.
Ghasemzadeh, A., Ahmed, M.M., 2018. Utilizing naturalistic driving data for in-depth analysis of driver lane-keeping behavior in rain: non-parametric MARS and parametric logistic regression modeling approaches. Transp. Res. Part C Emerg. Technol. 90.
Ghasemzadeh, A., Hammit, B.E., Ahmed, M.M., Young, R.K., 2018. Parametric ordinal logistic regression and non-parametric decision tree approaches for assessing the impact of weather conditions on driver speed selection using naturalistic driving data. Transp. Res. Rec. 2672, 2.
Greene, W., 2008. Econometric Analysis. Pearson Prentice Hall, Upper Saddle River, New Jersey.
Haghighi, N., Liu, X.C., Zhang, G., Porter, R.J., 2018. Impact of roadway geometric features on crash severity on rural two-lane highways. Accid. Anal. Prev. 111.
Halton, J., 1960. On the efficiency of evaluating certain quasi-random sequences of points in evaluating multi-dimensional integrals. Numer. Math. 2, 84–90.
Hankey, J.M., Perez, M.A., McClaff ;erty, J.A., 2016. Description of the SHRP 2 Naturalistic Database and the Crash, Near-Crash, and Baseline Data Sets. Virginia Tech Transportation Institute.
Hassan, H.M., Shawky, M., Kishta, M., Garib, A.M., Al-Harthei, H.A., 2017. Investigation of drivers' behavior towards speeds using crash data and self-reported questionnaire. Accid. Anal. Prev. 98.
Henclewood, D., 2014. Safety Pilot Model Deployment – One Day Sample Data Environment Data Handbook. Research and Technology Innovation Administration. Research and Technology Innovation Administration, McLean, VA.
Hu, W., Donnell, E.T., 2010. Median barrier crash severity: some new insights. Accid. Anal. Prev. 42 (6), 1697–1704. https://doi.org/10.1016/j.aap.2010.04.009.
Jovanis, P., Aguero-Valverde, J., Wu, K.-F., Shankar, V., 2011. Analysis of naturalistic driving event data: omitted-variable bias and multilevel modeling approaches. Transp. Res. Rec. 2236.
Khattak, Z.H., Fontaine, M.D., 2020. A Bayesian modeling framework for crash severity effects of active traffic management systems. Accid. Anal. Prev. 145 https://doi.org/10.1016/j.aap.2020.105544.
Khattak, Z.H., Magalotti, M.J., Fontaine, M.D., 2017a. Estimating safety effects of adaptive signal control technology using the empirical bayes method. J. Safety Res. https://doi.org/10.1016/j.jsr.2017.12.016.
Khattak, Z.H., Magalotti, M.J., Miller, J.S., Fontaine, M.D., 2017b. Using new mode choice model nesting structures to address emerging policy questions: a case study of the Pittsburgh Central Business District. Sustainability. https://doi.org/10.3390/su9112120.
Khattak, Z.H., Fontaine, M.D., Boateng, R.A., 2018. Evaluating the impact of adaptive signal control technology on driver stress and behavior using real-world

experimental data. Transp. Res. Part F 58, 133–144. https://doi.org/10.1016/j.trf.2018.06.006.

Khattak, Z., Magalotti, M., Fontaine, M., 2019a. Operational performance evaluation of adaptive traffic control systems: a Bayesian modeling approach using real-world GPS and private sector PROBE data. J. Intell. Transp. Syst. https://doi.org/10.1080/15472450.2019.1614445.

Khattak, Z.H., Fontaine, M.D., Smith, B.L., Ma, J., 2019b. Crash severity effects of adaptive signal control technology: An empirical assessment with insights from Pennsylvania and Virginia. Accid. Anal. Prev. 154, 151–162. https://doi.org/10.1016/j.aap.2019.01.008.

Khattak, Z.H., Smith, B.L., Fontaine, M.D., 2020a. Exploratory investigation of disengagements and crashes in autonomous vehicles under mixed traffic: an endogenous switching regime framework. IEEE Trans. Intell. Transp. Syst. https://doi.org/10.1109/TITS.2020.3003527.

Khattak, Z.H., Smith, B.L., Park, H., Fontaine, M.D., 2020b. Cooperative lane control application for fully connected and automated vehicles at multilane freeways. Transp. Res. Part C 111, 294–317. https://doi.org/10.1016/j.trc.2019.11.007.

Klauer, S., Dingus, T.A., Neale, V.L., Sudweeks, J.D., Ramsey, D.J., 2006. The Impact of Driver Inattention on Near-Crash/Crash Risk: an Analysis Using the 100-Car Naturalistic Driving Study Data. Department of Transportation, US.

Kluger, R., Smith, B.L., Park, H., Dailey, D.J., 2016. Identification of safety-critical events using kinematic vehicle data and the discrete fourier transform. Accid. Anal. Prev. 96, 162–168. https://doi.org/10.1016/j.aap.2016.08.006.

Lefèvre, S., Vasquez, D., Laugier, C., 2014. A survey on motion prediction and risk assessment for intelligent vehicles. ROBOMECH J. https://doi.org/10.1186/s40648-014-0001-z.

