NATURALISTIC DRIVING DATABASE DEVELOPMENT AND ANALYSIS OF CRASH AND NEAR-CRASH TRAFFIC EVENTS IN HONOLULU

FINAL REPORT

by
Luana Carneiro Pereira, MSCE
Panos D. Prevedouros, PhD

Department of Civil and Environmental Engineering
University of Hawaii at Manoa

for

Center for Safety Equity in Transportation (CSET)
USDOT Tier 1 University Transportation Center
University of Alaska Fairbanks
ELIF Suite 240, 1764 Tanana Drive
Fairbanks, Alaska 99775-5910

In cooperation with U.S. Department of Transportation,
Research and Innovative Technology Administration (RITA)

rural · isolated · tribal · indigenous

DISCLAIMER

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	**2. Government Accession No.**	**3. Recipient's Catalog No.**
4. Title and Subtitle		**5. Report Date**
NATURALISTIC DRIVING DATABASE DEVELOPMENT AND ANALYSIS OF CRASH AND NEAR-CRASH TRAFFIC EVENTS IN HONOLULU		September 15, 2021
		6. Performing Organization Code
7. Author(s) and Affiliations		**8. Performing Organization Report No.**
Panos D. Prevedouros, PhD and Luana Carneiro Pereira, MSCE Department of Civil and Environmental Engineering University of Hawaii at Manoa		INE/CSET 21.08
9. Performing Organization Name and Address		**10. Work Unit No. (TRAIS)**
Center for Safety Equity in Transportation ELIF Building Room 240, 1760 Tanana Drive Fairbanks, AK 99775-5910		11. Contract or Grant No. Grant # 69A3551747129
12. Sponsoring Organization Name and Address U.S. Department of Transportation Research and Innovative Technology Administration 1200 New Jersey Avenue, SE Washington, DC 20590		**13. Type of Report and Period Covered** July 2019 to August 2021
		14. Sponsoring Agency Code
15. Supplementary Notes		
Report uploaded to:		
16. Abstract		

Dashboard cameras and sensors were installed in 233 taxi vans on Oahu, Hawaii which produced several hours of events classified as naturalistic driving data (NDD) in a period of seven months between fall 2019 and spring 2020. The study achieved its objectives to: (1) collect data from NDD events where driving maneuvers caused an acceleration of 0.5g or higher; (2) develop a database suitable for statistical analysis; (3) derive basic statistics for all variables; (4) investigate correlations between variables; and (5) further investigate correlations (which may represent causality effects) for the most frequent types of events, using stepwise linear regression models. The database included a total of 402 harsh events, of which were 398 near-crashes and four were crashes. Several variables such as road, environmental, driver and vehicle characteristics were coded for each event. The installation of Samsara by the CTL company proved to be a successful tool for coaching drivers, and for providing useful insights into traffic safety factors relating to near-miss events.

17. Key Words		**18. Distribution Statement**	
Naturalistic driving studies, near-crash, rear-end, regression, distraction			
19. Security Classification (of this report)	**20. Security Classification (of this page)**	**21. No. of Pages**	**22. Price**
Unclassified.	Unclassified.	65	N/A
Form DOT F 1700.7 (8-72)	Reproduction of completed page authorized.		

SI* (Modern Metric) Conversion Factors

APPROXIMATE CONVERSIONS TO SI UNITS				
Symbol	**When You Know**	**Multiply By**	**To Find**	**Symbol**
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		NOTE: volumes greater than 1000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS				
Symbol	**When You Know**	**Multiply By**	**To Find**	**Symbol**
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380. (Revised March 2003)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

This study collected naturalistic driving data derived from collaboration between the University of Hawaii of Manoa (UHM) and Charley's Taxi and Limousine (CTL). Dashboard cameras and sensors were installed in 233 taxi vans on Oahu, Hawaii which produced several hours of events classified as naturalistic driving data (NDD) during a period of seven months between fall 2019 and spring 2020. The data collection was halted by the shutdown due to the Covid-19 pandemic. The main goals of this study were to develop a statistical database from the NDD by coding selected near-crash events, and then identify factors that relate to the near-crash/crash events.

Several studies done previously in different parts of the world have shown that through NDD, it is possible to analyze near-crash events to understand crash risk factors. This is important because near-crash events are numerous, whereas crash numbers are low. This study developed a database with a total of 402 harsh events, of which 398 were near-crashes and 4 were crashes. Several variables such as road, environment, driver and vehicle characteristics were coded for each event.

The objectives were to: (1) collect data from NDD events where driving maneuvers caused an acceleration of 0.5g or higher; (2) develop a database suitable for statistical analysis; (3) derive basic statistics for all variables; (4) investigate correlations between variables; and (5) further investigate correlations (which may represent causality effects) for the most frequent types of events, using stepwise linear regression models. The main findings of this study were as follows:

- Nearly 18% of events occurred on Thursdays and only 12% occurred on Sundays.
- Only 17.2% of the events occurred on a curvy road segment, while 82.8% occurred on a straight road segment.
- Approximately 7.5% of the total events occurred on a construction zone or a blocked lane.
- Approximately 31% of the events occurred on a road segment with no traffic control and 16% occurred on a freeway.
- Signal traffic control was present in about 40% of the events.
- Approximately 47% of the events occurred on roads with medium traffic congestion, and 29% with heavy traffic congestion.
- Pavement surface was wet in 10% of the events.
- Lighting conditions were as follows: 13% cloudy, 16% dark, 63% sunlight/clear, 6% glare/sunset/sunrise, and 1% tunnel.
- Pedestrians were present in 24% of the events.
- The most frequent type of event was V1 (the taxi van) almost rear-ended V2 (the vehicle ahead of the taxi van), followed by lane changing related events, and events with pedestrians.
- Near rear-end events on freeways are common. The most influencing factors were light traffic conditions which had a negative effect on near rear-end events, and mobile phone use which had a positive effect on near rear-end events.
- Lane changing events on freeways were strongly affected by mobile phone use and absence of automated driver aids in the vehicle.

A more detailed look on pedestrian, rear end and lane changing types of events using stepwise linear regressions provided additional insights as follows:

- Uninterrupted flow facilities reduce the risk of near rear-end events.
- Wider expressways come with a higher the risk for near rear-end events.
- Roads without parking lower the risk of near rear-end events.

- The risk of near rear-end event is lower on Sundays.
- Light traffic density significantly reduces the risk of rear-end events on freeways.
- Cellphone usage has a positive and significant increase to the risk of highway rear-end events.
- Cellphone usage increases the risk of lane-chancing near-crash events.
- Vehicles without an auto-braking system installed have a higher risk of lane-changing near-crash events on a freeway. Note that auto braking systems also include warnings of occupied adjacent lanes to help the driver avoid erroneous lane changing maneuvers.
- The presence of pedestrian crossings significantly increases the crash and near-crash events with pedestrians.

The installation of Samsara by the CTL company proved to be a successful tool for coaching drivers and the company proceeded with the installation of a different, and more advanced system in 2022. The study was interrupted by the Covid-19 pandemic which reduced the actual sample size to about half of the original target. For future research, it is desirable to study non-professional drivers, as they represent most of the drivers in the real world.

CHAPTER 1. INTRODUCTION

1.1 Background

Improving traffic safety is an important goal. According to data released by the National Highway Traffic Safety Administration (NHTSA) 36,560 people were killed in traffic crashes in 2018, including 1,038 children (14 and younger), 9,378 speeding-related deaths, 4,985 motorcycle fatalities, 6,283 pedestrian deaths, 857 bicyclist deaths and 885 large-truck occupant deaths. In May 2021, NHTSA reported that there were 6,721 pedestrian deaths in 2020, which represents a 4.8% increase over the 6,412 deaths recorded in 2019. The annual death toll in the US was well above 30,000 in the last two decades which is a major concern [1]. Collecting data to understand the major causes of crash incidents is a challenge due to issues with reliable data and not enough cases for statistical analyses. Importantly, most crash analysis approaches predict the crash causes using the data collected after the crash happened. Thus, the data do not describe the exact actions and conditions during the crash [2].

Naturalistic Driving Data (NDD) are collected from sensors installed inside a vehicle. The sensors make possible the observation of a variety of conditions and actions in the seconds before an incident. Conditions and variables such as driver behavior, behavior of other actors before the crash, speed, acceleration and deceleration profiles, location, road characteristics and environment characteristics can be both observed on video and coded into a database, both simultaneously by the onboard unit and via post processing by trained analysts and driver coaches. Data from a detailed NDD system largely eliminate the need for picture and measurement taking, the making of assumptions and for gathering perceptions from eyewitnesses and survivors of the crash, all of which are necessary for the typical post-crash police investigation and accident reconstruction. Furthermore, the number of near-crash incidents is substantially larger than the number of actual crashes, and large data bases with near-crash incidents have been used in various studies for traffic safety analysis, both as the main object of analysis and as a surrogate to crashes [3].

There are several NDD systems for collecting naturalistic driving data such as: Azuga[1], Lytx[2], Mobileye[3], NetraDyne[4], Samsara[5], and Teletrac Navman[6]. Usually, these systems are composed of on-board cameras that record the view in front of a vehicle (and sometimes the interior of the vehicle), GPS tracking, speed sensors and sometimes a facial recognition system. Typically, the data are saved on a cloud-based platform where administrators and driver coaches access information about incidents, typically triggered by a specified change in acceleration.

Each NDD system has specific features. Azuga [4] has a driver's coaching and rewarding system. Lytx [5] presents an integrated advanced machine vision and artificial intelligence that captures and categorizes risky driving behaviors. Mobileye Shield+ [6] monitors buses and truck blind spots, for the purpose of warning for the presence of pedestrians and bicyclists in these areas. It also provides information about the city's dangerous zones and potential hazards along specific routes. Netradyne Driveri [7] measures and alerts speed according with following distance, relative speed to traffic and visual detection of road sign. Samsara [8] provides a safety report system that gives labels to classify the near miss events. Teletrac Navman [9] offers

[1] https://www.azuga.com/
[2] https://www.lytx.com/en-us/
[3] https://www.mobileye.com/us/fleets/
[4] https://www.netradyne.com/
[5] https://www.samsara.com/
[6] https://www.teletracnavman.com/

on board Quad camera that shows a 360-degree display, providing forward-facing, driver-facing and left and right-side views.

For this study, the University of Hawaii at Manoa (UHM) research team partnered with a taxicab company located in Honolulu, Hawaii. In order to improve its fleet safety and reduce insurance costs through driver safety training, Charley's Taxi and Limousine (CTL), Oahu's older taxicab company, acquired a system that collects real time data with a camera system that records the front view and the inside of each vehicle, including its driver. Several systems were reviewed, three of them were field tested and in mid-2019 CTL choose the Samsara system for installation in their nearly 200 taxis and limousines.

Samsara [8] offers a system called Samsara AI Dash Cams that can detect real-time incidents and prevent incidents trough the coaching platform. The system uses artificial intelligence (AI) and the data provided by the accelerometer to issue real-time audible alerts to notify the driver of rough maneuvering that may denote unsafe driving.

1.2 Study Purpose and Objectives

The purpose of this study was to identify the main factors affecting near miss incidents in Honolulu using naturalistic driving data. CTL partnered with UHM and made the Samsara data directly available for analysis, in an effort to improve the state-of-the-art in real world naturalistic driving data collection and analysis, and potentially improve driving guidance and laws in Hawaii and the nation. This project used the database created to identify and quantify effects of human factor characteristics (e.g., driver gender and behavior), environmental conditions (e.g., weather, rural, lighting), road characteristics (e.g., road alignment and surface conditions), traffic operations variables (e.g., speed limit and operating speeds), and safety variables (e.g., severity of event, near-crash, or crash outcome) on recorded near-miss and crash events.

The following tasks were set to achieve the study objectives:

1- Collect data from NDD events where driving maneuvers caused an acceleration of 0.5g or higher,
2- Develop a database suitable for research analysis,
3- Derive basic statistics for all variables,
4- Investigate correlations between variables; and,
5- Further investigate the variables related to the most frequent types of events.

1.3 Report Outline

Figure 1 outlines the main parts of this research. The report is organized into seven chapters. Following the introductory Chapter 1, Chapter 2 contains a detailed literature review on road safety studies and previous naturalistic driving studies. Chapter 3 presents the research methodology, data collection, database description, and data analyses techniques. Chapter 4 describes the basic analysis as variables frequency and correlations. Chapter 5 the linear regression models. Chapter 6 contains the main conclusions, and Chapter 7 presents recommendations for future research.

