

# Quantifying the Effect of Roadway, Driver, Vehicle, and Location Characteristics on the Frequency of Longitudinal and Lateral Accelerations

Gibran Ali<sup>1</sup>, Shane McLaughlin<sup>1</sup>, and Mehdi Ahmadian<sup>2</sup>

<sup>1</sup>Division of Data and Analytics, Virginia Tech Transportation Institute,  
Blacksburg, VA 24061

<sup>2</sup>Department of Mechanical Engineering, Virginia Tech, Blacksburg, VA 24061

## Abstract

The purpose of this study is to understand and quantify the simultaneous effects of roadway speed category, driver age, driver gender, vehicle class, and location on the rates of longitudinal and lateral acceleration epochs. The rate of usual as well as harsh acceleration epochs are used to extract insights on driving risk and driver comfort preferences. However, an analysis of acceleration rates at multiple thresholds incorporating various effects while using a large scale and diverse dataset is missing. This analysis will fill this research gap. Data from the 2<sup>nd</sup> Strategic Highway Research Program Naturalistic Driving Study (SHRP2 NDS) was used for this analysis. The rate of occurrence of acceleration epochs was modeled using negative binomial distribution based generalized linear mixed effect models. Roadway speed category, driver age, driver gender, vehicle class, and location were used as the fixed effects and the driver identifier was used as the random effect. Incidence rate ratios were then calculated to compare subcategories of each fixed effect. Roadway speed category has the strongest effect on longitudinal and lateral accelerations of all magnitudes. Acceleration epoch rates consistently decreases as the roadway speed category increases. The difference in the rates depends on the threshold and is up to three orders of magnitude. Driver age is another significant factor with clear trends for longitudinal and lateral acceleration epochs. Younger and older drivers experienced higher rates of longitudinal accelerations and decelerations. However, the rate of lateral accelerations consistently decreased with age. Vehicle class also has a significant effect on the rate of harsh accelerations with minivans consistently experiencing lower rates.

## Nomenclature

$\beta$	Vector of regression parameters
$\lambda_i$	Rate of events per mile
$\mu_i$	Number of epochs experienced by the $i^{\text{th}}$ driver
$LCL$	The lower limit of the 95% confidence interval

$m_i$	Number of miles driven by the $i^{\text{th}}$ driver
$RR_{category\ A/category\ B}$	The ratio of rate of events observed for subcategory A versus subcategory B
$UCL$	The upper limit of the 95% confidence interval
$X_i$	Matrix of covariates for the $i^{\text{th}}$ driver
AD	Automated Driving
ADAS	Advanced driver-assistance systems
CAN	Controller Area Network
DAS	Data Acquisition System
GPS	Global Positioning System
SHRP2 NDS	The Second Strategic Highway Research Program Naturalistic Driving Study
SUV	Sports Utility Vehicle

## 1 Introduction

Every year in the United States, motor vehicle crashes cause thousands of fatalities, millions of injuries, and economic losses nearing a quarter trillion dollars (Webb 2018; Blincoe et al. 2015). The critical factor in a majority of crashes is the driver (93.5%), followed by environment (2.4%), vehicle (2%), and other reasons (2.1%) (Singh 2018; Dingus et al. 2016). Therefore, significant research has focused on understanding driver behavior and its relationship to safety critical events. Driver behavior has been studied using various observable measures such as acceleration, speeding, following distance, and driver attentiveness. All of these are important and shed light on different aspects of driving safety.

A number of studies have shown that a positive correlation exists between the crash rates of drivers and the frequency of harsh acceleration epochs (Elvik 2006; Klauer et al. 2009; Bagdadi and Várhelyi 2011; Simons-Morton et al. 2012; Wu and Jovanis 2013a; Wu and Jovanis 2013b; Bagdadi 2013; Tselentis, Yannis, and Vlahogianni 2017; Stipancic, Miranda-Moreno, and Saunier 2018). Most of these studies are based on naturalistic driving data as it provides insight into normal driving as well as safety critical events for the same set of drivers. The positive correlation between crash and harsh acceleration rates has been shown for both longitudinal as well as lateral acceleration events (Simons-Morton et al. 2012). However, when trying to explore these correlations, varied acceleration thresholds have been used to define harsh acceleration events and a consistent methodology does not exist (Kamla, Parry, and Dawson 2019). Moreover, most studies analyzing harsh acceleration events do not simultaneously control for the various factors that could be affecting driver acceleration behavior such as roadway properties, driver demographics, vehicle characteristics, and environmental features.

Age and driving experience are an important factor in crash risk. Considerable research has shown that young drivers under the age of 25 years and older drivers above the age of 74 years experience higher driving risk than the drivers between the ages of 25 to 74 years (Massie, Campbell, and Williams 1995; Ryan, Legge, and Rosman 1998; Curry et al. 2015; McCartt et al. 2009; *Traffic*

*safety facts annual report tables*; Webb 2018; Williams 2003; Simons-Morton et al. 2014; Ouimet et al. 2014; Guo et al. 2017; Tefft 2017; Choi et al. 2017; Ali et al. 2020a; Vankov, Schroeter, and Twisk 2021). Studies have also shown that driver age, experience, and other driver demographic factors affect the rate of harsh acceleration epochs (Simons-Morton et al. 2009; Lyu et al. 2018; Paefgen, Staake, and Fleisch 2014; Ali et al. 2020a). However, there is a need for further investigation to quantify the effect of age on the rates of acceleration events at multiple thresholds based on real-world naturalistic driving data.

A number of studies have shown that roadway properties affect driving risk as well the rate of harsh acceleration epochs (Paefgen, Staake, and Fleisch 2014; Petraki, Ziakopoulos, and Yannis 2020; Park et al. 2021). Roadway properties such as speed limit, type of access (e.g, controlled or not controlled), and functional class are major predictors for traffic speed and influence the frequency of stops and turns. Similarly, vehicle properties such as age and weight have been shown to affect driving risk and harsh acceleration frequency (Broughton 2008; Høye 2019). Therefore, it is important that these factors also be simultaneously taken into account with the driver demographics when making comparisons about driving behavior and risk.

The aggregated driving complexity of a location can be influenced by several factors such as population density, traffic conditions, urban versus rural proportions, weather, intersection density, etc. Since these factors vary by location, so does driving complexity. Several studies have shown that location of driving can also affect the driving risk and acceleration behavior (Kmet and Macarthur 2006; Paefgen, Staake, and Fleisch 2014; Park et al. 2021). This, again, shows that analyses comparing acceleration based driving behavior should control for the the location in which the driving occurs.

Even though many studies explore the different factors affecting driving behavior, some major gaps still exist. Most of the studies simultaneously account for only one or two factors affecting driver behavior. Also, when identifying harsh acceleration behavior, most studies use a simple threshold in various driving conditions. To fill these gaps in our understanding of acceleration behavior, a comprehensive analysis based on a large scale and representative dataset and accounting all the major factors was needed. This study aims to fulfill this need. In this paper, the longitudinal and lateral acceleration rates of drivers at multiple thresholds were modeled using generalized linear mixed effect models. The various factors such as roadway speed, driver demographics, vehicle properties and location were used as fixed effects. These models were then used to calculate the rate ratios for acceleration frequency at different thresholds between subcategories of roadway speed, vehicle type, location, driver age, and driver gender. The value of the rate ratio quantified how each factor affects acceleration behavior.

This research will be beneficial in a number of fields. With better understanding of age effects on acceleration behavior, policy makers can devise programs to better detect at risk younger and older drivers and devise training programs to help them. Vehicle system designers can use the rate ratio based quantification to better specify or tune their systems for different user demographics as well as roadway conditions. Various organizations that collect and process cohort based driving data such as insurance companies, transportation providers, and national traffic safety agencies can use this analysis to compare the trends in their user populations with a diverse and representative dataset.

## 2 Literature Review

Table 1 summarizes some of the major studies analyzing the relationship between acceleration based driving behavior, crash risk, driver demographics, vehicle characteristics, and roadway properties. Even though these studies have many major insights, Table 1 only summarizes the insights relevant to our discussion. The data source used by the researchers is also of importance and therefore has been briefly described as well.

### 2.1 Data Sources Used in Literature

The data sources described in Table 1 belong to one of the following categories:

1. Crash datasets such as the Fatal Accident Reporting System. These datasets are often maintained by federal agencies or police departments and offer details about hard to find crashes. However, they often need to be augmented with other datasets to estimate mileage based rates.
2. Naturalistic driving studies such as the 100 car NDS and the SHRP2 NDS. Large naturalistic driving studies are a great resource for studying driver behavior. However, finding rare and specific scenarios requires advanced data mining methods. Extremely rare driving events such as fatal crashes are often not captured in such studies.
3. Driving studies conducted on simulators such as the CARRS-Q Advanced Driving Simulator. Such studies are ideal when prior real world data does not exist or when behavior during rare and complex scenarios needs to be studied.
4. GPS traces from smartphones or in-vehicle data recorders. These studies offer a cheaper way to acquire datasets with many participants and large mileages. However, these datasets often lack the context around specific incidents offered by naturalistic driving studies. These studies can be augmented by other data sources such as insurance records to fill the missing context around crashes.
5. Field operation tests. In such studies, usually one vehicle is instrumented for data acquisition and multiple participants drive the same vehicle on the same roadways. These studies have been used to study specific predetermined scenarios in real world traffic conditions. However, the effect of the participant driving a unfamiliar vehicle and being observed by an experimenter may influence their behavior.

It is important to consider the data source along with the insights from each study as it affects the application of the results and the confidence in the conclusions.

### 2.2 Relationship Between Acceleration Behavior and Driving Risk

There is ample evidence to show that drivers with high crash rates have different acceleration behavior than drivers with low crash rates. Using the 100 Car Naturalistic Driving Study data, Klauer et al. 2009 show that drivers with higher rates of crashes and near crashes also have significantly higher rates of decelerations and lateral accelerations  $\geq 0.3g$ . While analyzing the effect of jerky driving on crash rates using two different naturalistic driving studies, Bagdadi and Várhelyi 2011 and Bagdadi 2013, show that there is a relationship between critical jerk and self

Table 1: A summary of existing literature about acceleration behavior and its relationship to various factors of interest.