Lento, C., Gradojevic, N., Wright, C.S., 2007. Investment information content in Bollinger Bands? Appl. Financ. Econ. Lett. 3, 4.

Mannering, F., 2009. An empirical analysis of driver perceptions of the relationship between speed limits and safety. Transp. Res. Part F 12.

McFadden, D., 1981. Econometric models of probabilistic choice. In: Manski, D., McFadden (Eds.), A Structural Analysis of Discrete Data with Econometric Applications. The MIT Press, Cambridge, MA.

Milton, J.C., Shankar, V.N., Mannering, F.L., 2008. Highway accident severities and the mixed logit model: an exploratory empirical analysis. Accid. Anal. Prev. 40 (1), 260–266. https://doi.org/10.1016/j.aap.2007.06.006.

Mitchell, R., Bambach, M., Friswell, R., 2014. Work and non-work-related vehicle crashes: the contribution of risky driving practices. Saf. Sci. 68.

Naik, B., Tung, L.-W., Zhao, S., Khattak, A.J., 2016. Weather impacts on single-vehicle truck crash injury severity. J. Safety Res. 58.

NHTSA, 2015. Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey. National Highway Traffic Safety Administration, US Department of Transportation.

Osman, O.A., Hajij, M., Bakhit, P.R., Ishak, S., 2019. Prediction of Near-Crashes From Observed Vehicle Kinematics Using Machine Learning. https://doi.org/10.1177/0361198119862629.

Perez, M.A., Sudweeks, J.D., Sears, E., Antin, J., Lee, S., Hankey, J.M., Dingus, T.A., 2017. Performance of basic kinematic thresholds in the identification of crash and near-crash events within naturalistic driving data. Accid. Anal. Prev. 103, 10–19. https://doi.org/10.1016/j.aap.2017.03.005.

Rakauskas, M.E., Gugerty, L.J., Ward, N.J., 2004. Effects of naturalistic cell phone conversations on driving performance. J. Safety Res. 35, 4. https://doi.org/10.1016/j.jsr.2004.06.003.

Richard, C.M., Lee, J., Atkins, R., Brown, J.L., 2020. Using SHRP2 naturalistic driving data to examine driver speeding behavior. J. Safety Res.

Scott-Parker, B., Oviedo-Trespalacios, O., 2017. Young driver risky behaviour and predictors of crash risk in Australia, New Zealand and Colombia: same but diff ; erent? Accid. Anal. Prev. 99.

Smorti, M., Guarnieri, S., 2014. Sensation seeking, parental bond, and risky driving in adolescence: some relationships, matter more to girls than boys. Saf. Sci. 70.

Train, K., 2003. Discrete Choice Methods with Simulation, First ed. Cambridge University Press, Cambridge, United Kingdom.

TRB, 2013. The 2nd Strategic Highway Research Program Naturalistic Driving Study Dataset. SHRP 2 NDS InSight Data Dissemination.

Wali, B., Khattak, A.J., 2020. Harnessing ambient sensing & naturalistic driving systems to understand links between driving volatility and crash propensity in school zones – a generalized hierarchical mixed logit framework. Transp. Res. Part C 114. https://doi.org/10.1016/j.trc.2020.01.028.

Wali, B., Khattak, A.J., Bozdogan, H., Kamrani, M., 2018. How is driving volatility related to intersection safety? A Bayesian heterogeneity-based analysis of instrumented vehicles data. Transp. Res. Part C 92, 504–524. https://doi.org/10.1016/j.trc.2018.05.017. May 2017.

Wali, B., Khattak, A.J., Karnowski, T., 2019. Exploring microscopic driving volatility in naturalistic driving environment prior to involvement in safety critical events—Concept of event-based driving volatility. Accid. Anal. Prev. 132 (August), 105277 https://doi.org/10.1016/j.aap.2019.105277.

Wang, X., Khattak, A.J., Liu, J., Masghati-amoli, G., Son, S., 2015. What is the level of volatility in instantaneous driving decisions? Transp. Res. Part C 58, 413–427. https://doi.org/10.1016/j.trc.2014.12.014.

Washington, S.P., Karlaftis, M.G., Mannering, F.L., 2011. Statistical and Econometric Methods for Transportation Data Analysis. CRC Press Taylor adn Francis Group. https://doi.org/10.1198/tech.2004.s238.

Weng, J., Meng, Q., 2012. Eff ;ects of environment, vehicle and driver characteristics on risky driving behavior at work zones. Saf. Sci. 50.

Xie, M., Cheng, W., Gill, G.S., Zhou, J., Jia, X., Choi, S., 2018. Investigation of hit-and-run crash occurrence and severity using real-time loop detector data and hierarchical Bayesian binary logit model with random effects. Traffic Inj. Prev. 19, 2. https://doi.org/10.1080/15389588.2017.1371302.

Yan, X., Abdel-Aty, M., Radwan, E., Wang, X., Chilakapati, P., 2008. Validating a driving simulator using surrogate safety measures. Accid. Anal. Prev. 40.

Ye, M., Osman, O.A., Ishak, S., Hashemi, B., 2017. Detection of driver engagement in secondary tasks from observed naturalistic driving behavior. Accid. Anal. Prev. 106 (June), 385–391. https://doi.org/10.1016/j.aap.2017.07.010.