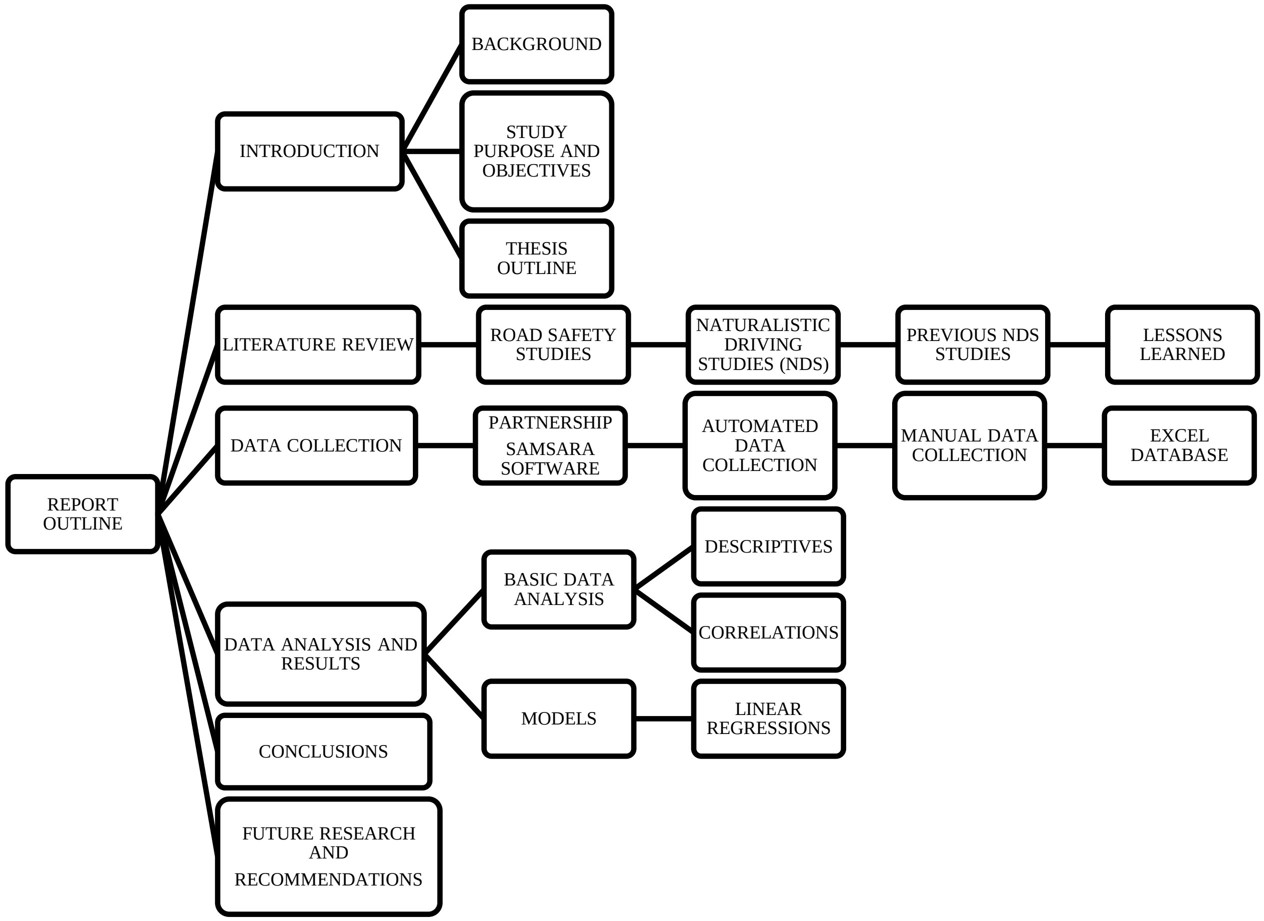

Figure 1.1 Outline of this report

CHAPTER 2. LITERATURE REVIEW

2.1 Road Safety Studies

Several interrelated factors affect road safety. Traffic crashes, especially those involving injuries and fatalities, are the focus of many analyses by enforcement agencies, transportation agencies, and researchers at universities, manufacturers, and safety foundations. Despite all the studies that have been published on traffic safety over the years by authors from diverse areas, there is still a lack of good quality and detailed crash data; typically, the data is collected after the crash. For this reason, measuring safety risks and factors precisely and assigning specific countermeasures to safety risks is not possible [10].

Several advanced statistical models for traffic safety analysis have been developed over the years; persistent issues with small size data samples, questionable causality of post-crash data, and data aggregations (both to protect anonymity and reveal trends) are some of the remaining weaknesses of road safety studies. [10] It takes several years of data collection to have a rich database, e.g., about ten years to collect 1,000 fatality events in Hawaii. A sample size of 1,000 fatal crashes is of modest size because events typically must be analyzed separately by island, type of road, time of day, type of crash, while controlling for environmental and other conditions.

Additional techniques have been developed to obtain more crash-related data and improve the significance of the road safety studies. Traffic conflict observations can be made from a fixed site or from a moving position. In the early studies, the conflicts were analyzed from a fixed location by human observation. The results of these studies showed that in this approach, data subjectivity and inaccuracy were weaknesses. Recent studies use modern techniques such as video, radar and LiDAR sensors from various perspectives such as fixed (e.g., a city's CCTV system) or moving (e.g., in-vehicle driver monitoring system) [10]. The latter gave birth to Naturalistic Driving Studies (NDS.)

Dingus et al. [11] compared the NDS with other driving safety research methods and noted that this approach can fill gaps from previous methods. NDS provide much more detailed observations about crash and pre-crash conditions than existing databases from epidemiological data collection approaches. Furthermore, the naturalistic driving approach provides more natural data than the studies that use empirical data collection, such as test tracks and simulators.

Through this comparison, it was observed that empirical data collection approaches, such as test tracks and simulators, are proactive and provide important crash risk information. However, this approach is imprecise for relying on unproven safety surrogates and for causing driver behavior changes due to the specific experimental conditions, which typically result in no or benign negative outcomes [11]. On the other hand, the large-scale naturalistic data collection approach presents detailed pre-crash and crash information. Long term NDS record the natural driver behavior thus making it possible to observe human factors such as distraction, drowsiness, and driver errors. The in-vehicle sensors also provide vehicle dynamic data [11,4,5,6,7,8,9].

According to Tarko [10], autonomous and other vehicles equipped with advanced driver assistance technology improves traffic safety analysis by providing direct data in a fashion similar to a commercial airliner's black box. The confidence of outcomes obtained from post-crash and conflicts count analyses is affected by weaknesses such as the improper aggregation of data from heterogeneous sites, crash underreporting, ambiguous definitions of crashes and conflicts, assumed but not confirmed causality, etc. NDS eliminates most of the sources of low confidence in the traditional methods [10].

2.2 Naturalistic Driving Studies

NDS is a type of investigation of driver performance and behavior, where the driver, vehicle and the environment are continuously recorded over a long period of time, through video cameras and sensors installed on the vehicles of participants [12]. Klauer et al. [13] stated that the naturalistic driving data approach provides potent tools for the studies that merge empirical data analysis approaches with some characteristics of the epidemiological ones. The combination of the approaches is very beneficial and generates new analytical methods for better driver safety studies, focusing on driver behavior.

NDS was described by Van Schagen et al. [12] as a research method that provides insights in to everyday driver behavior, where equipment such as small cameras and sensors are installed in the vehicles registering vehicle maneuvers, driver behavior, and external conditions. Through the data, it is possible to observe and analyze how driver, vehicle, road, and conflict, crashes and normal traffic situations are interrelated. Backer-Grøndahl et al. [14] defined NDS observations relevant for understanding not only the driver's behavior but also for analyzing crash events.

The observation of the pre-crash scenarios and natural reactions of drivers during traffic conflicts is important for identifying risky driving behaviors. Commercial transportation companies have been using data from on-board cameras and sensors to monitor their drivers and develop feedback and coaching programs [15,16,17]. Hickman et al. [15] evaluated the impact of a commercially available onboard safety monitoring system and its feedback on commercial truck drivers. Bell et al. [16] analyzed if two different types of feedback from in-vehicle monitoring system would decrease risky driver behaviors. Mase et al. [17] investigated the influence of camera monitoring on truck driver risky behaviors. The study also analyzed the effect of camera monitoring and coaching on the driving errors and violations.

The large amount of data provided by NDS can be a disadvantage. Researchers have to observe a large number of drivers for an extensive period of time to select the crashes and near-crashes, this process takes time, and it is costly and laborious. Because a well-designed fleet monitoring system can reduce the crashes by 60%, such systems have been adopted by several passenger transport agencies and the freight industry [18]. There are many systems in the market from US and foreign suppliers, including a low-cost, easy-to-install driver monitoring and assistance system developed as part of Transportation Research Board's (TRB) IDEA program [19].

Fitch et al. [20] summarized reasons why NDS can be an effective tool for the design, testing, and evaluation of driver assistance systems. The first reason is that NDS provides data about driver errors leading to safety-critical events, from which it is possible to identify the driver's needs from a driver assistant system. Second, the observations about driver behavior and performance are useful for working prototype tests. Third, the frequency of the driver's involvement in safety-critical events over the time can be used to evaluate the effectiveness of NDS-based driver assistance and coaching. NDS makes it possible to relate the events to behaviors and other contributing factors. However, in order to achieve meaningful results, the study should present a large data sample, with enough miles driven, with different drivers over a long period of time. [14] The original 100-vehicle NDS conducted by NHTSA and the Virginia DOT, proposed the use of near-crash data for understanding crashes and improving road safety. According to Dingus et al. [11], near-crashes have two important advantages over crashes: Near-crashes happen fifteen times more often than crashes and they give details about successful evasive maneuvers, which provide insights into safer driving techniques.

A report developed at the Virginia Tech Transportation Institute (VTTI) presents the following operational definitions for crash and near-crash events [11]:

Crash – Any contact with an object, either moving or fixed, at any speed in which kinetic energy is measurably transferred or dissipated. Includes other vehicles, roadside barriers, objects on or off the roadway, pedestrians, cyclists, or animals.
Near-crash – Any circumstance that requires a rapid, evasive maneuver by the subject vehicle, or any other vehicle, pedestrian, cyclist, or animal to avoid a crash. A rapid, evasive maneuver is defined as a steering, braking, accelerating, or any combination of control inputs that approach the limits of the vehicle capabilities (e.g., a longitudinal deceleration of at least 0.5 g).

Considering that the amount of naturalistic crash data is usually insufficient, Guo et al. [3] conducted three different analyses to study the use of near crashes as a surrogate measure to crashes. First, the sequential factor analysis concluded that no relevant differences in the number of contributing factors were present. The second analysis showed a positive relationship of the presence of behaviors between crashes and near crashes. Third, the sensitivity analysis indicated that increasing the data sample size by adding near-crash events increased the precision of the results. Guo et al. [3] analyzed the 100-Car NDS data and concluded that the near-crash data used for measuring crashes produces conservative risk estimates; especially in small-scale studies, where the number of crashes is small, the near-crash incidents can provide information to lead the researchers in the right direction.

2.3 Overview of Previous Naturalistic Driving Studies

The 100-Car Naturalistic Driving Study was the first study that used instrumented-vehicles to collect pre-crash data[11]. This study was followed by the Second Strategic Highway Research Program (SHRP2) program, which is the largest scale naturalistic study to date, with over 3,000 participating drivers [21]. These two main studies introduced the benefits of NDS, and then several studies with the same approach were developed. NDS systems with on-board cameras and sensors were installed in different countries for research purposes and for commercial fleet management purposes [15,16,17]. Selected relevant highlights of these studies are summarized below. These studies provided the basis for the development of the NDD database of our project.