Study	Data source	Driving behavior insights relevant to this study
Relationship between acceleration behavior and safety critical events		
Klauer et al. 2009	100 Car Naturalistic Driving Study with 109 participants and data collected for over a year	Drivers with higher rates of crashes and near crashes showed significantly higher rates of decelerations $\geq 0.3g$ and lateral accelerations $\geq 0.3g$ .
Bagdadi and Várhelyi 2011; Bagdadi 2013	Naturalistic driving study with 166 participants. 100 Car Naturalistic Driving Study with 109 participants and data collected for over a year.	A relationship exists between self reported accidents and jerky driving. For each additional critical jerk that the driver causes, a regression model showed that the number of accidents increases by about 1.13.
Simons-Morton et al. 2012	Teenage Naturalistic Driving Study with 42 newly licensed participants and data collected for 18 months.	Elevated g-force event rates can predict the crash and near crash likelihood in the near future. The correlation between the crashes/near crashes and elevated g-force rates was 0.60.
Stipancic, Miranda-Moreno, and Saunier 2018	Smartphone GPS data from 4,000 drivers for 21,000 trips map matched to roadways	Hard braking and hard acceleration events were positively correlated to historical crash frequency at the roadway level.
Relationship between driver behavior and driver demographics, roadway properties, and vehicle characteristics.		
Massie, Campbell, and Williams 1995	Passenger vehicle travel data from the U.S. 1990 Nationwide Personal Transportation Survey, fatal crash data from 1990 Fatal Accident Reporting System (FARS), and police accident reports from the 1990 General Estimates System to produce crash involvement rates.	Rates of police reported crashes as well as fatalities is higher for younger and older drivers in the U.S. Drivers between the ages of 30 and 64 years are the safest.
Ryan, and Legge, Rosman 1998	Rates of crash involvement from the Road Injury Database of the Road Accident Prevention Research Unit at the University of Western Australia. The database contains all crashes that resulted in police reports or hospital admissions.	Younger and older drivers in Western Australia had the highest rates of crashes. Women had higher rates of crashes than men for all age groups.

Table 1: A summary of existing literature about acceleration behavior and its relationship to various factors of interest.

Study	Data source	Driving behavior insights relevant to this study
Curry et al. 2015	A linked dataset that contains licensing and crash data for drivers in New Jersey, U.S.A. 410,230 drivers were selected who got their intermediate license at 17-20 years of age from 2006-2009.	Crash rates are influenced by a combined effect of age and driving experience. For young drivers, crash rates reduce with increased driving experience.
Webb 2018	A dataset cataloging the leading causes of death in the US available through the Web-based Injury Statistics Query and Reporting System (WISQARS) maintained by the Center of Disease Control and Prevention (CDC) USA.	Establishes that motor vehicle crashes are the leading cause of death for persons 16-24 years of age.
Guo et al. 2017	The SHRP2 naturalistic driving study with 3,542 participants accumulating 34.5 million miles of driving data over 3 years.	Younger and older drivers are more adversely affected by secondary-task engagement during driving than middle-aged drivers.
Choi et al. 2017	Dataset containing 1,614 police reported crashes that happened on 51 intersections in Seoul, South Korea.	Ageing drivers experience more crashes at intersections than younger drivers. This difference is greater when turning movements are involved.
Simons-Morton et al. 2009	Teenage Naturalistic Driving Study with 42 newly licensed participants and data collected for 18 months in the US.	Young drivers experienced higher rates of hard braking events during first month of licensure. The rates were also significantly higher when no passengers were present as compared to with adult passengers. Male drivers experienced higher rates of hard braking events as compared to female drivers.
Lyu et al. 2018	Field operation test with 46 participants based in Wuhan, China. Each participant drove the same vehicle once over the same loop with a 14 km warm up.	This study concludes that there are gender based differences in drivers. Male drivers recognize more driving risk but also show more aggressive driving tendencies.

Table 1: A summary of existing literature about acceleration behavior and its relationship to various factors of interest.

Study	Data source	Driving behavior insights relevant to this study
Ali et al. 2020a	A driving simulator study in the CARRS-Q Advanced Driving Simulator with 78 participants who drove in four driving conditions. (Ali et al. 2020b)	Age and gender-related effects are observed with young and male drivers having a higher probability of engaging in a hard-braking event when driving without driving aids.
Sharma et al. 2020	A driving simulator study in the CARRS-Q Advanced Driving Simulator with 78 participants who drove in two driving environments.	Acceleration noise reduces in a connected environment. Young drivers take more advantage of connected environment relative to the middle-aged or old drivers.
Paefgen, Staake, and Fleisch 2014	Data from in vehicle data recorders collected by a Pay as You Drive insurance provider. 600 vehicles were sampled from a larger set based on involvement in crashes. All the selected vehicles were from Italy. The data recorders aggregated GPS signals every 2,000 meters before transmitting the data for storage.	Crash risk varies by time of day, day of week, road type, and vehicle speed. Driving is riskier at night versus during the day, during weekdays versus the weekends, on low speed roads versus high speed roads. This study also shows that driving is the safest in the mid-range velocities of 60-90 km/h as compared to speeds above or below.
Petraki, Ziakopoulos, and Yannis 2020	Driving behavior data collected via smartphone sensors for 303 drivers while driving on two urban expressways in Athens, Greece. Traffic characteristics were obtained by 26 inductive loops on the two expressways.	Traffic characteristics such as traffic flow and speed have the most statistically significant impact on the rates of harsh acceleration events.
Broughton 2008	The British STATS19 national road accident reporting system in combination with the Vehicle Registration Mark to find the data of first registration, make, model, and weight of vehicles involved in an accident.	The driver casualty rate falls significantly with the size of the car. The rate of casualty increases with the time since first registration.
Høyen 2019	Norwegian crash data from 2000 to 2016.	There are fewer drivers killed or seriously injured in newer and heavier cars.

reported crash rates. Using GPS traces from smartphone data, Stipanovic, Miranda-Moreno, and Saunier 2018 show that a positive correlation exists between historical crash frequency and hard braking/acceleration events by analyzing the location of the events. Therefore, understanding the

factors that affect the frequency of harsh longitudinal and lateral acceleration events can help understand their role in crashes and near crashes as well.

## **2.3 Relationship Between Driving Behavior, Driver Demographics, Vehicle Characteristics, and Roadway Properties**

A multitude of studies have examined the factors affecting driver behavior, especially when it concerns driving risk. Massie, Campbell, and Williams 1995; Ryan, Legge, and Rosman 1998; Curry et al. 2015; Choi et al. 2017 and Webb 2018 use aggregated crash datasets to show that age is an important factor affecting crash risk in the US, Australia, and South Korea. In all the studies younger and older drivers had higher crash rates than middle aged drivers. Guo et al. 2017 use the SHRP2 NDS data to show that younger and older drivers are more adversely affected by secondary-task engagement during driving than middle-aged drivers.

Using the Teenage Naturalistic Driving Study, Simons-Morton et al. 2009 show that younger drivers experienced higher rates of hard braking events during the first month of licensure. The same study also shows that male driver experienced higher rates of hard braking events as compared to female drivers. Lyu et al. 2018 used data from a field operation test to show that male drivers recognize more driving risk but also show more aggressive driving tendencies. Data from a driving simulator based study is used by Ali et al. 2020a to show that young and male drivers have a higher probability of engaging in hard-braking events when driving without driving aids. Therefore, substantial evidence exists to show that driver age and gender can affect acceleration based driving behavior.

Using summarized GPS traces collected from vehicle data recorders Paefgen, Staake, and Fleisch 2014 show that crash risk varies by time of day, day of week, road type, and vehicle speed. Petraki, Ziakopoulos, and Yannis 2020 uses GPS traces collected via smartphone sensors and traffic data collected via inductive loops to show that traffic flow and speed have statistically significant impact on the rates of harsh acceleration events. Broughton 2008 and Høye 2019 use national crash data from Britain and Norway to show that vehicle size has an effect on driver casualty rates. Even though all these studies do not directly link to acceleration behavior to vehicle characteristics or roadway properties, sufficient evidence exists to show that such factors should be taken into account when studying acceleration behavior.

## **2.4 Gaps in Previous Studies**

There are four major gaps in our current understanding of acceleration behavior. First, to the best of the author’s knowledge, none of the existing studies examining acceleration behavior simultaneously account for the multiple factors that affect behavior such as driver age, driver gender, roadway properties, vehicle class, and location. Second, most of the current studies use a simple constant threshold for analyzing harsh acceleration events which cannot be accurate in all driving conditions. For example, the median driver in SHRP2 NDS experiences a 0.3 g braking event once every mile or so while driving on low speed neighborhood roads but once every 200 miles while driving on an interstate (Ali, McLaughlin, and Mehdi 2021). Third, none of the existing studies have relied on large scale and representative driving studies such as the SHRP2 NDS. Even when larger datasets were available, only a small sample was analyzed. Finally, to the best of the author’s knowledge, no existing studies analyze how the effect of earlier mentioned factors changes as the acceleration requirements for inclusion are increased from mild to harsh. This study aims to fill these gaps in our understanding of acceleration behavior.



### 3 Data

SHRP2 NDS is the ideal data set to effectively study the influence of roadway characteristics, driver demographics, and vehicle class on acceleration behavior. It is the single largest naturalistic driving study, consisting of over 34 million miles of driving conducted by about 3,500 participants recruited in six locations across the United States of America (Antin et al. 2011; Dingus et al. 2014; Hankey, Perez, and McClafferty 2016). In addition to its scale, the SHRP2 NDS was designed to capture diverse driver populations and driving conditions (Antin et al. 2015). Finally, the data richness created through high frequency vehicle kinematics, driver inputs, video feeds, and GPS signals make the study an unparalleled data source for this analysis.

Table 2 summarizes the number of participants and corresponding mileage driven for each factor and its subcategories under consideration in this analysis. Almost all subcategories have more than a million miles of driving which illustrates the scale, diversity, and richness of the data.

Most of the participants have driven on every roadway speed category subtype and the mileages are uniformly distributed. The roadway speed category represents the usual driving speeds on a road segment and was obtained from HERE.com digital map data through map matching (Here Technologies 2019)(McLaughlin and Hankey 2015). The map matching process uses latitude, longitude, and other vehicle timeseries data and finds the most appropriate matching segments from HERE.com digital maps. Since there are driving segments when the GPS system did not acquire meaningful data due to start up delay or low signal availability, such driving is marked as “Unknown” and has been excluded from the final analysis.

Even though there is a slightly higher number of female participants, SHPR2 NDS mileage is almost evenly split by gender. However, when it comes to age, this study was designed to over represent younger and older drivers in comparison to the national driving population. The difference is significant but will not affect the inferences drawn in this analysis as both age range and gender and examined as fixed effects in the analysis. The date of birth and gender were declared by the participants at the time of enrollment. Each driver was assigned an age group based on their age at the time of enrollment in the study. This was an additional measure to further obfuscate any personally identifiable information (PII) about the participants. It should be noted that the minimum age for issuance of driving license in the United States can vary by state. However, all collection sites for SHPR2 NDS were in states that allowed 16 year old drivers to apply for a restricted license.

All the vehicles were assigned one of four vehicle classes with cars being the most common subcategory, followed by SUVs/crossover, pickup trucks, and vans/minivans. These categories are informative but should not be considered monolithic as seemingly different vehicles can have the same classification. There were 6 collection sites setup to recruit participants and swap filled hard drives. It should be noted that the actual location of driving strongly correlates with these collection sites but may, in some cases, be different. Given that there is enough data in each subcategory and that vehicle class and location are examined as fixed effects, the over representation of one category should not affect the inferences drawn during analysis.