2.3.1 The 100-Car Naturalistic Study

The study was conducted by VTTI and sponsored by NHTSA, Virginia Tech, Virginia DOT, and Virginia Transportation Research Council. This study was the first one to collect detailed information on near-crash events [11]. The 100-Car Naturalistic Study addressed ten high priority goals:

Goal 1: Characterization of crashes, near-crashes, and incidents for the 100-Car study.
Goal 2: Quantification of near-crash events.
Goal 3: Characterization of driver inattention.
Goal 4: Driver behavior over time.
Goal 5: Rear-end conflict causal factors and dynamic conditions.
Goal 6: Lane change causal factors and dynamic conditions.
Goal 7: Inattention for rear-end lead-vehicle scenarios.
Goal 8: Characterization of the rear-end scenarios in relation to Heinrich's Triangle.
Goal 9: Evaluation of the performance of hardware, sensors, and the data collection system.
Goal 10: Evaluation of the data reduction plan, triggering methods, and data analysis.

The study accumulated almost 43,000 hours of data, from 241 primary and secondary drivers, with each vehicle having collected data for 12 to 13 months. The events were classified in crashes, near crashes and other "incidents." A total of 69 crashes, 761 near-crashes, and 8,295 incidents were recorded during the study. Data were classified by pre-event maneuver (e.g., changing lanes, decelerating in traffic lane, maneuvering to avoid a vehicle, turning left, turning right), precipitating factor (e.g., lost control, subject vehicle in changing lanes, subject vehicle off the roadway), event type (e.g., single vehicle, lead vehicle, following vehicle, object/obstacle, parked vehicle, animal), contributing factors (e.g., driving environment, infrastructure, secondary task) and the avoidance maneuver exhibited. Parameters such as vehicle speed, vehicle headway, time-to-collision, and driver reaction time were also recorded [11].

After the study was completed, it was reported that from the 82 collisions observed, only 15 were reported to the police. Low severity collisions occurred much more frequently than severe crashes. The NDS data collection system increased the crash data not only by detecting more crash incidents, but also by allowing the analysis of near-crash events which helped to better understand risk factors [11].

2.3.2 SHRP 2 Naturalistic Driving Study

The SHRP 2 NDS was the largest naturalistic driving study conducted to date among NDS with publicly available data. The study lasted for three years, and had the participation of more than 3,500 volunteer drivers of passenger-vehicles, ages 16 to 98, with most of the drivers participating for 12 to 13 months. The main goal of this study was to understand the interaction between drivers, vehicles, traffic environment, roadway characteristics, traffic control devices, and the environment. The study also sought to assess the association between these factors, the interactions and collision risk. The study contributed to the SHRP 2 goal of addressing the role of driver performance and behavior in traffic safety [21].

Figure 2.1 shows the complete Data Acquisition System (DAS) that was installed in the project vehicles (radar unit, cables to front turn signals, radar interface box, OBD connector cables, head unit, rear looking camera, GPS/cell antenna and DAS Main Unit). The system consisted of a head unit that contained four cameras, accelerometers, an illuminance sensor, an infrared illuminator, a passive alcohol sensor, and a GPS sensor; a radar interface box, GPS and cellphone antennae, rear-looking camera, cables and the DAS main unit [22].

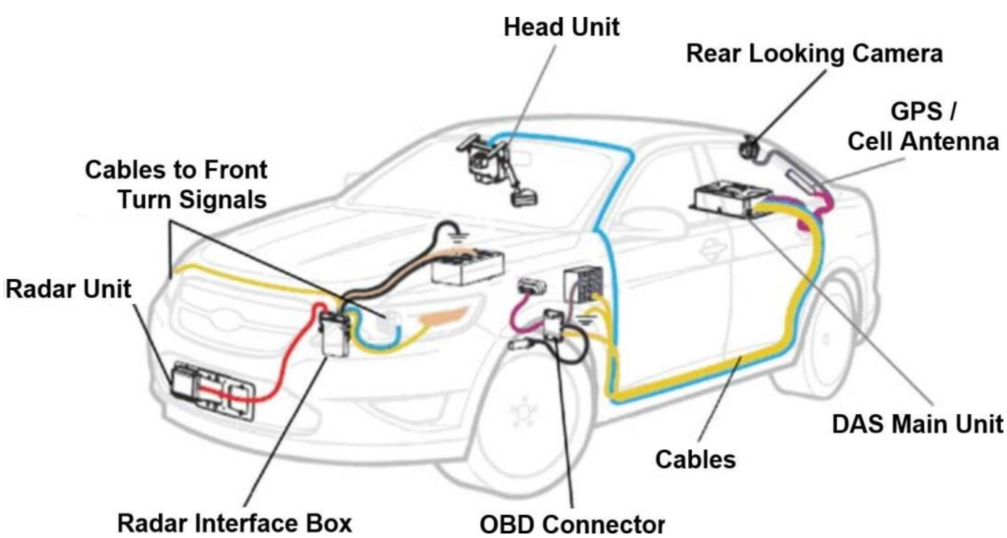

Figure 2.1 Data acquisition system installation SHRP 2 [22]

The study collected approximately 5.4 million trips, 35 million vehicle-miles of naturalistic driving data in six states: Florida, Indiana, New York, North Carolina, Pennsylvania, and Washington. Throughout the study, 2,705 near-crashes and 1,541 crashes were recorded. The database is available for certified researchers on a dedicated website [21].

The VTTI team of data reductionists reviewed the possible crashes and coded up to 75 variables for each identified crash and near-crash. The study classified the data collected in the SHRP 2 NDS project in the following categories [21]:

- Participant assessments: Demographic questionnaire, driving history, driving knowledge, medical conditions and medications, ADHD screening, risk perception, frequency of risky behavior, sensation seeking behavior, sleep habits, visual, physical and cognitive test results, and exit interview.
- Vehicle information: Make, model, year, body, style, vehicle condition, safety and entertainment systems.
- Continuous data: Face, Forward, rear, and instrument panel video, vehicle network data, accelerometer/gyroscopic, forward radar, GPS, additional sensor data.
- Trip summary data: Characterization of trip content, start time and duration of trip, min, max, mean sensor data, time and distance driven at various speeds, headways, vehicle systems usage.
- Event data: crashes, near crashes, baselines, 30 second events with classifications, post-crash interviews, other crash data.
- Cellphone records: Subset of participant drivers, call time and duration, call type (text, call, pic, etc).
- Roadway data: Matching trip GPS to roadway database, roadway classifications, other roadway data.

The SHRP 2 database can be combined with the Roadway Information Database (RID) that provides road characteristic and environmental characteristics, such as time of the day and weather. Driver data about demographics, driving history and some health conditions were also collected through questionnaires and tests [23]. The system allows the researchers to filter the data relevant to their research questions. Several researchers have used the SHRP 2 NDS data in various studies related to traffic safety; some of their findings are presented later in this chapter.

2.3.3 Other Studies

NDS has been adopted by researchers in various countries. In Japan, a fleet of 202 cars and trucks participated of a study to collect traffic accident data from the driver's point of view. Video drive-recorders (VDR) equipped with a video recording unit were installed for a test period, in which 30 crashes were recorded. The study assessed the effectiveness of video driver recording systems in providing details of traffic events [24].

A VDR system was installed in 50 taxis in Beijing, with 48 male and 2 female taxi drivers. The equipment recorded images of the vehicle front view. The study lasted for one year and recorded 51 crash events and 3,010 near-crash events. The researchers used the data to analyze the pre-event maneuvers and for observing the braking operations and driver reaction times[25]. Lin et al. [26] used the same 50 taxi database to study the characteristics of vehicle to pedestrian/bicycle conflicts.

In Iran, a naturalistic driving study [27] was conducted to analyze the occurrence and the severity of vehicle-pedestrian conflicts. A total of 52 drivers participated of the study that had one-year duration. The data was collected from a camera data system installed in each participant's private vehicle.

Smaller NDS were conducted over time to study specific factors or a specific group of drivers. For example, to understand the impact of age and driving experience over traffic conflicts involvement, personal vehicles driven by 42 novice teen drivers in the state of Virginia were instrumented and recorded for 18 months. Based on their involvement in traffic conflicts over the first six months of driving, the participating drivers were grouped into high and low risk categories [28].

Several passenger and freight transportation companies have been using NDS systems to study the behavior of their drivers. As one of the advantages of this approach is the huge amount of data collected, private companies have been partnering with research groups to extract factors and lessons for improving traffic safety [10]. Some studies such as the one funded by the Federal Motor Carrier Safety Administration evaluated a commercially available onboard safety monitoring system. The system composed of two cameras and three accelerometers. It was installed on a truck that was driven by different drivers during their regular deliveries, for seventeen weeks. After the first four weeks, the researchers started giving feedback to the drivers and then observed the impact of coaching strategies on the number of traffic incidents [15].

2.4 Main Findings of Naturalistic Driving Studies

NDS are known for providing a large amount of data. Researchers must observe several drivers for a long period of time to select the crash and near-crash cases that meet their criteria. Relevant NDS applications on demographics and human factors, type of traffic conflicts, and commercial motor vehicles are summarized below.

2.4.1 Demographics and Human Factors

Several studies have analyzed the relationship between demographic factors such as age and gender for near-crashes and crash events [28,29,30]. Human factors such as distraction, inattention and drowsiness were also the subject of various analyses [29,31]. According to Tarko, previous studies showed that young males (aged 16-25) are frequently involved in rear-end or near rear-end crashes [10]. Using the 100-vehicle NDS data, Dingus et al. [11] observed that in almost 80% of all the near-crashes, the driver was looking away from the roadway prior to the conflict. Secondary task distraction, extreme drowsiness, driving inattention, and non-driving related eye glances were observed in 93% of the crashes of the subject vehicles.

The 100-vehicle study also detected a high rate of drowsiness contributing to the conflicts. Drowsiness was a contributing factor in 10% of the near-crashes and 12% of the crashes; these rates are higher than most of regular crash databases which usually contain drowsiness as a contributing factor in only 2-4% of the cases [11]. Klauer et al. [31] analyzed the impact of inattention using the 100-vehicles database. They showed that drowsiness can increase the individual crash/near-crash risk by four to six times; secondary tasks seem to increase the risk two times, while driving related inattention, such as checking the rear-view mirror was shown to be safer than normal. The 100-vehicle NDS database also showed that inattention is strongly related to age. Drivers aged 18 to 20 are four times more involved in inattention-related crashes than older driver groups. Cell phones and other hand-held devices were highly associated with secondary task distraction-related events [11]. Analysis using the naturalistic teenage driving study database observed that teenagers were involved more than five times in high g-force events than adults.

2.4.2 Type of Traffic Conflicts

Several researchers used NDS databases to understand the main factors related to specific types of traffic conflicts, such as rear-end conflicts, lane departure conflicts, vehicle-pedestrian conflicts, etc. Tarko [10] observed that in the SHRP 2 data, only three groups of drivers were involved in rear-end conflicts, young males (16-25) were related to 31 rear-end crashes, mature males (45-64) were related to 10 rear-end crashes, and mature females (45-64) were related to 2 rear-end crashes. In the 100-vehicle NDS database, males are overrepresented in near-crashes, being 120% more likely to be in a rear-end crash than females. Drivers aged 25-34 were 190% more involved in rear-end events than other age groups. Most of the rear-end incidents occurred in clear weather conditions, straight alignment and on dry roads. No significant influence of environmental light on the rear-end events was detected, and 60% of the rear-end crashes occurred at intersections, intersection-related areas, or highway entrance/exit ramp locations [32].

McLaughlin et al. [33] analyzed all run-off-road (ROR) events in the 100-vehicle NDS database. Of the 122 incidents, 94 were near-crashes and 28 were crashes. The ROR events were more likely to happen under poor-visibility and low-friction conditions. Rain, fog and other precipitation increased the likelihood of a ROR event 250% compared with clear conditions. Wet roads increased the likelihood of a ROR event by approximately 180% compared with dry conditions. Snow or ice increased the likelihood by 700% compared with dry conditions. The driver applied the brakes in approximately half of the events.

Fitch et al. [34] analyzed all the lane-change events in the 100-vehicle NDS database. Of the 135 events 3 were crashes and 132 were near crashes. They observed that this type of event occurred in less than two seconds. Most of the lane-changing events could be avoided if the driver had properly monitored their surroundings. The traffic conflicts related to right-turn maneuvers were also analyzed.

Lv et al. [29] analyzed right-turn events with distracted drivers at intersections using the SHRP 2 database. The study selected 581 events, of which 208 included distracted driving and 373 did not include distracted driving. They concluded that the traffic control and the lane that the vehicle occupied were correlated to the distracted driving. The driving time and the traffic density also had an influence over the distracted driving behaviors.