To analyze the effect of the above mentioned factors on acceleration behavior, a comprehensive dataset of acceleration epochs in SHRP2 NDS was needed. This requirement was fulfilled by The Surface Accelerations Reference which was created to catalog all longitudinal and lateral acceleration epochs experienced by drivers in SHRP2 NDS (Ali, McLaughlin, and Mehdi 2021). Using an algorithm, researchers discovered over 1.2 billion epochs and summarized them into multi-dimensional data points stored in an SQL queryable DB2 database. Each datapoint has multiple measures for acceleration, vehicle speed, driver inputs, and roadway properties. These summarized

Table 2: SHRP2 NDS participant and mileage summary by roadway speed category, driver gender, driver age, vehicle class, and collection site.

Factor	Sub categories	Number of participants	Number of million miles
Roadway speed category	0-30 MPH	3,545	4.79
	31-40 MPH	3,540	5.39
	41-54 MPH	3,535	6.48
	55-64 MPH	3,527	7.79
	65-80 MPH	3,159	5.17
	Unknown	3,546	4.89
Driver gender	Male	1,563	16.85
	Female	1,705	16.52
	Did not specify	278	1.12
Driver age range	16-19	547	4.45
	20-24	742	8.10
	25-29	278	3.29
	30-34	164	1.89
	35-39	128	1.32
	40-44	116	1.42
	45-49	147	1.72
	50-54	167	1.77
	55-59	144	1.47
	60-64	151	1.43
	65-69	211	2.11
	70-74	174	1.64
	75-79	269	1.96
	80-84	156	1.19
	85+	70	0.351
	Did not specify	82	0.40
Vehicle class	Car	2,557	24.48
	Pickup Truck	170	1.84
	SUV - Crossover	721	6.84
	Van - Minivan	152	1.34
Collection site	Bloomington, IN	280	2.69
	Buffalo, NY	784	7.77
	Raleigh, NC	576	6.19
	Seattle, WA	835	7.36
	State College, PA	285	2.34
	Tampa, FL	787	8.15

data points were then used to calculate three types of measures for each driver separated by roadway properties:

- Rate of events stronger than a threshold per mile. For example, rate of deceleration events stronger than 0.5 g per mile experienced on roadways with <30 mph speed category.
- Event magnitude at key percentiles. For example, magnitude of 95th percentile acceleration experienced on roadways with 31-40 mph speed category.
- Strongest event magnitude experienced at key measures of distance. For example, strongest right leaning lateral acceleration experienced in 100 miles of driving.

An online interactive tool was created to let users analyze and download the data (Ali, McLaughlin, and Mehdi 2021). Through this tool, user can visualize the above mentioned measures and also visually compare the effect of roadway properties, driver demographics, and vehicle class.

Using the same data, Figure 1 compares the distribution of longitudinal and lateral acceleration rates. This is done by plotting the rates for 5<sup>th</sup>, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup>, and 95<sup>th</sup> percentile drivers along the y-axis. The x-axis represents the acceleration threshold at which the rates are calculated and the data is separately plotted for each speed category. Since the y-axis spans over five orders of magnitude, it shows that even for the median drivers, some acceleration events are experienced about 10 times per mile whereas other can be as rare as once every 10,000 or more miles. The purpose of this figure is to set the general context of rarity by threshold and roadway type. This context is important when used with rate ratio plots in the results section.

For example, consider the deceleration rates for the median driver on  $\leq 30$  mph roadways, which mostly represent driving on neighborhood and minor collector roads. Deceleration  $\geq 0.1$  g and  $\geq 0.2$  g occur multiple times per mile and are more representative of the traffic conditions than driver preferences.  $\geq 0.3$  g deceleration occur about once every two miles and may be influenced by driving style and traffic conditions equally.  $\geq 0.4$  g,  $\geq 0.5$  g and  $\geq 0.6$  g deceleration epochs occur every 20, 100, and 1,000 miles respectively. These are likely to be created by the driver’s response to unusual or rare incidents. Drivers with lower rates for such thresholds are either driving in safer environments, better at observation and prediction, or have superior vehicle control skills. The median driver in SHRP2-NDS never experienced a  $\geq 0.7$  g deceleration.

Figure 1 also illustrates that irrespective of acceleration type, as the threshold is increased, the distribution changes towards rates becoming lower. Also, the difference between the 25<sup>th</sup> and the 75<sup>th</sup> percentile driver grows with higher thresholds. Therefore, the rate distribution for each of the four acceleration types at every threshold should be modeled separately.

Even though the rates data provided in the accelerations reference informs on the differences by roadway, it does not quantify the individual contributions of various other factors in the rates of events. For example, if drivers in a particular age group have elevated rates of deceleration, is it because they primarily drove on lower speed roads or because that age group has a higher likelihood of producing such rates? Therefore, there is a need for quantifying the independent contributions of the roadway, vehicle, location, and driver attributes. This has been achieved using generalized linear mixed effect models in the analysis described in the next few sections.

## 4 Methods

The purpose of this analysis was to quantify the difference in rates of acceleration epochs based on roadway speed category, driver age, driver gender, vehicle class, and collection site location. To

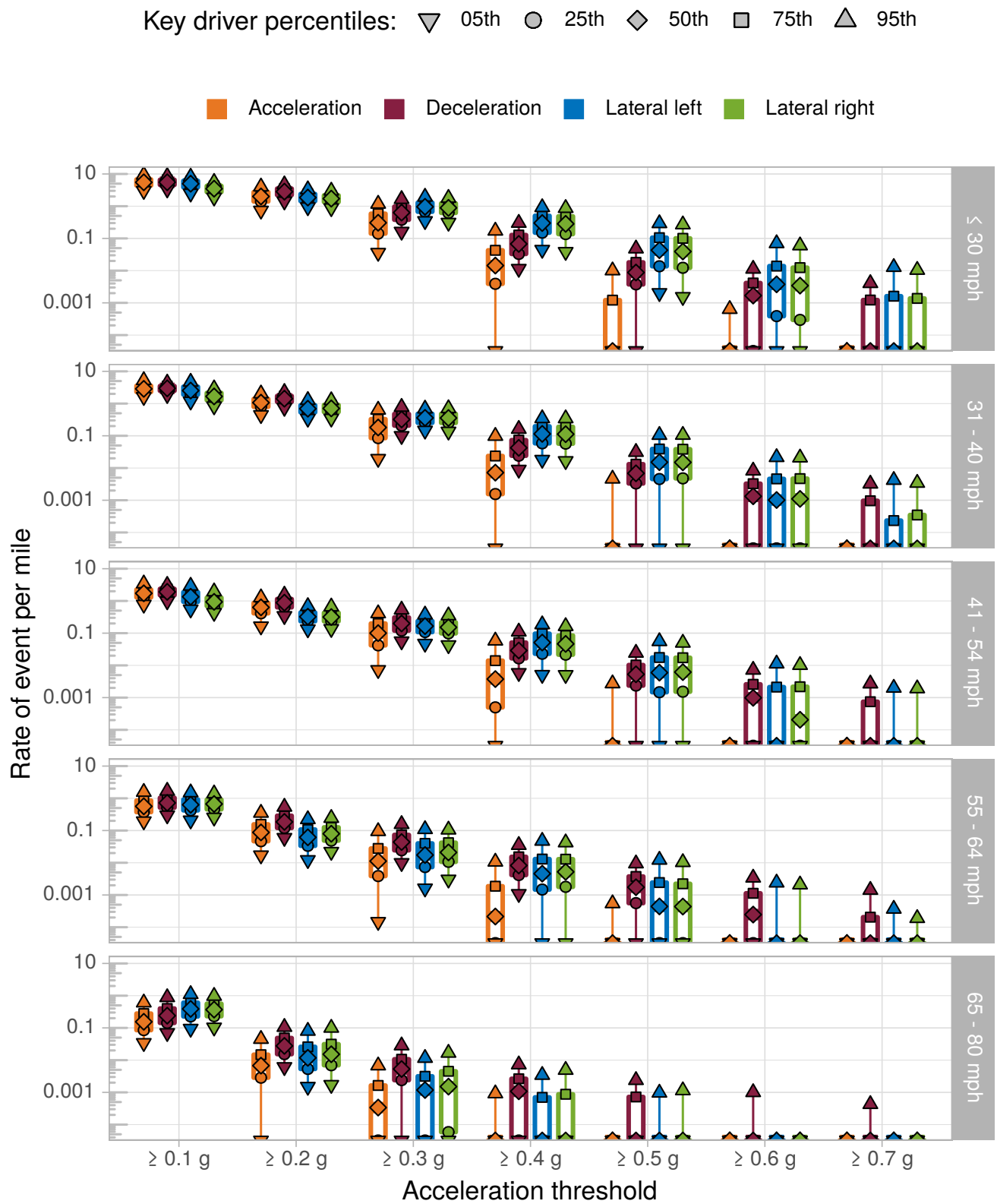


Figure 1: Distribution of acceleration rates for SHRP2 NDS drivers at various thresholds separated by roadway speed category.

illustrate the methodology, consider the simpler analysis of comparing rate of strong deceleration on two types of roadways with speed categories of “ $\leq 30$  mph” and “65 - 80 mph”.

Let  $\lambda_{\leq 30 \text{ mph}}$  be the rate of deceleration epochs  $\geq 0.5$  g on roadways with speed category “ $\leq 30$  mph” and let  $\lambda_{65 - 80 \text{ mph}}$  be the rate of deceleration epochs  $\geq 0.5$  g on roadways with speed category “65 - 80 mph”. Therefore, the rate ratio between the two is given by,

$$RR_{\frac{65 - 80 \text{ mph}}{\leq 30 \text{ mph}}} = \frac{\lambda_{65 - 80 \text{ mph}}}{\lambda_{\leq 30 \text{ mph}}} \quad (1)$$

To show that “65 - 80 mph” roadways produce lower rates of deceleration epochs than “ $\leq 30$  mph” roadways, the value of  $RR_{65 - 80 \text{ mph}/\leq 30 \text{ mph}}$  needs to be less than 1. However, to show that this would be true for 95% of such studies conducted, the upper limit of a 95% confidence interval for the rate ratio also needs to be less than 1 as well. Therefore, for conclusive evidence of difference in behavior, both the rate ratio as well as the 95% confidence interval need to be calculated. For this analysis, these measures have been calculated using negative binomial distribution based generalized linear mixed effect models which are often used for over dispersed data (Lord and Mannering 2010; Guo and Fang 2013).

There are four acceleration types under consideration and six thresholds ranging from  $\geq 0.1$  g to  $\geq 0.6$  g. For each combination of acceleration type and threshold, one mixed effect model was fitted to the data. The fixed effects were:

- roadway speed category
- vehicle class
- collection site location
- driver age
- driver gender

Since each driver drove over multiple roadways, the unique driver and vehicle identifier was used as the random effect. If  $\lambda_i$  is the rate of a particular acceleration type for a given threshold, the mixed effect model can be described as

$$\log \lambda_i = X_i\beta + \mathcal{Z}u_i \quad (2)$$

where  $X_i$  is the matrix of predictor variables for the driver,  $\beta$  is the column vector of fixed-effects regression coefficients for all drivers,  $\mathcal{Z}$  is the design matrix for random effects and  $u_i$  is the random effect for the driver. The rate  $\lambda_i$  can also be expressed as

$$\lambda_i = \frac{\mu_i}{m_i} \quad (3)$$

where  $\mu_i$  is the number of relevant epochs experienced and  $m_i$  is the number of miles driven. Substituting equation 3 in equation 2:

$$\log \mu_i = X_i\beta + \mathcal{Z}u_i + \log m_i \quad (4)$$

This is a mixed effect count regression problem and the outcome variable can be represented by Poisson or Negative Binomial distributions. Since high threshold counts can be over dispersed, the negative binomial family was better suited and hence chosen for this analysis. Equation 5 symbolically represents the model for **lme4** package based **glmer.nb** function in R (Bates et al. 2014).

$$\log(\text{number of acceleration events}) \sim \text{roadway speed category} + \text{vehicle class} + \text{location} + \text{gender} + \text{age group} + (1|\text{driver vehicle id}) + \text{offset}(\log(\text{distance traveled})) \quad (5)$$

To facilitate incident rate ratio comparison, the model in equation 2 can also be represented as:

$$\lambda_i = e^{X_i\beta + Z_i u_i} \quad (6)$$

To carry the simpler example in equation 1 forward, the two roadway speed subcategories can be compared using the above form of the predicted rate equation. To study incident rate ratios in different subcategories of a fixed effect, we can assume identical random effects. Also, since all fixed effects are categorical, the matrix of predictor variables is made up of 1's and 0's further simplifying the equation as:

$$\begin{aligned} RR_{\substack{65-80 \text{ mph} \\ \leq 30 \text{ mph}}} &= \frac{\lambda_{65-80 \text{ mph}}}{\lambda_{\leq 30 \text{ mph}}} \\ &= e^{\beta(X_{65-80 \text{ mph}} - X_{\leq 30 \text{ mph}})} \\ &= e^{(\beta_{65-80 \text{ mph}})} \end{aligned} \quad (7)$$

There is a baseline level for each fixed effect that corresponds to 0, in this case  $X_{\leq 30 \text{ mph}}$ , and an indicator variable that corresponds to 1. The rate ratio only needs  $\beta_{65-80 \text{ mph}}$  as  $\beta_{\leq 30 \text{ mph}}$  is already included in the intercept being the first subcategory of the fixed effect. To calculate the 95% confidence interval, we can calculate the upper confidence limit (UCL) and lower confidence limit (LCL) as

$$RR_{\substack{65-80 \text{ mph} \\ \leq 30 \text{ mph}}} \text{ UCL} = e^{(\beta_{65-80 \text{ mph}} + 1.96 \times S.E\beta_{65-80 \text{ mph}})} \quad (8)$$

$$RR_{\substack{65-80 \text{ mph} \\ \leq 30 \text{ mph}}} \text{ LCL} = e^{(\beta_{65-80 \text{ mph}} - 1.96 \times S.E\beta_{65-80 \text{ mph}})} \quad (9)$$

where  $S.E\beta_{65-80 \text{ mph}}$  is the standard error associated with the  $\beta_{65-80 \text{ mph}}$  estimate. All other incident rate ratio comparisons discussed in Section 5 are calculated using similar methodology. For each of combination of the four acceleration types and six thresholds, the data was fitted with a separate model making a total of 24 models used in the results section. A good fit was ensured by trying multiple optimizers and comparing results to ensure similar convergence.

## 5 Results

The purpose of this analysis is to understand the effect of roadway speed category, driver age, vehicle class, location, and driver gender on the rates of common as well as rare acceleration epochs. To achieve this, the rate of the four acceleration types at thresholds of “ $\geq 0.1g$ ” to “ $\geq 0.6g$ ” was estimated using generalized linear mixed effect models with the the above mentioned factors as fixed effects and the driver-vehicle identifier as a random effect. Then the regression coefficients of the models were used to compare the subcategories of each fixed effect. Significant influence of roadway speed category and driver age is observed with consistent trends across all acceleration types and thresholds. Certain vehicle types and location also showed consistently different behavior. Driver gender had the smallest influence among all the fixed effects. These results are discussed in greater detail in the following subsections.

Figures 2 to 6 have 24 subplots each with every subplot derived from a separately fitted generalized mixed effect model. For each plot, the y-axis represents the incidence rate ratio along with the 95% confidence interval shown as a circle with error bars and the x-axis represents the subcategories of the fixed effect being examined. For each subplot, the first subcategory on the left

is the basis of comparison and hence has a rate ratio of 1. These plots are visualizing the relative frequency among the various subcategories and not the actual frequency of epochs. Therefore, Figure 1 should be also be used in conjunction for contextualizing the inferences. The underlying data produced by the analysis is also available for download with this paper. Subsections 5.1 to 5.5 discuss the important results.

## 5.1 Effect of Roadway Speed Category

Roadway speed category shows the most significant impact on the rates of acceleration epochs for all acceleration types and thresholds. Figure 2 shows the effect of roadway speed category on rate ratios with respect to the “ $\leq 30$  mph” speed category. The main takeaways are:

- For norms as well as extreme acceleration thresholds, as the roadway speed increases, the frequency of acceleration epochs decreases. For example, drivers on “ $\leq 30$  mph” roads experience about 100 times more  $\geq 0.2$  g deceleration than on “65 – 80 mph” roadways. The much lower rates for high speed roadways can be explained in two ways. First, the driving is much more stable with speed or direction changing less frequently and the mixing of traffic traveling at different speeds is rare. Secondly, these rates are mileage based, and higher speed roadways accumulate more mileage in a lesser time.
- For higher thresholds, the drop in the relative rate is sharpest in longitudinal accelerations followed by lateral left and right accelerations. Deceleration epochs have a slightly different trend with higher speed roads also producing relatively higher rates than other acceleration types. This is due to the vehicle’s ability to produce higher deceleration than acceleration and the more frequent need to do so because of traffic lights and slowdowns.
- For all roadway speed comparisons, the 95% confidence intervals are relatively small when compared to other fixed effects. This indicates a strong characteristic influence of each subcategory. The tight confidence intervals are also due to most drivers having driven on all roadway speed categories, and hence a larger sample size as compared to other fixed effect subcategories.

## 5.2 Effect of Driver Age Range

Driver age group has significant effect on the relative rates of all acceleration types, especially at higher thresholds. Figure 3 shows the effect of driver age range on the rate ratios with respect to the “16 - 19” years age range. Some important observations are:

- For longitudinal acceleration at higher thresholds, as the age range increases from “16-19”, the rate ratio keeps decreasing until the “70 - 74” group, after which it starts increasing again. “16-19” year old drivers experience the highest rates as compared to any other group. Other younger and older drivers also experience more frequent high-g longitudinal acceleration epochs when compared with ages 30 to 70. However, the 95% confidence intervals are wider than other acceleration types suggesting more variance within age groups.
- For longitudinal deceleration at higher thresholds, a similar trend is observed. Hard braking is experienced more frequently by younger and older drivers. For example, a  $\geq 0.5$  g deceleration is experienced more than twice as often by teenagers than by “50 - 54” year old drivers.

In case of high-g deceleration rate ratios, there are certain bands with similar behavior within the general age based trends. “16 - 19” age range has the highest rates followed by the “20 -

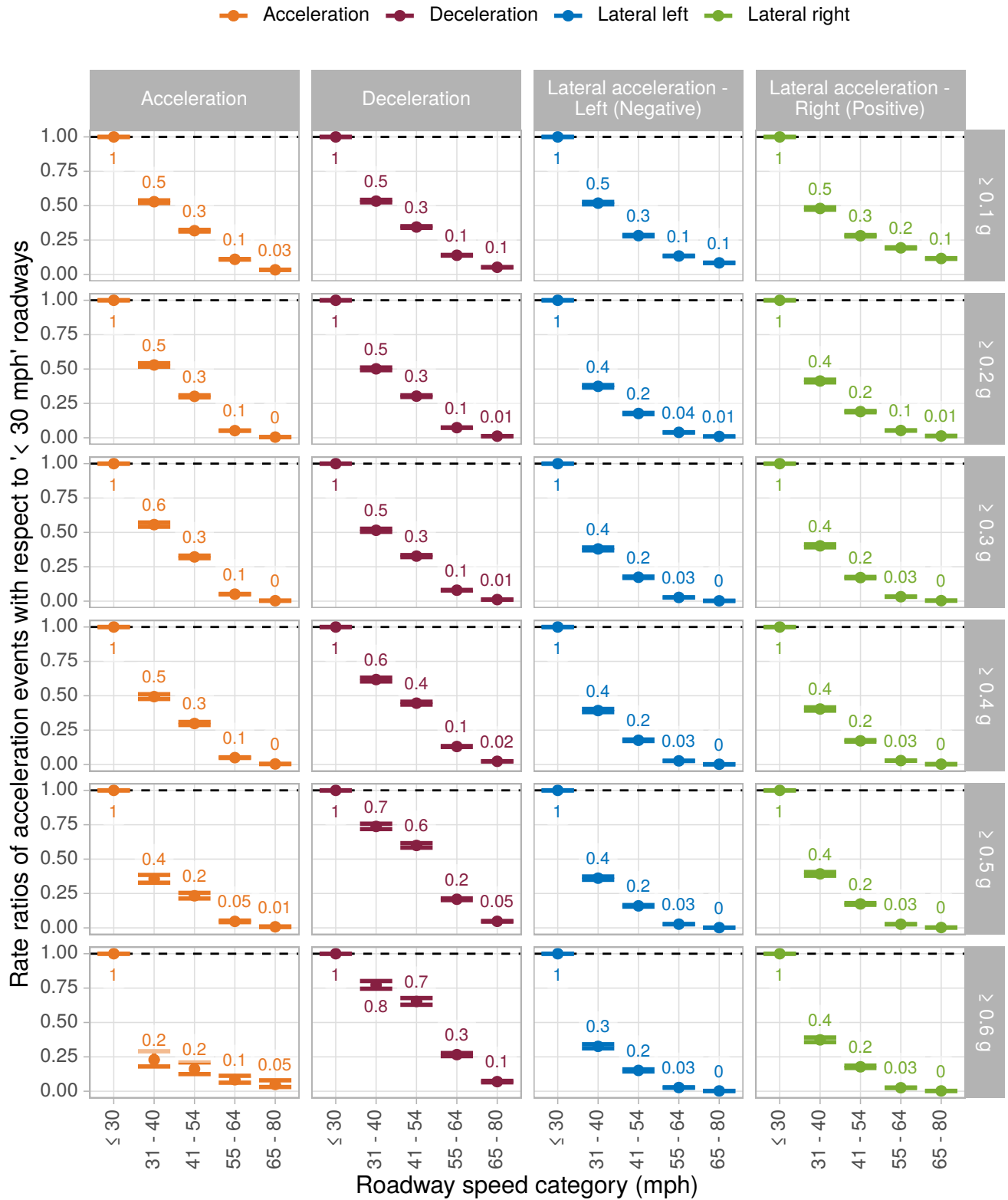


Figure 2: Comparing the effect of roadway speed category on incident rate ratios of the 4 acceleration types at the 6 thresholds with respect to speed category " $\leq 30$  mph".