Das et al. [35] used the SHRP 2 database to analyze the impact of weather conditions on driver lane-keeping ability. They found that foggy weather conditions reduced lane-keeping performance. Other variables such as traffic conditions, lane changing, geometrical characteristics, and driver marital status influence lane keeping performance.

Alshatti [36] analyzed the driver behavior risk factors in roadway departure crashes/near-crashes using the SHRP2 database. The author found a correlation between driver attentiveness and roadway departure conflicts. Road alignment and a driver's total mileage per year were significant variables affecting the risk for roadway departure crashes/near-crashes.

2.4.3 Commercial Motor Vehicles

NDS has also been an important tool for improving traffic safety in the commercial transportation industry. Hickman et al. [15] analyzed the response of two carriers after an on-board camera system was installed and a behavioral coaching technique was applied. The study observed a relevant reduction in the mean of safety-related events/10,000 miles in both carriers. Another study focused on two companies that used NDS data to provide supervisory coaching. They observed that monitoring and coaching had a higher difference in reducing harsh braking events compared with camera-only monitoring. The monitoring and coaching approach was effective in decreasing driving errors [17]. A third study observed that risky driving behaviors had a higher reduction rate when a feedback system was combined with supervisory coaching and alerts [16].

CHAPTER 3. METHODOLOGY

This study seeks to find factors that negatively impact traffic safety by observing and analyzing naturalistic driving data collected by the CTL Taxicab Company using the Samsara onboard NDS system. The data were recorded between January of 2020 and April of 2020 on Oahu, Hawaii, primarily in the central portions of Honolulu. The analysis was based on two foundations: (1) A literature review about road safety studies, previous naturalistic driving studies and their main findings, and (2) Data collection consisting of: data collection tool description, explanation of the coding process, and database description that includes the definition of all variables. After both foundations were completed, the analysis techniques were defined, and the actual analysis commenced in two stages: (i) basic statistics and correlations; and (ii) model development. Figure 3.1 represents the methodology of the study.

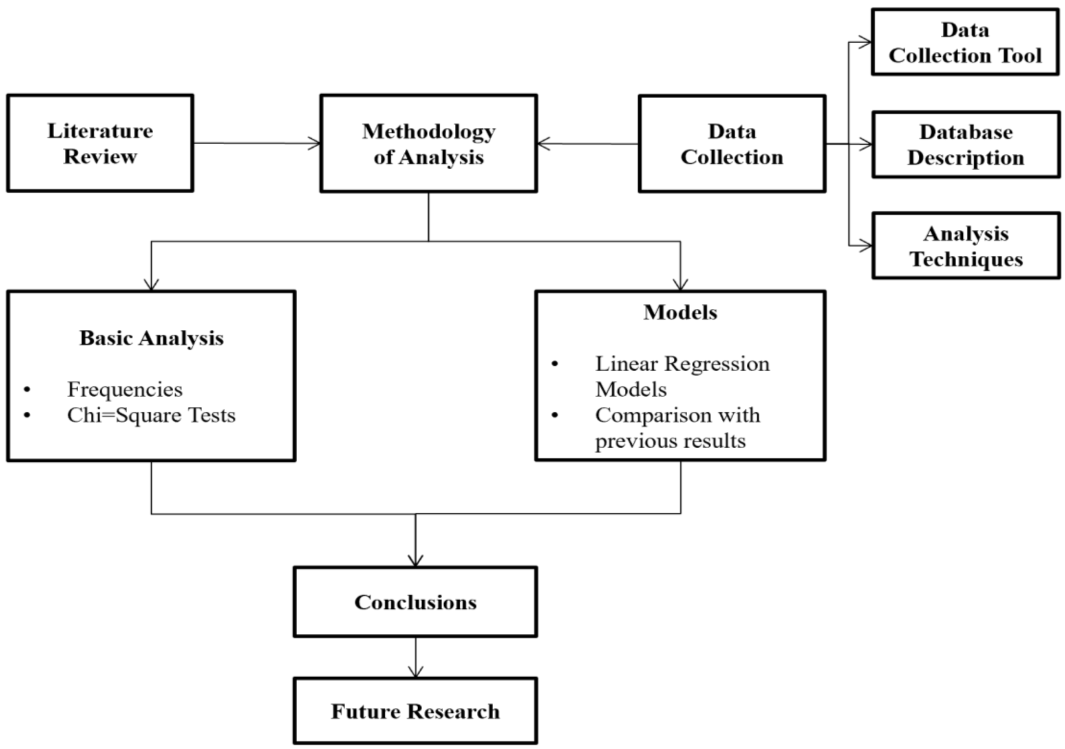

Figure 3.1 Outline of methodology

Two types of approaches were used for coding the database: (i) automated data collection by Samsara system, and (ii) manual data collection by user input. The automatic data collection tool and the database inputs are described below.

3.1 Automatic Data Collection Tool

The Samsara system automatically provides vehicle location on a map, recent routing, instant speed and speed profile (from -10 to +10 seconds from the accelerometer's triggering of a "harsh event"), front view, and vehicle interior including recognition of the driver's attention level by facial orientation, as shown in Figure 3.2.

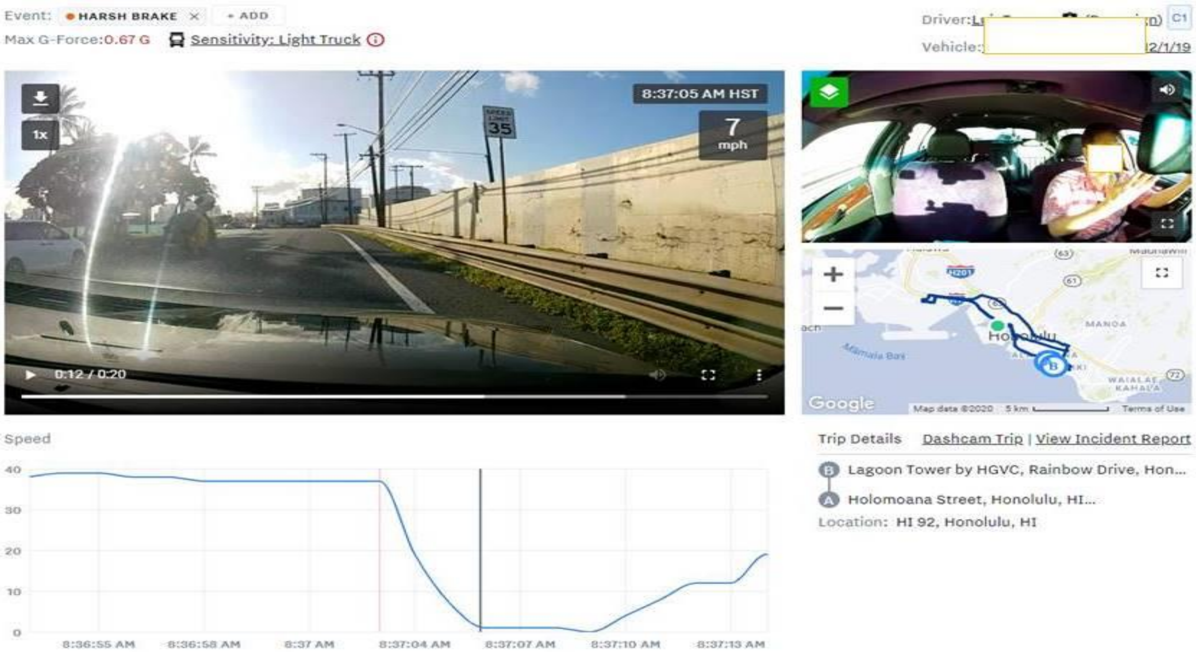

Figure 3.2 Screen capture of interface used to record naturalistic driving database

Figure 3.2 shows a sample of the screen capture pertaining to one harsh event. There are four main sources of information on this interface. Top left is the 20 second video with the red vertical line demarcating the triggering of the harsh event. Below the video is the corresponding speed profile over time. The interface includes a video of the interior of the vehicle with typically a good view of the driver, and the route of the taxi. Other details include an exact time stamp and GPS derived speed overlaid on the forward-looking video, the maximum g-force, driver and vehicle info, and the date of the event. Both forward and interior view videos can be played on full screen mode for better visual recognition of the events. Some of the limitations can be seen in this video such as glare (which makes it difficult to see the pedestrian crossing the highway illegally) and the number of lanes on this highway segment. The latter is easy to fix by observing the exact location on a contemporary GIS or mapping tool.

The system is on continuous recording mode and its storage of events is based on an accelerometer that detects harsh events. Unlike the 100-car and the SHRP 2 studies where the analyst had to view hundreds of vehicle miles of video and manually select and store noteworthy events for subsequent consideration and analysis, in the Samsara system of the CTL taxi vans, the system administrator set an acceleration or deceleration which in the system's nomenclature is referred to as the G-force level. To be consistent with the 100-car NDS (Dingus et al [11]), the threshold in the CTL taxi van units was set at 0.5g and we refer to it herein as the G-force. After that setting in input, all events that meet or exceed the stated acceleration or deceleration level are recorded from -10 to +10 seconds around the time stamp of the G-force triggering event. The system categorizes events as "harsh brake", "harsh turn" and "harsh acceleration", and includes additional labels such as "distracted," "near collision", "mobile usage" and others, as applicable. The user can filter a specific type of incident in a study based on the labels attached to the harsh event. Some labels are related to the driver behavior, noted by the facial recognition system, and labelled as distracted driver, mobile usage and seat belt use.

Samsara's system has an automatic traffic safety system that works as shown in Figure 3.3. The device installed in the vehicle identifies risky behaviors through the facial recognition and the accelerometer sensors. The system alerts the driver in real time with a voice message describing the unsafe behavior. The sensors assign labels to each incident and save all the incident information onto a cloud system.

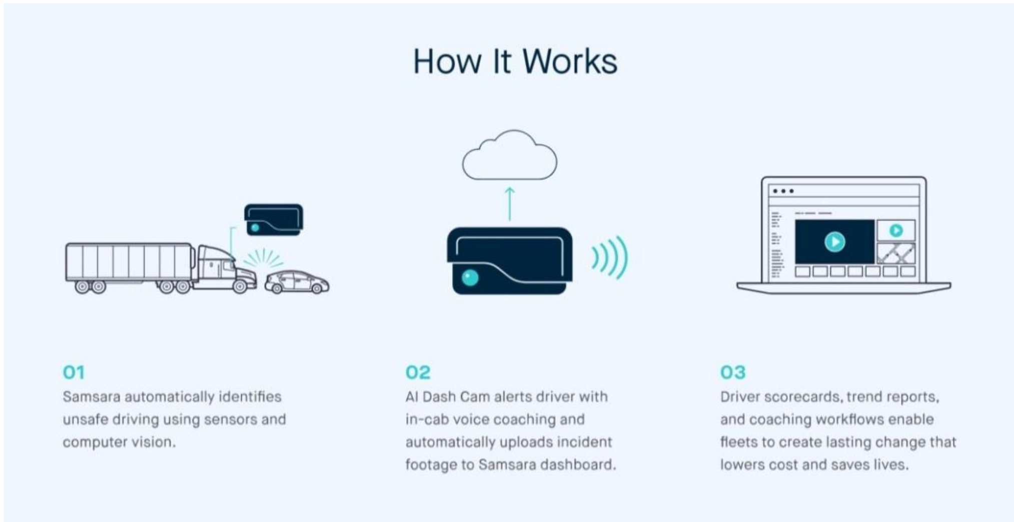

Figure 3.3 The Samsara system

The system's online platform is where the coaches can access a safety report that shows the fleet or a driver's safety trends and data. Coaches can track safety trends and improvements over time, identify risky driving practices and measure changes in safety culture over time. A driver's safety report shows his or her safety score which is calculated based on the frequency of harsh events, and amount of time driven over the speed limit, as shown in Figure 3.4. According to the Samsara's website [8]: "Customers using Samsara's technology platform—which includes tools for driver safety, fleet management, and compliance—have seen marked improvements in both severity and volume of accidents, reducing overall costs by as much as 50%."

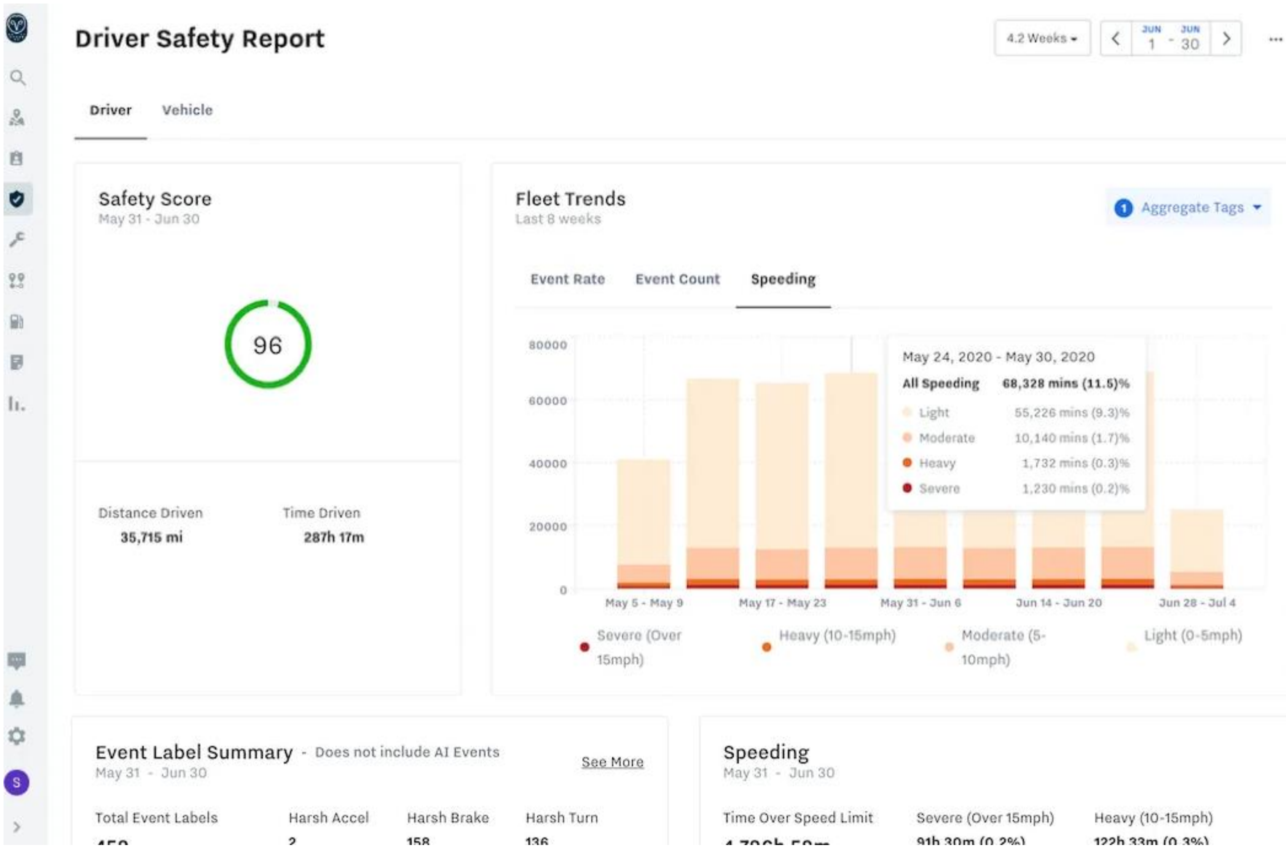

Figure 3.4 Safety report of Samsara online fleet management dashboard

3.2 Database Coding and Description

The database for the study by the UHM team was manually selected from the CTL full safety report provided by the Samsara system. We selected the harsh events that met the 0.5 g threshold or that included a "near-collision" or "collision" label. Harsh events for which at least two UHM observers could not observe any reason for the triggered recording were put in a separate bin and were not included in the analysis herein. The database coding was carried out in two phases.

Phase I: Test. The CTL Company gradually started to install the system in their fleet in October 2019 and the process took approximately three months; it completed in December 2019. The UHM research team utilized the data from this period to define the variables and to create the codes that describe the event types. A total of 444 event videos were extracted from this period, out of which 207 were coded and 237 were not coded because no near-miss event could be observed. This phase was crucial to finalizing the data descriptions and to standardizing the data coding.

Phase II: Final Database. The database was coded with data recorded from January 2020 to April 2020 at which point the data collection was interrupted due to the Covid-19 pandemic (taxi operations were largely shut down.) During this phase, the whole fleet had the system installed, and a total of 533 events were extracted from which 402 events were coded, and 131 events were not coded because no near-miss event could be observed. The final database is described below.

3.3 Database Description

The CTL/UHM database consists of 402 events, of which 398 are near-crashes and four are actual crashes, recorded by 233 vehicles between January and April of 2020 on Oahu, Hawaii. The database consists of two different types of data, the automatically coded data provided by the Samsara system for each event, and the manually inputted data by the UHM team. The Samsara variables include the date and time, driver data, the type of event, kinematics data, and labels according to what happened, such as "harsh brake", "harsh turn" and "harsh acceleration". It also includes additional labels such as "distracted," "near collision", "mobile usage" and others, as applicable. Our work included the verification of each of the labels and the addition of more after repeat observations of each event video.

The UHM variables were divided in five groups. The first group of variables describes the event characteristics, the second describes the road characteristics, the third represents the environmental characteristics, the fourth describes the driver characteristics and, the last one describes the vehicle characteristics. We developed 63 event descriptions (or reasons) that describe a harsh event. Examples of this are given later. Figure 3.5 presents a summary of the database variables; all the variables are defined in the following sections.

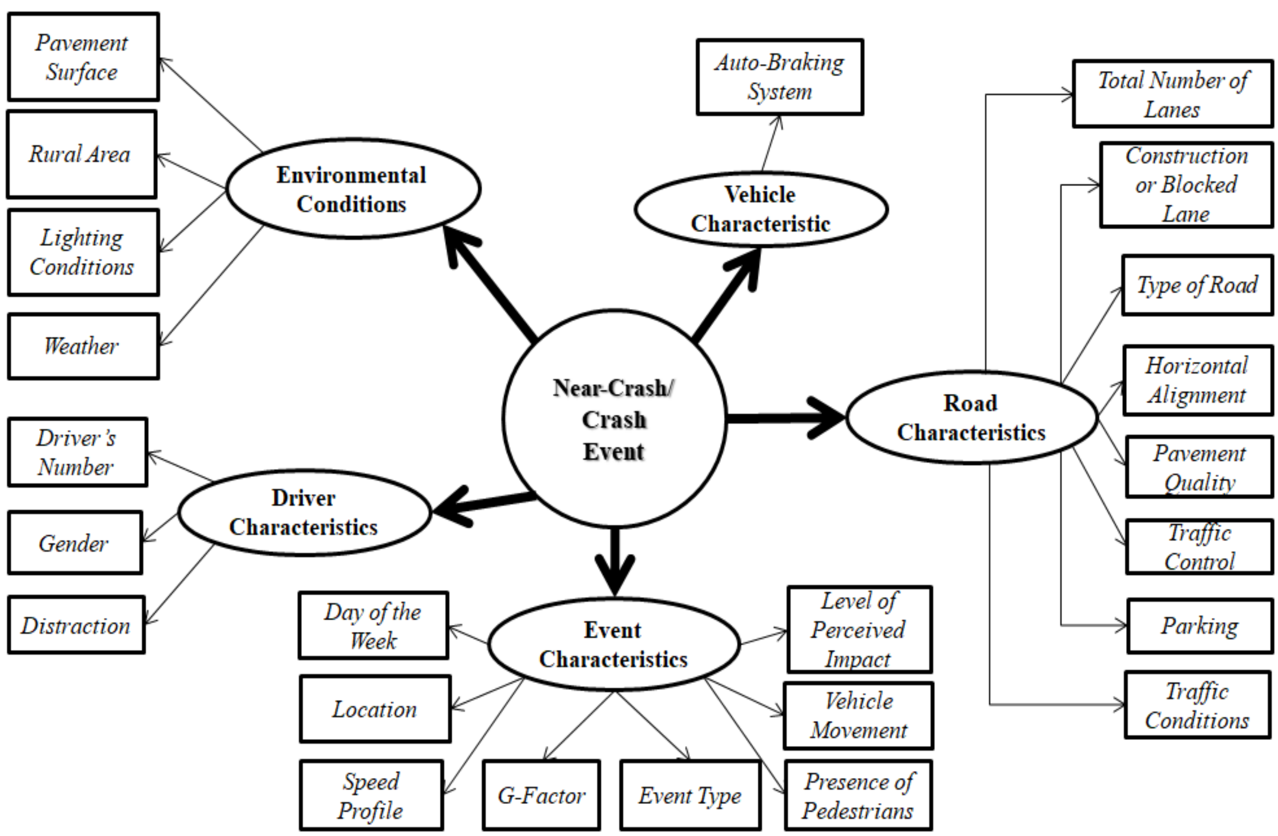

Figure 3.5 Summary of the CTL/UHM database variables

3.3.1 Event Characteristics

- Day of the Week: The System provides the event's exact day and time from which we found and inserted the day of the week.

- Location: Samsara includes a GPS system that shows the location where the near miss event happened on a map and provides the corresponding street address.

- Speed Profile: The near miss event is recorded as a 20 second video and it provides the vehicle's speed per second, 10 seconds before and 10 seconds after the harsh event was detected.

- G-Factor: The acceleration, deceleration or lateral force or g-force was measured by a sensor and the system uses the g-force as a parameter to classify the event as a harsh event, based on administrator setting of the g-force (0.5g in the case of the Samsara devices of CTL). During the early part of the study, in Fall 2019, when the taxis were gradually outfitted with the Samsara devices, CTL coaches and the UHM team observed many events and concluded that events with a g-force of less than 0.5g did not present risks and they did not provide coachable lessons or research challenges. Notably, the 100-car NDS also set the minimum g-force at 0.5 for selecting near-crash and crashes events.

- Presence of Pedestrians: This variable indicates whether there is a pedestrian involved in the near crash event.

- Vehicle Movement: This variable indicates the left turning, right turning or going through movement of the CTL vehicle.

- Event Type: The UHM team developed many codes describing what happened during the near miss events. This variable is tailored to ask the question What Happened? In all cases, up to three reasons (event types) were enough to fully describe in code the near-miss event. In order to identify the vehicles, the subject vehicles with the onboard Samsara system were called V1, and all the other vehicles were called V2. If there were multiple other vehicles involved (a rare occasion), they were numbered in the order they were involved in the scene, such as V2, V3, etc. Pedestrians, motorcycles, bicycles and police officers were coded according to their respective status. The events were divided in three different categories: events caused by V1, events caused by V2 and events that involved non-vehicular participants. All the codes are represented in Table 3.1: codes for type of event.