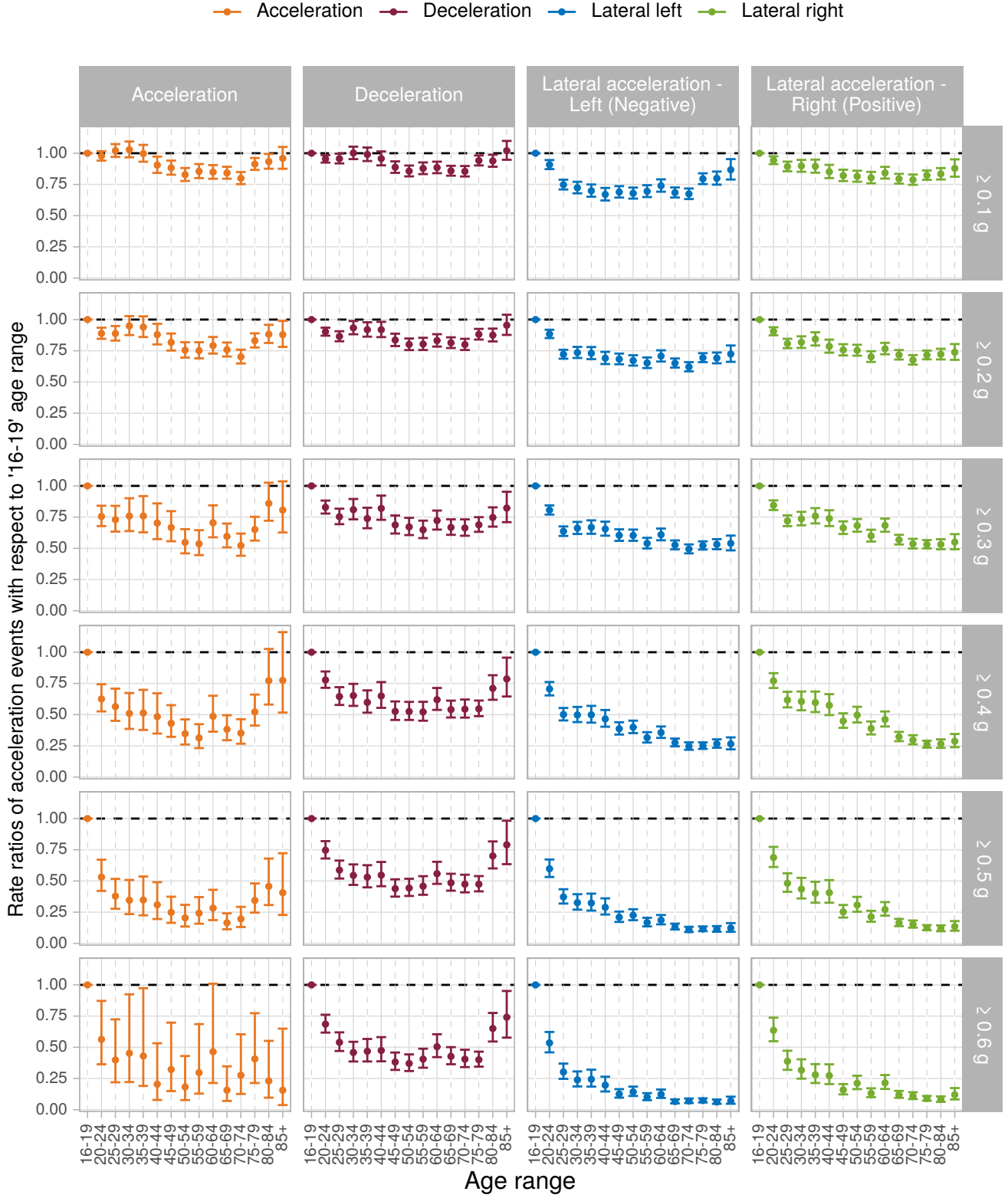


Figure 3: Comparing the effect of driver age on incident rate ratios of the four acceleration types at the six thresholds with respect to age range “16 - 19” years.

24” age group. Then groups within ages 25 to 44 seem to have similar values. Ages 45 to 79 also show similar but slightly lower values than groups within ages 25 to 44. 80 and above drivers show increased rates of hard deceleration epochs. Exploration of the exact mechanism of the age based differences is outside the scope of this paper. However, it should be noted that the mechanism for higher rates in younger and older drivers is probably different. For example, younger drivers could produce higher rates due to inexperience whereas older drivers may produce it due to reduced perceptual capabilities and slower reaction times.

- The rates of high-g lateral acceleration epochs keeps decreasing as the driver age increases. The rate is highest for the teenage group followed by the “20 - 24” driver age group. Drivers between 25 to 44, 45 to 64, and 65 to 85 + form three bands of age groups that have progressively decreasing high-g lateral acceleration rates. For example, teenage drivers experience about 10 times as many  $\geq 0.5$  g left leaning lateral accelerations as “70 - 74” year old drivers. This is a stark difference in driving behavior purely based on the driver age.

The driver age based trend is true for both left as well as right leaning lateral accelerations. This is in contrast to longitudinal acceleration and deceleration rates which showed higher rates for younger and older drivers. The difference in age based trends for high-g longitudinal versus lateral accelerations are due to the differences in the mechanism by which they occur. High-g longitudinal acceleration and deceleration epochs occur due to rapidly changing vehicle speed. However, high-g lateral acceleration epochs occur during rapidly changing vehicle direction such as turns at high speeds, tightly curved roads, and swerving maneuvers. Therefore, if certain groups are predisposed to driving at slower speeds, they may naturally experience lower rates of high-g lateral acceleration epochs.

- For longitudinal acceleration and deceleration rates at lower thresholds of  $\geq 0.1$  g and  $\geq 0.2$  g, the differences are significant but less severe. Drivers from 16 to 44 years show similar rates whereas drivers from ages 45 to 74 show slightly lower rates.
- For lateral acceleration rates at lower thresholds of  $\geq 0.1$  g and  $\geq 0.2$  g, the differences are significant but less severe. Also, the trends are slightly different than longitudinal accelerations. Teenage drivers have the highest rates followed by the “20 - 24” age group. Drivers between the ages of 25 to 75 have similar rates which are slightly lower than the two youngest groups.

Exploring age effects on the rate ratios of high g-force acceleration rates could provide insight about change of driving expertise with age. This could be especially useful for designing age based training and certification programs for drivers.

### 5.3 Effect of Vehicle Class

Vehicle class has significant effect on the relative rates of all acceleration types at higher thresholds. Figure 4 shows the effect of vehicle class on the rate ratios with respect to the “Car” vehicle class. Some important observations are:

- At the lower thresholds of  $\geq 0.1g$  and  $\geq 0.2g$ , all vehicle classes have similar rate ratios with minor differences.
- For moderate to strong acceleration thresholds, minivans have consistently lower rates of acceleration, deceleration, and both types of lateral accelerations. For example, on average cars experience 2.5 times the  $\geq 0.5$  g lateral accelerations that minivans do.

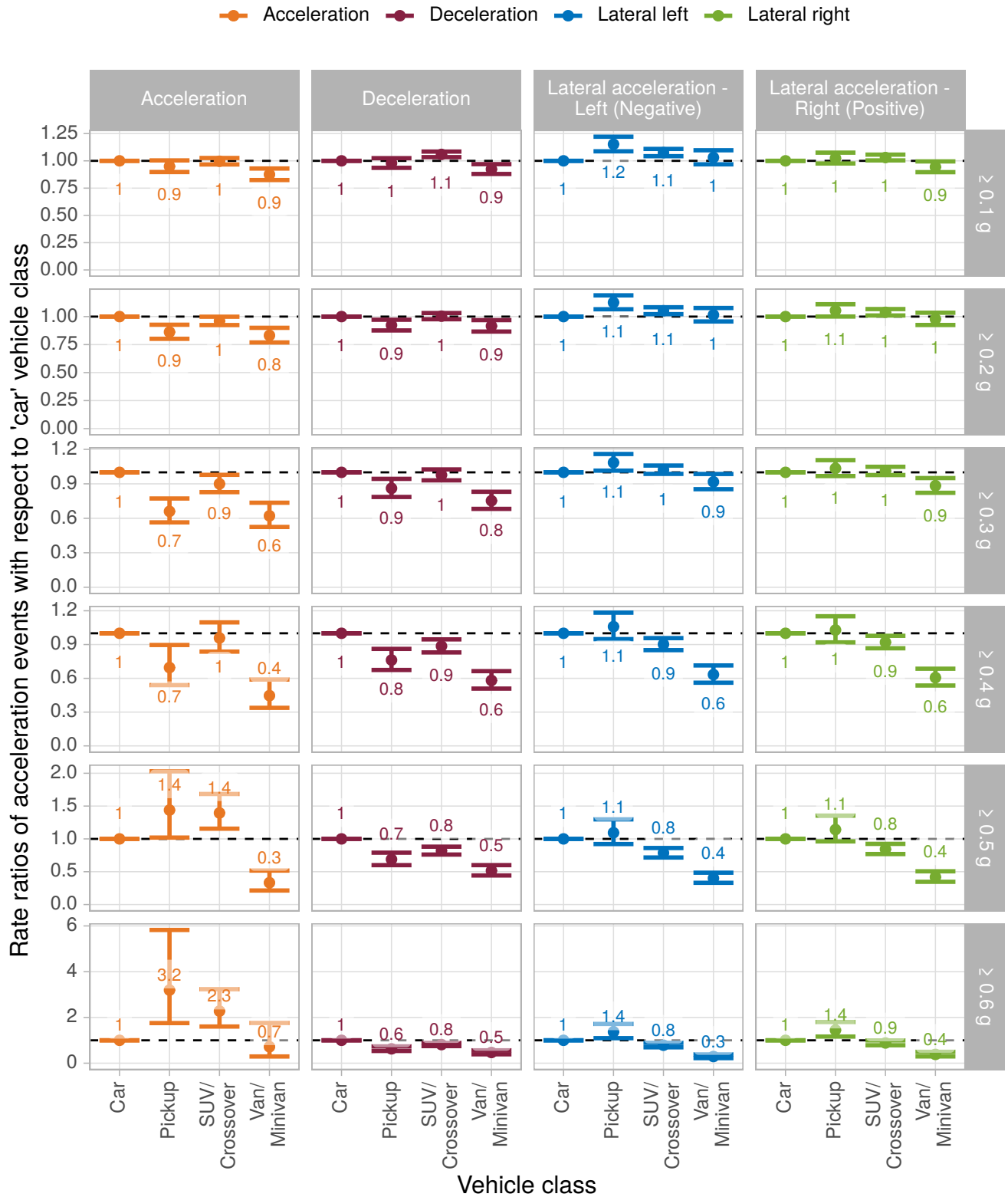


Figure 4: Comparing the effect of vehicle class on incident rate ratios of the 4 acceleration types at the six thresholds with respect to vehicle class “Car”.