Table 3.1 Codes for Type of Event

Vehicle 1	Vehicle 2	Non-Vehicle
V1 ABRUPT BRAKING OR STOPPING	V2 ABRUPT BRAKING OR STOPPING	ANIMAL ON THE ROAD
V1 ABRUPT LEFT TURN	V2 ABRUPT LEFT TURN	BICYCLE ON LANE
V1 ABRUPT RIGHT TURN	V2 ABRUPT RIGHT TURN	DRIVEWAY INCIDENT
V1 ALMOST HEAD-ON V2	V2 ALMOST HEAD-ON V1	MOTORCYCLE CRASH
V1 ALMOST REARENDED V2	V2 ALMOST REARENDED V2	PARKING INCIDENT
V1 ALMOST SIDESWIPED V2	V2 ALMOST SIDESWIPED V1	PEDESTRIAN IN MID-BLOCK XING
V1 ALMOST T-BONED V2	V2 ALMOST T-BONED V1	PEDESTRIAN MID-BLOCK ILLEGAL
V1 ALTERCATION/ROAD RAGE	V2 CUT IN FRONT OF V1	PEDESTRIAN XING IN DONTWALK
V1 CUT IN FRONT OF V2	V2 DID NOT YIELD	PEDESTRIAN XING IN WALK
V1 DID NOT YIELD	V2 ERRATIC BEHAVIOR	POLICE OFFICER
V1 ERRATIC BEHAVIOR	V2 LANE CHANGE/WEAVING	
V1 EXHIBITION OF ACCELERATION	V2 LATE RESPONSE TO RED	
V1 LANE CHANGE/WEAVING	V2 LT FROM TH LANE	
V1 LATE RESPONSE TO BRAKE	V2 LT THRU OPPOSING TRAFFIC	
V1 LATE RESPONSE TO RED	V2 RAN RED LIGHT	
V1 LATE RESPONSE TO STOP SIGN	V2 REAR ENDED	
V1 LATE RESPONSE TO YIELD	V2 RT FROM TH LANE	
V1 LT FROM TH LANE	V2 RTOR	
V1 LT THRU OPPOSING TRAFFIC	V2 SIDESWIPED V1	
V1 LTOR	V2 STOPPED ON YELLOW	
V1 RAN RED LIGHT	V2 WENT THRU STOP	
V1 REARENDED V2	V2 WRONG WAY	
V1 RIGHT ANGLE NEAR COLLISION		
V1 RT FROM TH LANE		
V1 RTOR		
V1 SPEEDING		
V1 STOPPED ON YELLOW		
V1 TH FROM LT LANE		
V1 WENT THRU STOP		
V1 WRONG WAY		

- Perceived Impact: This factor represents the impression of the event on the observer of the video and his or her corresponding assessment of severity of the near miss event one a scale from 1 to 6. To help us during coding, we referred to this as the Wow Factor. The lower range represents less impressive incidents with impacts 1 to 3. The upper range represents incidents that clearly make an impression (i.e., give the observer a slight to major shock during the first time that he or she views the video of the event). These are impact scores from 4 to 6, with 6 representing an actual collision. Examples of Perceived Impact Factor from 1 to 6 are given below, based on actual events:
  - Sudden movement of V1 with small risk given the space available from other vehicles, people or objects.
  - Near rear end at lower speed ending at close proximity or at freeway speed ending at least one car length away from vehicle in front.
  - Near rear end at freeway speed ending at close proximity to the vehicle in front.
  - Vehicle suddenly weaves two or more lanes on freeway or highway, or vehicles stopped on lanes 1 and 2 for a pedestrian crossing in front of them, and another vehicle drove through the crossing on lane 3.
  - Motorcycle passenger falls off onto freeway lane, at speed.
  - Collision with another vehicle, person, or fixed object.

3.3.2 Road Characteristics

- Total Number of Lanes: The total number of lanes on both directions.

- Construction/Blocked Lane: Represents lane closed fully or partially due to construction activity or blocked lane due to other reasons such as maintenance activity, temporary security closure, police investigation, etc.

- Type of Road: Includes seven types of road in this database, largely in accordance with their official classification: freeway, highway, major arterial, minor arterial/collector, wide local street, regular local street and narrow local street; the last three categories breakdown the typical “local street” classification to more detailed classes to better reflect the variety of local streets in Honolulu. A narrow local street is approximately the same as the typical alleyway, with or without parking, as applicable.

- Horizontal Alignment: A road segment can be either straight or curvy; this is a binary 0/1 variable. (Vertical alignment was not assessed or coded because the up or down grade of a typical road segment is difficult to ascertain from the video recording.) However, if steep grade appeared to be a factor, we were prepared to make special notes, but such an event was not noted.

- Pavement Quality: Four types of pavement quality were used in this study to represent the visual quality of the road pavement: rough, rough spots, good and very good. The very good label is applied to pavement that appeared to be a newly resurfaced road.

- Traffic Control: Represents the traffic control device immediately adjacent or applicable to the near miss event. Highway and freeway events were classified as “free” (short for free flow), and “none” represents near misses where no traffic control was present nearby or related to the event.

- Parking: Identifies a road location where the event occurred had parked vehicles on the road.

- Traffic Conditions: This variable classifies the density of traffic in five levels. Traffic condition level 1 is very light, level 2 represents light, level 3 represents medium, level 4 represents heavy, and level 5 represents very heavy traffic conditions. These correspond to the Highway Capacity Manual's Levels of Service as follows: 1=A and B, 2=C, 3=D (i.e., restricted freedom of maneuver becomes clear at this level), 4=E and 5=F (i.e., crawling or stopping conditions.)

3.3.3 Environmental Characteristics

- Pavement Surface: Wet or dry pavement surface at the time of the event; 0/1 variable.

- Rural Road: This binary 0/1 variable identifies rural roads by analyzing the GPS location and the characteristics observed on the video recordings.

- Lighting Conditions: Five types of lighting conditions were identified in this study, cloudy, dark, sunlight/clear, glare/sunset/sunrise and tunnel lighting conditions.

- Weather: Several conditions were defined but only two were present in the database: clear and light rain conditions.

3.3.4 Driver Characteristics

- Gender: The taxicab company provided input for each driver's gender, for the drivers in the database.

- Driver's Number: In this is study, each driver has a specific number, the CTL driver identification number, because the driver names were removed from the database.

- Distraction: This variable describes whether the driver was distracted, whether he or she was using a handheld mobile device. The Samsara system identifies driver distraction through a facial recognition system.

3.3.5 Vehicle Characteristics

- Presence of Auto Braking System: The taxicab company provided us with the list of vehicles that had an autobrake system installed and a binary variable was created to represent the availability of this system. Notably the fleet is rather uniform; it consists of Honda Odyssey and Toyota Sienna vans only. Both vehicles have the same size and similar dynamics, but about half of them do not have an emergency braking assist system.

3.4 Analysis Techniques

After the database was created in Microsoft Excel, the SPSS statistical software was selected to analyze the data. The data analysis process was divided in two sections. First, the basic analysis presents the frequency distributions of all variables and selected results from correlations analysis. All the results are presented in Chapter 4 and Appendix A. The second section includes linear regression models that present the relationship between the variables and selected event types. All the results are presented in Chapter 5.

CHAPTER 4. BASIC ANALYSIS

This chapter presents the frequency analysis for all variables in the dataset, and correlations analysis among the variables using the χ^2 (chi-square) test. For the χ^2 correlation tests, the variables were selected based on previous studies reviewed in the literature and on our observations while coding the events in the dataset. The χ^2 tests were conducted separately for two groups: expressways (highways and freeways) and urban roads (arterials and local roads), because these two classes of road have different operating characteristics (i.e., they are governed by the uninterrupted and interrupted flow principles). Appendix A includes the results of all the χ^2 tests performed in this study.

4.1 Variable Frequencies and Distributions

The basic descriptive measurements and frequency of each variable of in the database is presented in this section separately for the characteristics of events, road segments, environment and driver.

4.1.1 Event Characteristics

4.1.1.1 Day of the week

Figure 4.1 shows the distribution of the events throughout the days of the week was similar. Thursday was the day of the week with the greatest number of events: 17.7% of the events occurred on Thursdays, while Sunday was the day with the lowest number of events at 11.9%.

Day of the Week

	No. of Obs.	%
SUNDAY	48	11.9%
MONDAY	51	12.7%
TUESDAY	55	13.7%
WEDNESDAY	60	14.9%
THURSDAY	71	17.7%
FRIDAY	55	13.7%
SATURDAY	62	15.4%

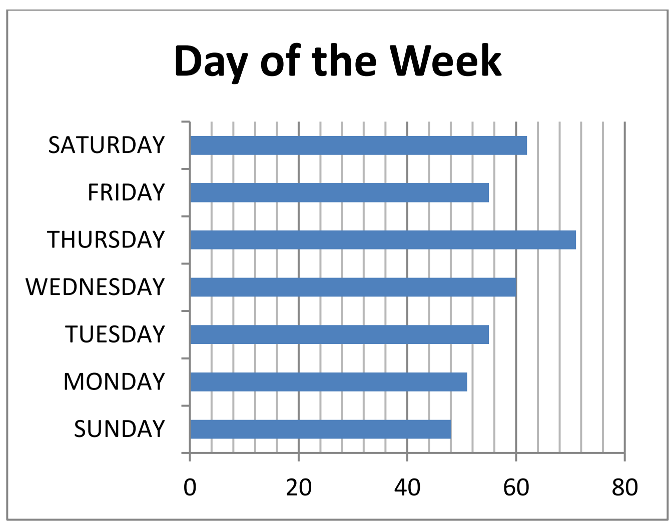

Figure 4.1 Event frequency per day of week.

4.1.1.2 Vehicle autobraking system

Figure 4.2 shows the event frequency per vehicle autobrake system. In only 31.1% of the incidents the subject vehicle (V1) had an automatic auto braking system installed. In most of the events related to vehicles with an auto braking system installed, the system did not seem to have a significant impact in the event occurrence, as explained later.

AUTO BRAKING SYSTEM

	No. of Obs.	%
NO	277	68.9%
YES	125	31.1%

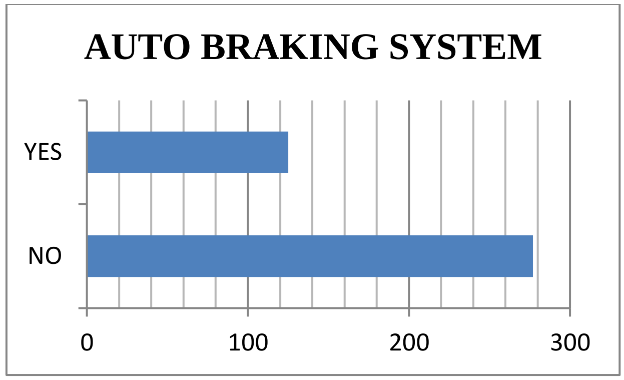

Figure 4.2 Event frequency per vehicle autobrake system.

4.1.1.3 Type of events

A total of 63 codes were created to describe what happened during a crash or near-crash event. Each event could have up to three event descriptors, as shown in Table 4.1, where Code 1 is always the code that represents the primary event description. The events were also classified in three types. First, events related to “non-vehicles” (here “vehicles” reflects those with four or more wheels) including bicycles and motorcycles as well as pedestrians and cases with animals on the road. Second, events in which the subject vehicle (V1) was primarily responsible for the near-crash or crash and third, events in which another vehicle (V2) was primarily responsible for the near-crash or crash.

Table 4.1 Event Type Codes Frequency

Event Category	Code	Type of Event	No. of Obs. Code 1	No. of Obs. Code 2	No. of Obs. Code 3	Total No. of Obs.	Total per Category	% in Category	% of Total No. of Obs.
Non-Vehicle	15	PEDESTRIAN IN MID-BLOCK XING	24	3	3	30	103	29%	17%
	18	PEDESTRIAN XING IN WALK	13	4	2	19		18%	0%
	16	PEDESTRIAN MID-BLOCK ILLEGAL	13	0	1	14		14%	0%
	5	BIKE ON LANE	5	4	0	9		9%	0%
	17	PEDESTRIAN XING IN DONTWALK	6	1	2	9		9%	0%
	7	DRIVEWAY INCIDENT	3	4	1	8		8%	0%
	14	PARKING INCIDENT	3	2	3	8		8%	0%
	19	POLICE OFFICER	2	1	0	3		3%	0%
	4	ANIMAL ON THE ROAD	1	1	0	2		2%	0%
	60	MOTORCYCLE CRASH	1	0	0	1		1%	0%
Vehicle 1	26	V1 ALMOST REARENDED V2	66	24	0	90	261	34%	15%
	10	V1 LATE RESPONSE TO RED	33	3	4	40		15%	6%
	1	V1 ABRUPT BRAKING OR STOPING	17	4	3	24		9%	4%
	9	V1 LATE RESPONSE TO BRAKE	7	3	7	17		7%	3%
	28	V1 ALMOST T-BONED V2	2	2	11	15		6%	2%
	8	V1 LANE CHANGE/WEAVING	4	6	4	14		5%	2%
	20	V1 RAN RED LIGHT	7	3	2	12		5%	2%
	2	V1 ABRUPT LEFT TURN	7	0	0	7		3%	1%
	3	V1 ABRUPT RIGHT TURN	2	3	0	5		2%	1%