In general, vehicle class does not have as big an effect as the roadway category or the driver age. However, many comparisons show significant differences in various vehicle classes. Since there are many factors that distinguish vehicles, such as body or engine specifications, some of the acceleration rate trends may be unclear due to the high dimensional space being represented by just four subcategories. Also, since some type of drivers may be more likely to drive a particular vehicle classes, there may be confounding factors that are outside the scope of this study.

## 5.4 Effect of Collection Site Location

Collection site location has significant effect on the relative rates for some comparisons. Figure 5 shows the effect of collection site on the rate ratios with respect to the “Tampa, FL” site. The collection site location can be seen to represent quite a few building blocks that constitute driving complexity in and around a location. This is in addition to the roadway speed category which is already a fixed effect in this analysis. For example, the layout of the location, whether it has curvy roads or perpendicular grids, proportion of rural versus urban areas, and average traffic conditions. Therefore, even when there are significant differences between locations, the mechanism causing these differences would be better understood by breaking out the driving into such constituent elements. However, as that is outside the scope of this analysis, only some significant observations are reported below:

- Bloomington, Indiana and State College, Pennsylvania have consistently lower rates of longitudinal acceleration and deceleration when compared to the other locations.
- Raleigh, North Carolina has the highest lateral acceleration rates for mild to moderate thresholds.

## 5.5 Effect of Driver Gender

For most comparisons, driver gender does not have significant effect on the relative acceleration rates. However, for some comparisons, there is a difference between male and female drivers. Figure 6 shows the effect of driver gender on the rate ratios with respect to female drivers. Some important observations are:

- For low thresholds, which represent most of driving, the rates are very similar across driver genders with slightly lower deceleration rates for male drivers.
- For higher thresholds, male drivers experienced higher longitudinal acceleration rates, equal deceleration rates, and slightly higher lateral acceleration rates.

## 5.6 Comparing the Influence Various Fixed Effects

Figure 7 compares the influence of various fixed effects on the rate of different acceleration types with a threshold of  $\geq 0.5$  g. The x-axis represents log of the ratio between the maximum rate and the minimum rate within the subcategories of each fixed effect shown on the y-axis. For example, in case of decelerations  $\geq 0.5$  g, roadway speed category has the largest range with the rate experienced on  $\leq 30$  mph roadways being 21 times the rate experienced on 60 – 80 mph roadways. Similarly, the ratio of maximum to minimum rate within each category is calculated for all acceleration types. It should be noted that the subcategories with maximum or minimum rates could be different across acceleration types, as is shown in the figure.

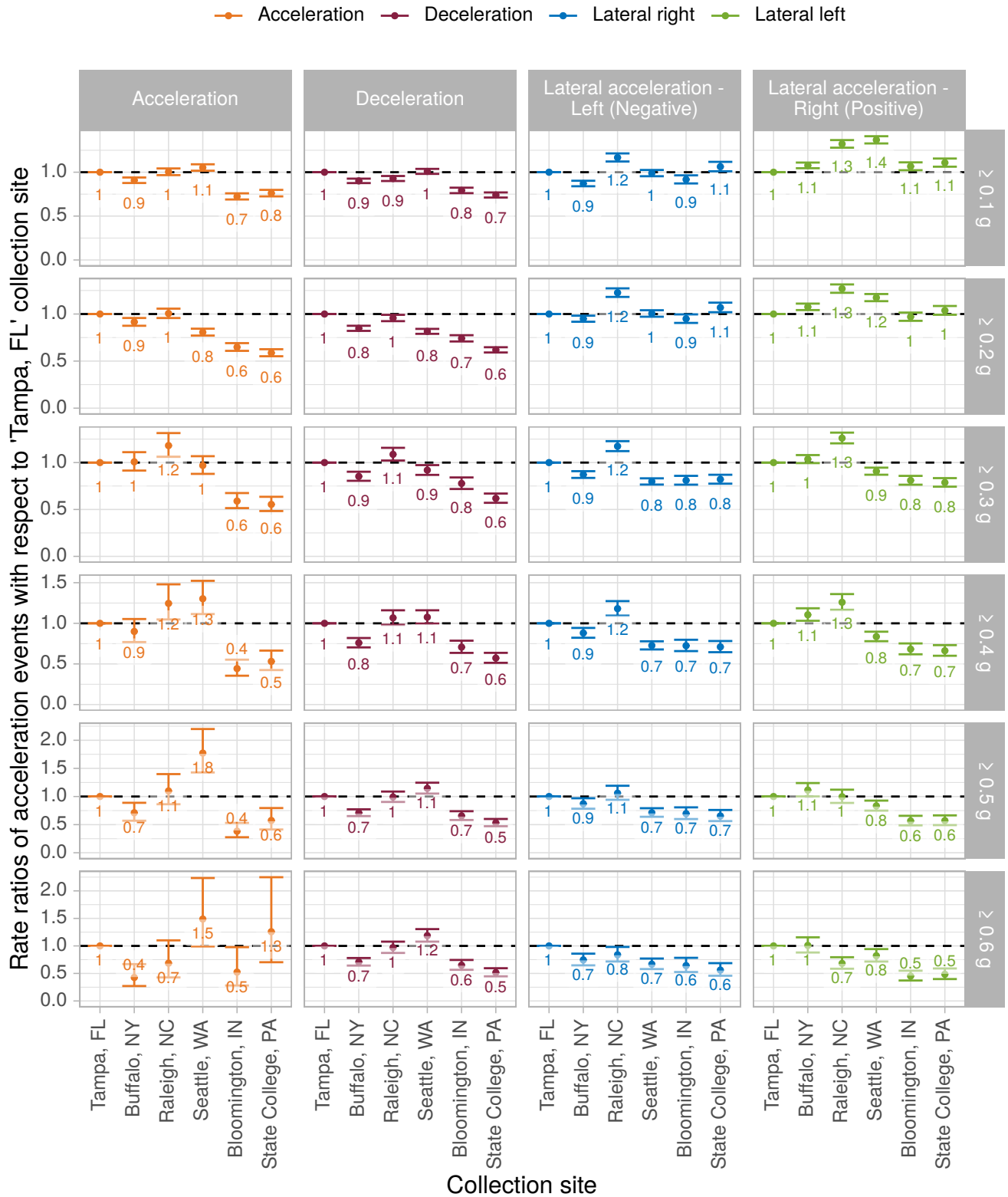


Figure 5: Comparing the effect of location on incident rate ratios of the four acceleration types at the six thresholds with respect to location “Tampa, FL”.

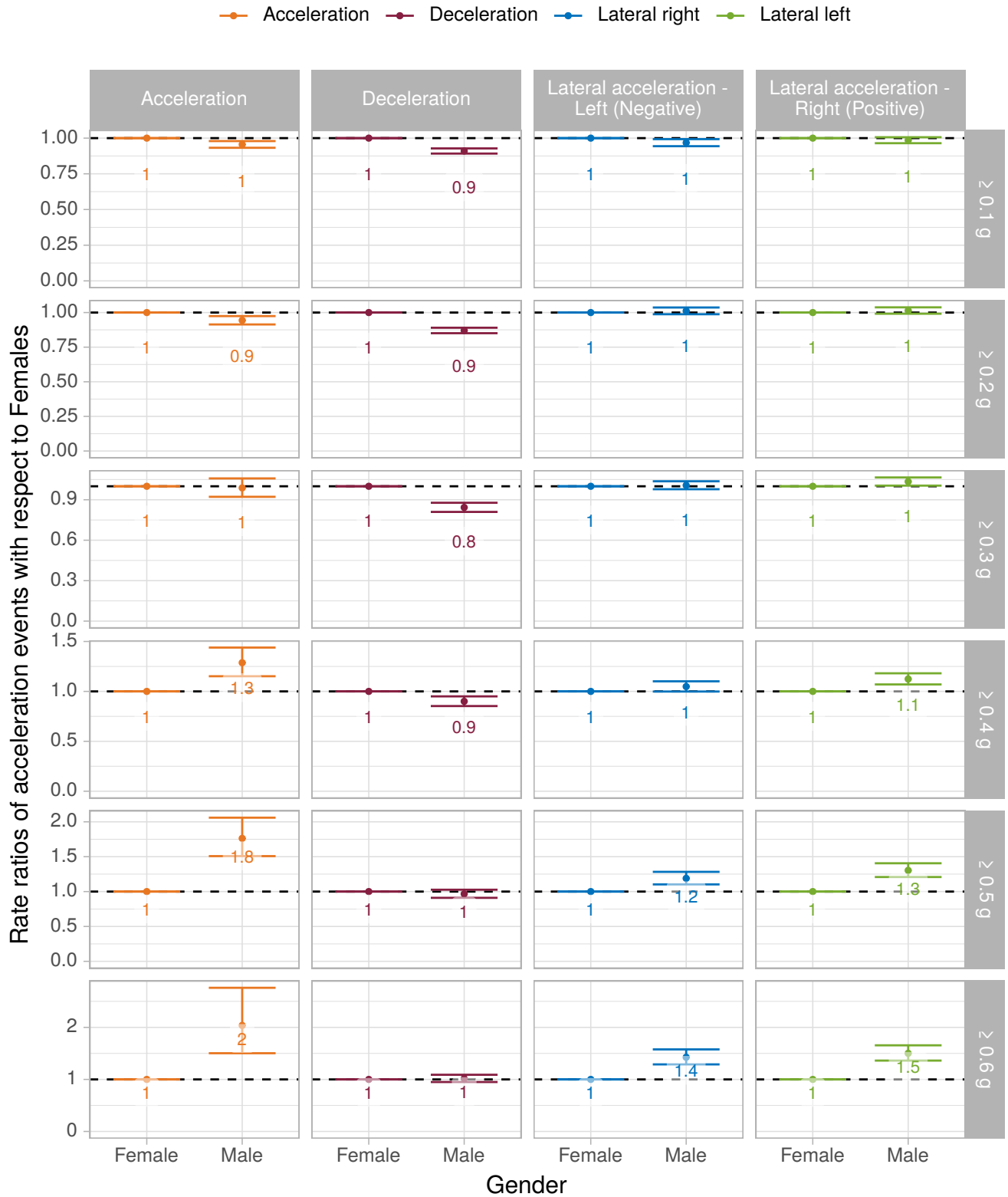


Figure 6: Comparing the effect of driver gender on incident rate ratios of the four acceleration types at the six thresholds with respect to driver gender “Female”.

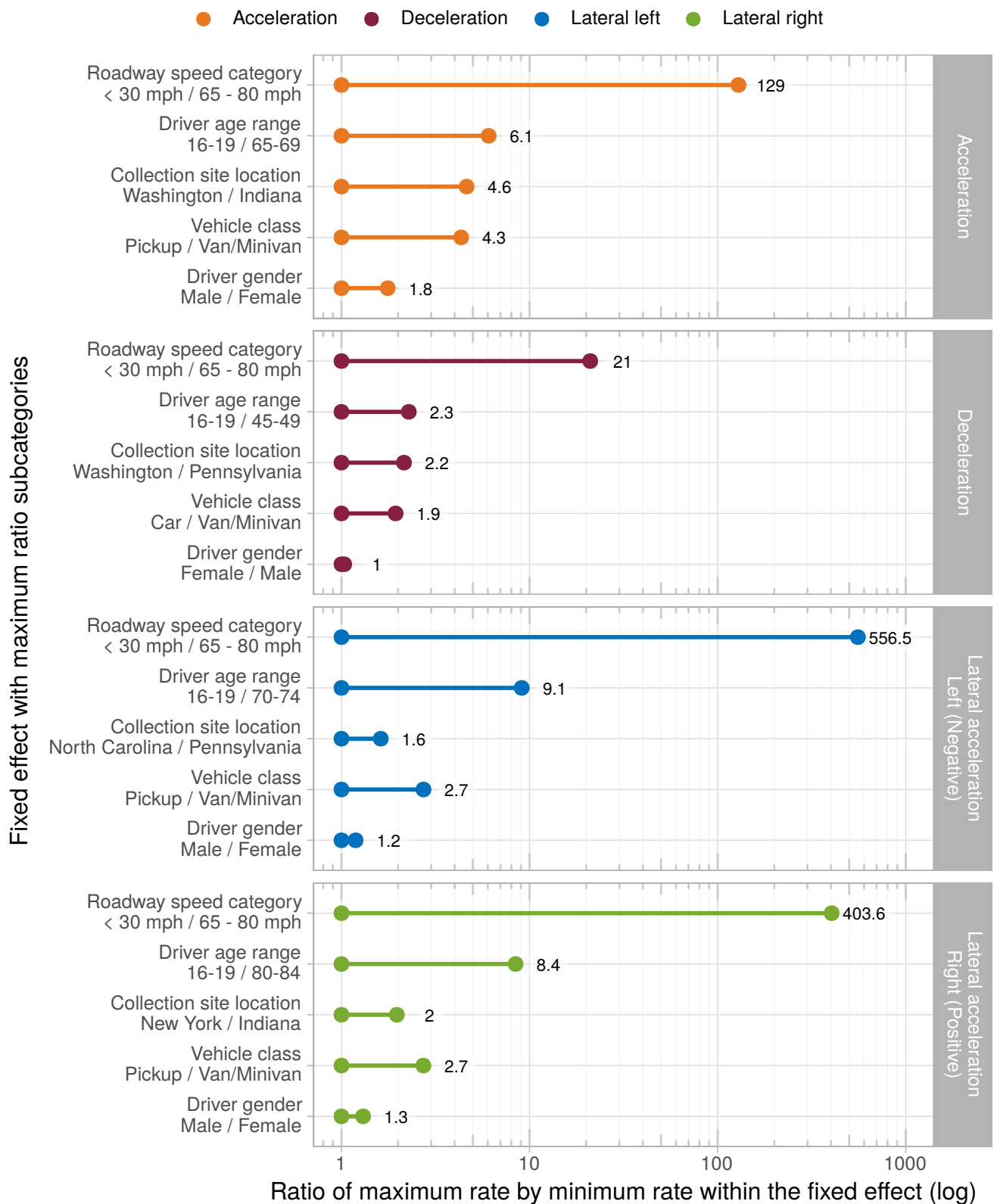


Figure 7: Comparing the ratio of maximum to minimum rate within each fixed effect for a threshold of 0.5 g.

The primary purpose of this figure is to illustrate the comparative magnitude of the effect size across various fixed effects. Even though this figure compares rates for a threshold of 0.5g, similar comparisons can be made at other thresholds using the supplementary data provided with this paper. There is consensus that driving within the acceleration values of  $\pm 0.05g$  is considered steady state driving (Ozaki 1993; Ali et al. 2019). However, various studies have used different thresholds for classifying harsh acceleration events (Kamla, Parry, and Dawson 2019). A threshold of around 0.5g has often been used in transportation studies to classify harsh braking and lateral accelerations events. For example, the following studies used a threshold between 0.4 to 0.6 g to classify some form of a harsh acceleration event. (Dingus et al. 2006; Klauer et al. 2006; Simons-Morton et al. 2009; Fitch et al. 2009; Alessandretti et al. 2009; Benmimoun et al. 2011; Simons-Morton et al. 2012; Geotab 2015; Haque et al. 2016). Therefore, Figure 7 was created for this particular threshold.

For each of the four acceleration types, roadway speed category has the most significant effect with the same subcategories having the highest and lowest rates. The ratio is 129 for longitudinal acceleration, 21 for longitudinal deceleration, 557 for left leaning lateral acceleration, and 403 for right leaning lateral acceleration. Age range is consistently the second most important factor determining the rate of all four acceleration types  $\geq 0.5$  g. Collection site location and vehicle class are the next two important factors but their order differs between the longitudinal and lateral acceleration types. Finally, driver gender has the smallest effect across all four acceleration types ranging between 1 and 1.8.

This study advances the understanding of acceleration behavior in three new ways. First, it simultaneously accounts for important factors such as roadway properties, driver characteristics, and vehicle classification. This ensures that the resulting differences are appropriately assigned to various factors and are not due to coincidences of data collection. Second, it simultaneously analyzes all four acceleration types. This ensures that the same methodology is applied to longitudinal as well as lateral accelerations and therefore the differences can be meaningfully compared. Finally, it analyzes all acceleration behavior at multiple thresholds ranging from  $\geq 0.1$  g to  $\geq 0.7$  g. This ensures that the behavior is investigated for mild, medium, and harsh events. Having all three regions enables researchers to examine traffic behavior, driving style, and safety critical events.

## 6 Conclusions

Acceleration epochs represent a change of state in driving. Be it the change of speed during longitudinal accelerations or the change of direction in case of lateral accelerations. The study of accelerations, their magnitudes, and their frequencies can reveal important insights into factors affecting driving. A comprehensive examination based on a large scale study and analyzing multiple factors has been missing from the literature. This analysis has quantified the simultaneous effects of various roadway, driver, and vehicle factors on the rate of acceleration epochs at various thresholds. It is based on SHRP2 NDS and the surface accelerations reference which are large scale, highly diverse, and context rich datasets. Generalized mixed effect models were used to predict the rates of acceleration epochs at various thresholds with the above mentioned factors as fixed effects and the vehicle-driver combination as a random effect. The regression coefficients from these models were then used to calculate incident rate ratios between subcategories of each fixed effect.

The analysis showed that roadway speed category had the strongest effect on rate of occurrence for mild as well as strong accelerations, with the maximum difference being as large as two orders



of magnitude between 65 - 80 mph and  $\leq 30$  mph roads. Driver age also plays a major role in how often drivers experience high-g acceleration, deceleration, and lateral acceleration epochs. For some comparisons such as strong lateral accelerations, teenage drivers could experience the same type of epoch 10 times as often as older drivers in their 70's. Vehicle class also shows significant effect with minivans providing consistently lower rates than other vehicle types. The collection site location showed significant effect for some comparisons but to better understand the underlying mechanism, the driving would need to be broken out by its constituent elements. Driver gender had the smallest effect with most comparisons showing nearly identical rates.

This study will be beneficial to a number of fields. Vehicle system engineers focused on ADAS and AD systems will be able to use the results presented in Figures 1 - 7 to inform the design of their systems to better represent or respond to real world driving. Safety researchers and transportation regulators will be able to use these results to develop targeted programs for at risk drivers using anomalous driver detection techniques. Roadway designers could use these data to recognize accelerations that are atypical for a given category of roadway. Insurance companies and vehicle cohort operators will also be able to use these results to compare their populations with a representative sample and draw conclusions about driver behavior and driving risk.

This study opens up several possibilities for future work. To better understand the role of vehicle characteristics in acceleration behavior, the vehicle classification could be broken into several constituent factors such as weight, power, and length. Similarly, to better understand the role of location, the current categories could be expanded to include better indicators of driving complexity such as road curvature, average traffic density, intersection density, and road width. Including such factors in the model will help researchers further understand what causes certain acceleration behaviors.

## 7 Acknowledgments

The authors would like to acknowledge Minh Le and Quyen Do from the Statistical Applications and Innovations Group at Virginia Tech for their guidance in selecting the appropriate statistical methods and programming tools for making this analysis possible.

All figures in this paper were created using the ggplot2 package in R (Wickham 2016). The dataset and scripts needed to recreate these figures are available at the GitHub repository accompanying this paper: <https://github.com/gibran-ali/quantifying-accelerations>.

## References

- Alessandretti, Giancarlo et al. (2009). “EuroFOT: European Large-Scale Field Operational Test on Active Safety Systems”. In: 16th ITS World Congress and Exhibition on Intelligent Transport Systems and Services ITS America ERTICO ITS Japan.
- Ali, Gibran, Shane McLaughlin, and Ahmadian Mehdi (2021). “The surface accelerations reference — A large-scale, interactive catalog of passenger vehicle accelerations”. In: *Manuscript in preparation*.
- Ali, Yasir et al. (Sept. 1, 2019). “A game theory-based approach for modelling mandatory lane-changing behaviour in a connected environment”. In: *Transportation Research Part C: Emerging Technologies* 106, pp. 220–242.