Event Category	Code	Type of Event	No. of Obs. Code 1	No. of Obs. Code 2	No. of Obs. Code 3	Total No. of Obs.	Total per Category	% in Category	% of Total No. of Obs.
	25	V1 ALMOST HEAD-ON V2	2	0	3	5		2%	1%
	34	V1 WRONG WAY	2	1	2	5		2%	1%
	6	V1 DID NOT YIELD	3	1	0	4		2%	1%
	61	V1 EXHIBITION OF ACCELERATION	3	0	0	3		1%	0%
	11	V1 LATE RESPONSE TO STOP SIGN	1	1	1	3		1%	0%
	21	V1 RIGHT ANGLE NEAR COLLISION	2	1	0	3		1%	0%
	27	V1 ALMOST SIDESWIPED V2	0	2	0	2		1%	0%
	57	V1 ALTERCATION/ROAD RAGE	1	0	1	2		1%	0%
	55	V1 ERRATIC BEHAVIOR	1	0	1	2		1%	0%
	32	V1 RT FROM TH LANE	1	0	1	2		1%	0%
	62	V1 SPEEDING	1	0	1	2		1%	0%
	58	V1 LATE RESPONSE TO YIELD	1	0	0	1		0%	0%
	30	V1 LT FROM TH LANE	0	0	1	1		0%	0%
	12	V1 LT THRU OPPOSING TRAFFIC	1	0	0	1		0%	0%
	23	V1 STOPPED ON YELLOW	0	0	1	1		0%	0%
	29	V1 CUT IN FRONT OF V2	0	0	0	0		0%	0%
	13	V1 LTOR	0	0	0	0		0%	0%
	31	V1 REARENDED V2	0	0	0	0		0%	0%
	22	V1 RTOR	0	0	0	0		0%	0%
	24	V1 TH FROM LT LANE	0	0	0	0		0%	0%
	33	V1 WENT THRU STOP	0	0	0	0		0%	0%
Vehicle 2	40	V2 CUT IN FRONT OF V1	23	5	13	41	147	28%	7%
	42	V2 LANE CHANGE/WEAVING	15	4	1	20		14%	3%
	56	V2 ERRATIC BEHAVIOR	4	6	8	18		12%	3%
	41	V2 DID NOT YIELD	3	5	3	11		7%	2%
	35	V2 ABRUPT BRAKING OR STOPING	4	0	4	8		5%	1%
	38	V2 ALMOST SIDESWIPED V1	1	1	5	7		5%	1%
	46	V2 RAN RED LIGHT	1	3	1	5		3%	1%
	45	V2 LT THRU OPPOSING TRAFFIC	1	2	1	4		3%	1%
	48	V2 RTOR	2	2	0	4		3%	1%
	51	V2 WENT THRU STOP	1	3	0	4		3%	1%
	36	V2 ABRUPT RIGHT TURN	0	1	2	3		2%	0%
	39	V2 ALMOST T-BONED V1	0	0	3	3		2%	0%
	44	V2 LT FROM TH LANE	1	1	1	3		2%	0%
	50	V2 STOPPED ON YELLOW	0	0	3	3		2%	0%
	37	V2 ALMOST HEAD-ON V1	0	0	2	2		1%	0%
	59	V2 ALMOST REARENDED V21	0	0	2	2		1%	0%
	43	V2 LATE RESPONSE TO RED	2	0	0	2		1%	0%
	54	V2 REAR ENDED V3	0	0	2	2		1%	0%
	47	V2 RT FROM TH LANE	0	0	2	2		1%	0%
	49	V2 SIDESWIPED V1	0	1	1	2		1%	0%
	52	V2 WRONG WAY	0	1	0	1		1%	0%
	53	V2 ABRUPT LEFT TURN	0	0	0	0		0%	0%
	99	MULTIPLE CODES	109	0	0	109	109	100%	18%
Grand Total			402	109	109	620	620		

In the first category, we observed that 43 of the 103 events had the presence of a pedestrian, only 5 events recorded had the presence of a bicycle and all the others were related to motorcycles, police officers and driveway incidents.

The second category had the largest number of codes and events recorded in this dataset. We observed that event type 26, vehicle 1 almost rear-ended vehicle 2, was recorded 90 times, which was the most frequent event type in the dataset, representing 34% of the total number of observations. V1 late response to red was recorded 40 times, representing 15% of the category, followed by V1 abrupt braking or stopping, which represents 9% of the category, V1 late response to brake which represents 7% of the category, V1 almost t-boned v2, which represents 6%, V1 lane changing/weaving representing 5%, V1 ran red light representing 5% of this category. Other types of incidents related to right turns and left turns, near head-on collisions, exhibition of acceleration, etc. represented a small percentage in this category.

In the third category, we observed that events related to lane changing were the most frequent ones. V2 cut in front of V1 was recorded 41 times, representing 28% of this category, followed by V2 lane changing/weaving represented 14% of the category. Erratic behavior was also a significant event in this category, representing 12% of it. All the other events represented a small percentage in the dataset.

4.1.1.4 Perceived impact factor

Figure 4.3 shows the event frequency per level of perceived impact. The perceived impact represents the severity of the event according to the researchers' point of view. A scale from 1 to 6 was used in which 1 is an event that has a minor impact on the researchers as they reviewed the incident, 5 had a startling or most severe impact, and 6 was an actual crash; this database includes four crashes. We observed that in this study most of the events have a medium impact and were classified as impact 2 and 3. Specifically, 10% of the incidents were classified as impact factor 1, 39.1% of the events were classified as impact factor 2, 39.6% of the events were classified as impact factor 3, 9% were classified as impact factor 4, 1.5% were classified as impact factor 5 and only 1% were classified as impact factor 6.

LEVEL OF PERCEIVED IMPACT FACTOR

	No. of Obs.	%
1	40	10.0%
2	157	39.1%
3	159	39.6%
4	36	9.0%
5	6	1.5%
6	4	1.0%

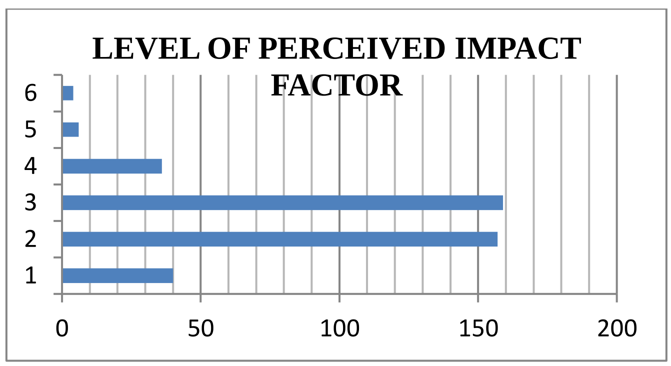

Figure 4.3 Event frequency per level of perceived impact

4.1.2 Road Characteristics

4.1.2.1 Horizontal alignment

Figure 4.4 shows the event frequency per road horizontal alignment. In this database, only 17.2% of the events occurred on a curvy road segment, and 82.8% occurred on a straight road segment.

ROAD HORIZONTAL ALIGNMENT

	No. of Obs.	%
STRAIGHT	333	82.8%
CURVY	69	17.2%

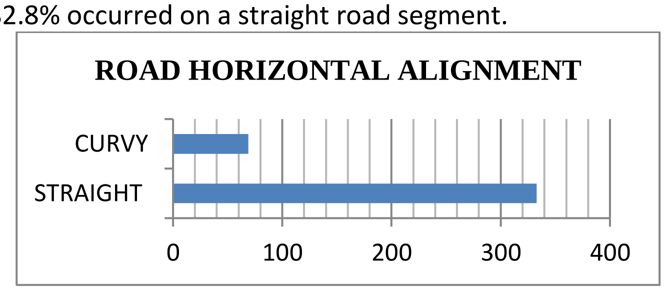

Figure 4.4 Event frequency per road horizontal alignment

4.1.2.2 Parking

Figure 4.5 shows the parking related event frequency. This study recorded 69 events that occurred on roads with vehicles parked. In 82.7% of the events recorded there was no influence of parking maneuvers or parked vehicles.

PARKING

	No. of Obs.	%
NO	333	82.8%
YES	69	17.2%

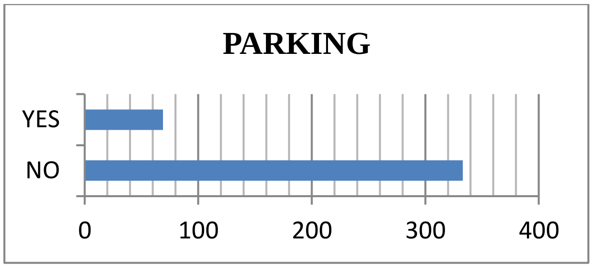

Figure 4.5 Parking related event frequency

4.1.2.3 Pavement quality

Figure 4.6 shows the event frequency per pavement quality. In 6% of the cases the pavement was rough, in 22% the pavement had rough spots, in 55% the pavement was good and in 16% of the cases the pavement was very good which represents that the pavement was visually flawless.

PAVEMENT QUALITY

	No. of Obs.	%
VERY GOOD	66	16.4%
GOOD	221	55.0%
ROUGH SPOTS	90	22.4%
ROUGH	25	6.2%

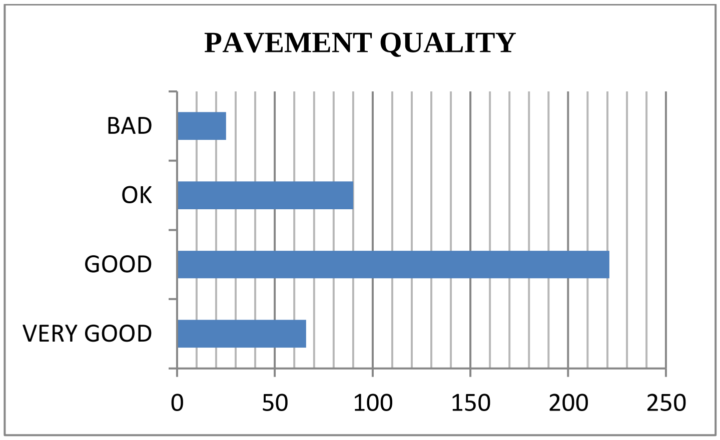

Figure 4.6 Event frequency per pavement quality

4.1.2.4 Construction, or other work zone or blocked lane

Figure 4.7 shows the construction, work zone and/or block lane related event frequency. About 7.5% of the total events occurred on a construction zone or a blocked lane.

CONSTRUCTION, OR OTHER WORK ZONE OR BLOCKED LANE	No. of Obs.	%
NO	372	92.5%
YES	30	7.5%

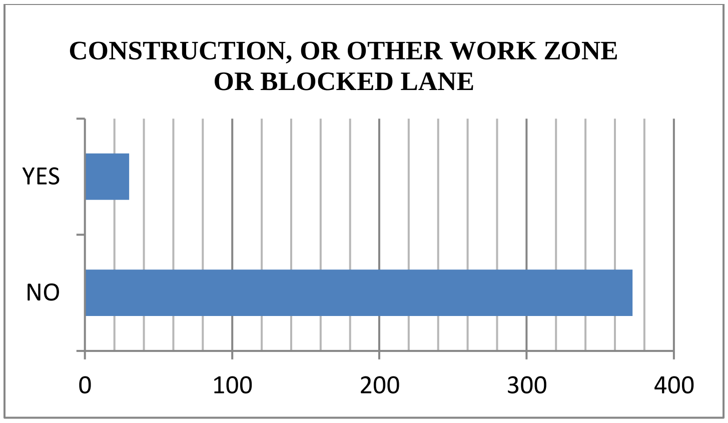

Figure 4.7 Construction, work zone and/or block lane related event frequency

4.1.2.6 Traffic control

Figure 4.8 shows the event frequency per traffic control. A traffic signal was present in 40.3% of the events; this was the most recorded traffic control in this study, followed by the pedestrian crossing, which was present in 10.7% of the events. Only 1.5% of the events were related to stop signs. All the other traffic control types represent a small percentage in this analysis. Clearly, each segment where and incident was observed is regulated by lane channelization and speed limits. We could only note speed limit violations of the V1 vehicles and those were within the "speed limit plus 5 mph" convention of acceptable speed. Lane keeping violations were recorded under event types and presented above.