- Ali, Yasir et al. (Dec. 1, 2020a). “Comparing the usefulness of real-time driving aids in a connected environment during mandatory and discretionary lane-changing manoeuvres”. In: *Transportation Research Part C: Emerging Technologies* 121, p. 102871.
- Ali, Yasir et al. (Sept. 1, 2020b). “The impact of the connected environment on driving behavior and safety: A driving simulator study”. In: *Accident Analysis & Prevention* 144, p. 105643.
- Antin, Jon et al. (Apr. 4, 2011). *Design of the in-vehicle driving behavior and crash risk study: In support of the SHRP 2 naturalistic driving study*. Final report S2-S05-RR-1. Washington, D.C.: Transportation Research Board of The National Academies.
- Antin, Jon et al. (Feb. 28, 2015). *Naturalistic driving study: descriptive comparison of the study sample with national data*. Washington, D.C.: Transportation Research Board.
- Bagdadi, Omar (Jan. 1, 2013). “Assessing safety critical braking events in naturalistic driving studies”. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 16, pp. 117–126.
- Bagdadi, Omar and András Várhelyi (July 1, 2011). “Jerky driving—An indicator of accident proneness?” In: *Accident Analysis & Prevention* 43.4, pp. 1359–1363.
- Bates, Douglas et al. (2014). “Fitting linear mixed-effects models using lme4”. In: *arXiv preprint arXiv:1406.5823*.
- Benmimoun, Mohamed et al. (2011). “Incident detection based on vehicle CAN-data within the large scale field operational test “euroFOT””. In: *22nd enhanced safety of vehicles conference (ESV 2011), washington, DC/USA*.
- Blincoe, Lawrence et al. (May 2015). *The economic and societal impact of motor vehicle crashes, 2010 (revised)*. NHTSA technical report DOT HS 812 013. Washington, DC: National Highway Traffic Safety Administration.
- Broughton, Jeremy (July 1, 2008). “Car driver casualty rates in Great Britain by type of car”. In: *Accident Analysis & Prevention* 40.4, pp. 1543–1552.
- Choi, Jaisung et al. (Nov. 1, 2017). “Turning movements, vehicle offsets and ageing drivers driving behaviour at channelized and unchannelized intersections”. In: *Accident Analysis & Prevention* 108, pp. 227–233.
- Curry, Allison E. et al. (July 1, 2015). “Young driver crash rates by licensing age, driving experience, and license phase”. In: *Accident Analysis & Prevention* 80, pp. 243–250.
- Dingus, T. A. et al. (Apr. 2006). *The 100-Car naturalistic driving study, phase II - results of the 100-Car field experiment*. Interim report DOT HS 810 593. Washington, D.C.: National Highway Traffic Safety Administration.
- Dingus, Thomas A. et al. (Mar. 22, 2014). *Naturalistic driving study: Technical coordination and quality control*. Washington, D.C.: Transportation Research Board.
- Dingus, Thomas A. et al. (Mar. 8, 2016). “Driver crash risk factors and prevalence evaluation using naturalistic driving data”. In: *Proceedings of the National Academy of Sciences of the United States of America* 113.10. tex.pmcid: PMC4790996, pp. 2636–2641.
- Elvik, Rune (July 1, 2006). “Laws of accident causation”. In: *Accident Analysis & Prevention* 38.4, pp. 742–747.
- Fitch, GM et al. (2009). “Analysis of lane-change crashes and near-crashes”. In: *US Department of Transportation, National Highway Traffic Safety Administration*.
- Geotab, Inc (2015). “Geotab management by measurement”. In:
- Guo, Feng and Youjia Fang (Dec. 1, 2013). “Individual driver risk assessment using naturalistic driving data”. In: *Accident Analysis & Prevention*. Emerging Research Methods and Their Application to Road Safety 61, pp. 3–9.

- Guo, Feng et al. (Feb. 1, 2017). “The effects of age on crash risk associated with driver distraction”. In: *International Journal of Epidemiology* 46.1, pp. 258–265.
- Hankey, Jonathan M, Miguel A Perez, and Julie A McClafferty (2016). *Description of the SHRP 2 naturalistic database and the crash, near-crash, and baseline data sets*. Virginia Tech Transportation Institute.
- Haque, M. Mazharul et al. (Jan. 2016). “Gap Acceptance Behavior of Mobile Phone-Distracted Drivers at Roundabouts”. In: *Transportation Research Record: Journal of the Transportation Research Board* 2602.1, pp. 43–51.
- Here Technologies (Mar. 2, 2019). *SpeedCategoryType - geocoder API - HERE Developer*. SpeedCategoryType - geocoder API - HERE Developer. URL: <https://developer.here.com/documentation/geocoder/topics/resource-type-speed-category.html> (visited on 03/02/2019).
- Høye, Alena (Feb. 1, 2019). “Vehicle registration year, age, and weight – Untangling the effects on crash risk”. In: *Accident Analysis & Prevention* 123, pp. 1–11.
- Kamla, Jwan, Tony Parry, and Andrew Dawson (Jan. 1, 2019). “Analysing truck harsh braking incidents to study roundabout accident risk”. In: *Accident Analysis & Prevention* 122, pp. 365–377.
- Klauer, Sheila G. et al. (Apr. 2006). *The impact of driver inattention on near-crash/crash risk: An analysis using the 100-Car naturalistic driving study data*. Technical report DOT HS 810 594. Washington, D.C.: National Highway Traffic Safety Administration.
- Klauer, Sheila G. et al. (Feb. 2009). “Comparing real-world behaviors of drivers with high versus low rates of crashes and near crashes”. In:
- Kmet, Leanne and Colin Macarthur (Jan. 1, 2006). “Urban–rural differences in motor vehicle crash fatality and hospitalization rates among children and youth”. In: *Accident Analysis & Prevention* 38.1, pp. 122–127.
- Lord, Dominique and Fred Mannering (June 1, 2010). “The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives”. In: *Transportation Research Part A: Policy and Practice* 44.5, pp. 291–305.
- Lyu, Nengchao et al. (Dec. 1, 2018). “The effect of gender, occupation and experience on behavior while driving on a freeway deceleration lane based on field operational test data”. In: *Accident Analysis & Prevention* 121, pp. 82–93.
- Massie, Dawn L., Kenneth L. Campbell, and Allan F. Williams (Feb. 1, 1995). “Traffic Accident involvement rates by driver age and gender”. In: *Accident Analysis & Prevention* 27.1, pp. 73–87.
- McCartt, Anne T. et al. (June 4, 2009). “Effects of age and experience on young driver crashes: Review of recent literature”. In: *Traffic Injury Prevention* 10.3, pp. 209–219.
- McLaughlin, Shane B. and Jonathan M. Hankey (Mar. 3, 2015). *Matching GPS records to digital map data: Algorithm overview and application*. 15-UT-033. National Surface Transportation Safety Center for Excellence.
- National Highway Traffic Safety Administration. *Traffic safety facts annual report tables*. URL: <https://cdan.nhtsa.gov/tsftables/tsfar.htm> (visited on 12/25/2019).
- Ouimet, Marie Claude et al. (June 1, 2014). “Higher crash and near-crash rates in teenaged drivers with lower cortisol response: An 18-Month longitudinal, naturalistic study”. In: *JAMA Pediatrics* 168.6, pp. 517–522.
- Ozaki, H (1993). “Reaction and anticipation in the car-following behavior”. In: *Proc. of 12th international symposium on theory of traffic flow and transportation*, pp. 349–366.

- Paefgen, Johannes, Thorsten Staake, and Elgar Fleisch (Mar. 1, 2014). “Multivariate exposure modeling of accident risk: Insights from Pay-as-you-drive insurance data”. In: *Transportation Research Part A: Policy and Practice* 61, pp. 27–40.
- Park, Eun Sug et al. (Feb. 1, 2021). “Exploration of the relationship among roadway characteristics, operating speed, and crashes for city streets using path analysis”. In: *Accident Analysis & Prevention* 150, p. 105896.
- Petraki, Virginia, Apostolos Ziakopoulos, and George Yannis (Sept. 1, 2020). “Combined impact of road and traffic characteristic on driver behavior using smartphone sensor data”. In: *Accident Analysis & Prevention* 144, p. 105657.
- Ryan, G. Anthony, Matthew Legge, and Diana Rosman (May 1, 1998). “Age related changes in drivers’ crash risk and crash type”. In: *Accident Analysis & Prevention*. Older road users 30.3, pp. 379–387.
- Sharma, Anshuman et al. (Sept. 1, 2020). “Is an informed driver a better decision maker? A grouped random parameters with heterogeneity-in-means approach to investigate the impact of the connected environment on driving behaviour in safety-critical situations”. In: *Analytic Methods in Accident Research* 27, p. 100127.
- Simons-Morton, B. G. et al. (May 15, 2012). “Do elevated gravitational-force events while driving predict crashes and near crashes?” In: *American Journal of Epidemiology* 175.10, pp. 1075–1079.
- Simons-Morton, Bruce G. et al. (June 22, 2009). “Hard braking events among novice teenage drivers by passenger characteristics”. In: *Proceedings of the ... International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design* 2009. tex.pmcid: PMC3019610, pp. 236–242.
- Simons-Morton, Bruce G. et al. (May 1, 2014). “Keep your eyes on the road: Young driver crash risk increases according to duration of distraction”. In: *Journal of Adolescent Health*. Driver distraction: A perennial but preventable public health threat to adolescents 54.5, S61–S67.
- Singh, S. (Mar. 2018). *Critical reasons for crashes investigated in the national motor vehicle crash causation survey*. Traffic safety facts crash stats DOT HS 812 115. Washington, DC: National Highway Traffic Safety Administration.
- Stipancic, Joshua, Luis Miranda-Moreno, and Nicolas Saunier (June 1, 2018). “Vehicle manoeuvres as surrogate safety measures: Extracting data from the gps-enabled smartphones of regular drivers”. In: *Accident Analysis & Prevention* 115, pp. 160–169.
- Tefft, Brian (June 2017). “Rates of motor vehicle crashes, injuries, and deaths in relation to driver age, united states, 2014 – 2015”. In:
- Tselentis, Dimitrios I., George Yannis, and Eleni I. Vlahogianni (Jan. 1, 2017). “Innovative motor insurance schemes: A review of current practices and emerging challenges”. In: *Accident Analysis & Prevention* 98, pp. 139–148.
- Vankov, Daniel, Ronald Schroeter, and Divera Twisk (Mar. 1, 2021). “Understanding the predictors of young drivers’ speeding intention and behaviour in a three-month longitudinal study”. In: *Accident Analysis & Prevention* 151, p. 105859.
- Webb, C. N. (Feb. 2018). *Motor vehicle traffic crashes as a leading cause of death in the united states, 2015*. Traffic safety facts crash stats DOT HS 812 499. Washington, DC: National Highway Traffic Safety Administration.
- Wickham, Hadley (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.
- Williams, Allan F (Jan. 30, 2003). “Teenage drivers: patterns of risk”. In: *Journal of Safety Research* 34.1, pp. 5–15.

- Wu, Kun-Feng and Paul P. Jovanis (Dec. 1, 2013a). “Defining and screening crash surrogate events using naturalistic driving data”. In: *Accident Analysis & Prevention*. Emerging research methods and their application to road safety 61, pp. 10–22.
- Wu, Kun-Feng and Paul P. Jovanis (Jan. 1, 2013b). “Screening naturalistic driving study data for safety-critical events”. In: *Transportation Research Record* 2386.1, pp. 137–146.