TRAFFIC CONTROL	No. of Obs.	%
SIGNAL	162	40.30%
NONE	126	31.30%
FREE	63	15.70%
PED XING	43	10.70%
STOP SIGN	6	1.50%
ROUNDABOUT	1	0.20%
SPEEDBUMP	1	0.20%

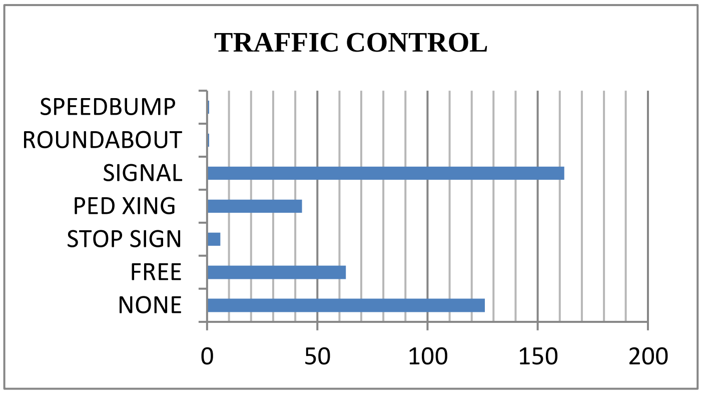

Figure 4.8 Event frequency per traffic control

4.1.2.7 Traffic conditions

Figure 4.9 shows the event frequency per traffic conditions. Most of the events occurred on roads with medium to heavy traffic conditions as shown below.

TRAFFIC CONDITIONS

	No. of Obs.	%
VERY LIGHT	1	0.2%
LIGHT	82	20.4%
MEDIUM	189	47.0%
HEAVY	116	28.9%
VERY HEAVY	14	3.5%

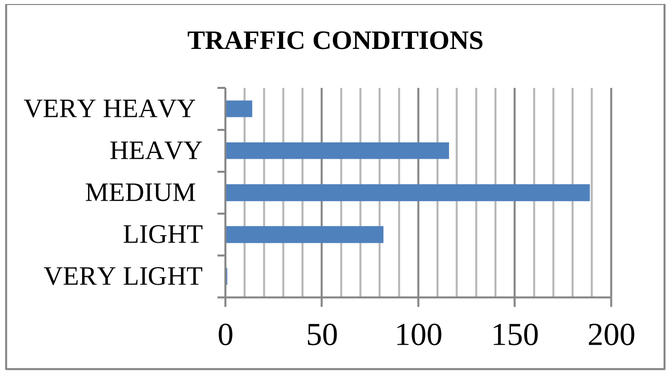

Figure 4.9 Event frequency per traffic conditions

4.1.2.8 Events by Type of Road

Figure 4.10 shows the event frequency per type of road. In this study, 32.3% of the events were recorded on a major arterial, 18.4% were recorded on a highway, 15.2% on a freeway, 14.4% on a minor arterial, 9% on a wide local street, 7% on a regular local street, and only 3.7% of the events were recorded on a narrow local street.

TYPE OF ROAD

	No. of Obs.	%
FREEWAY	61	15.2%
HIGHWAY	74	18.4%
MAJOR ARTERIAL	130	32.3%
MINOR ARTERIAL	58	14.4%
WIDE LOCAL STREET	36	9.0%
REGULAR LOCAL STREET	28	7.0%
NARROW LOCAL STREET	15	3.7%

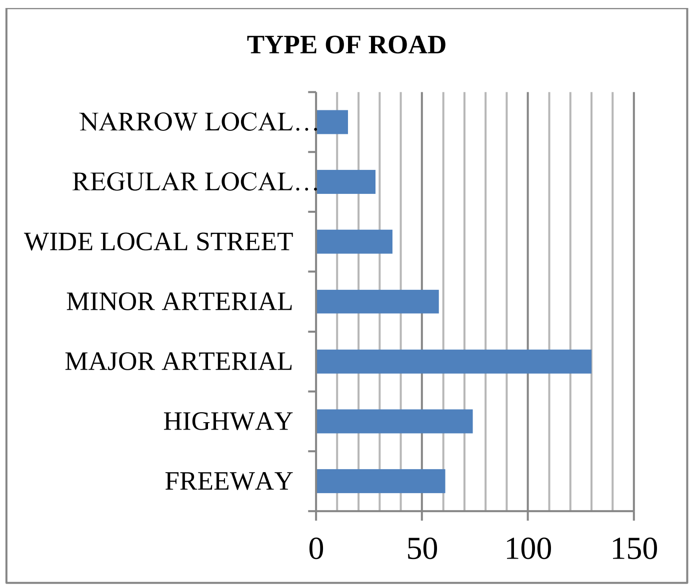

Figure 4.10 Event frequency per type of road

4.1.3 Environmental Characteristics

4.1.3.1 Weather conditions

Figure 4.11 shows the event frequency per weather conditions. Only 3% of the recorded events occurred in light rain; for all the rest the weather was clear.

WEATHER

	No. of Obs.	%
CLEAR	389	96.8%
LIGHT RAIN	13	3.2%

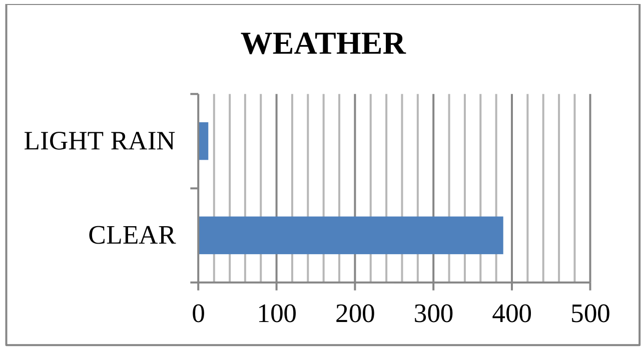

Figure 4.11 Event frequency per weather conditions

4.1.3.2 Rural roads

Figure 4.12 shows the event frequency on rural roads. As the taxicabs (V1 vehicles) operate mostly in urban areas, only 5% of the events were recorded on rural roads.

RURAL ROAD

	No. of Obs.	%
NO	381	94.8%
YES	21	5.2%

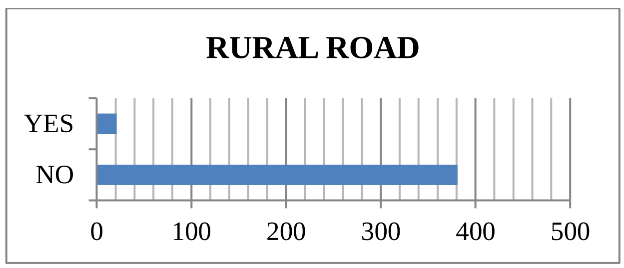

Figure 4.12 Event frequency on rural roads

4.1.3.3 Pavement surface conditions

Figure 4.13 shows the event frequency per pavement surface conditions. Only 10% of the recorded events occurred on wet pavement and all the rest on dry pavement.

PAVEMENT SURFACE

	No. of Obs.	%
DRY	362	90.0%
WET	40	10.0%

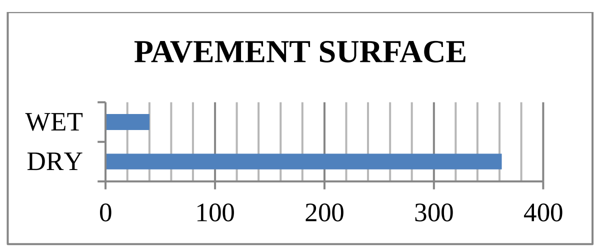

Figure 4.13 Event frequency per pavement surface conditions

4.1.3.4 Lighting conditions

Figure 4.14 shows the Event frequency per lighting conditions. Almost two thirds of the recorded events occurred under sunlight/clear conditions (63%).

LIGHTING CONDITIONS

	No. of Obs.	%
DARK	63	15.7%
SUNLIGHT CLEAR	254	63.2%
CLOUDY	54	13.4%
SUNSET/SUNRISE/GLARE	25	6.2%
TUNNEL	6	1.5%

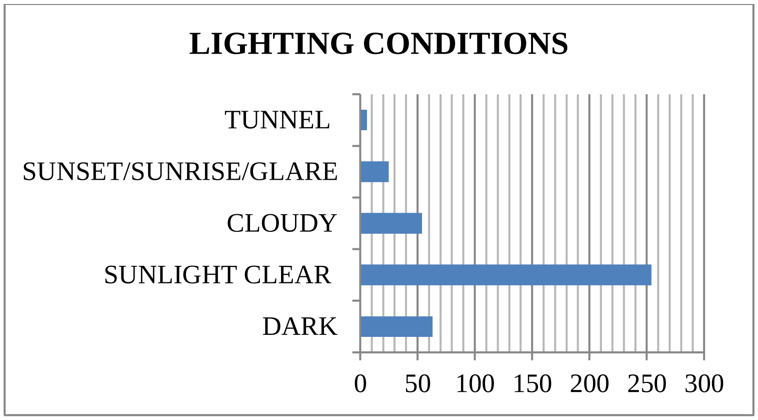

Figure 4.14 Event frequency per lighting conditions

4.1.3 *Driver Characteristics*

4.1.3.1 Gender

Figure 4.15 shows the event frequency per driver gender. During the study period, Charley's Taxi reported 217 drivers of whom 91% were males and 9% were females. However, in this study, 89% of the events recorded involved a male driver of the V1, and in 11% a female driver of the V1; that is, females were slightly over-represented in the near-miss event count and presumably slightly more risky drivers (which appears to go against traditional motorist risk behavior).

DRIVER GENDER

	No. of Obs.	%
FEMALE	44	10.9%
MALE	358	89.1%

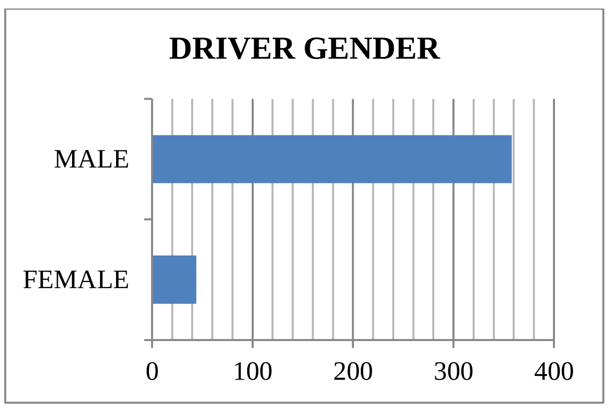

Figure 4.15 Event frequency per driver gender

4.1.3.2 Pedestrians

Figure 4.16 shows the pedestrian related event frequency. In this study 24% of the events recorded involved a pedestrian.

PEDESTRIAN

	No. of Obs.	%
NO	305	75.9%
YES	97	24.1%

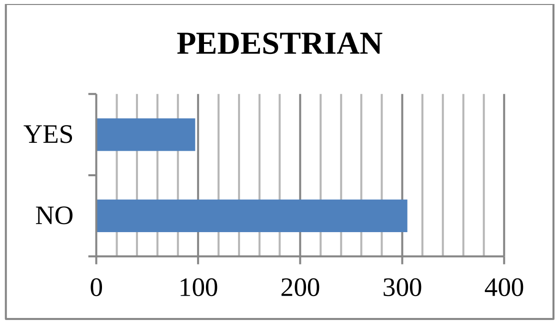

Figure 4.16 Pedestrian related event frequency

4.2 Chi-Square Tests

4.2.1 Urban Roads Versus Highways and Freeways

Due to the substantially different traffic flow characteristics between unterrupted and uninterrupted flow facilities, separate analyses were coducted in the two groups depicted in Figure 4.17.

Figure 4.17 Clusters by type of road

Pearson’s Chi-square correlation tests were performed to indentify the variables strongly related to the two group of roads. For this, the software SPSS was used, the variable road group was selected as the dependent variable and all the other variables were considered independent variables. The statistically significant correlations found are presented below. All these significant Chi-square factors indicate that the independent variable had a significantly different impact on each type of road.

Table 4.2 Type of Road and Construction and/or Blocked Lane Chi-Square Correlation

Count				
Group	Construction, or other work zone or blocked lane		Total	Pearson Chi-Square
	NO	YES		
HWY + FWY	116	19	135	0.000
URBAN ROADS	256	11	267	
Total	372	30	402	

Table 4.2 shows the type of road and construction and/or blocked lane Chi-square correlation. There was a

fairly small number of cases where construction or other work zones were present but events on highways/freeways were significantly more affected by closures than city streets.

Pavement Quality was depicted in four categories, with most of the pavements appearing to be good (221 cases). Table 4.3 shows type of road and pavement quality Chi-square correlation. The events on highways/freeways were significantly more affected by rough pavement conditions.

Table 4.3 Type of Road and Pavement Quality Chi-Square Correlation

Count						
Group	Pavement Quality				Total	Pearson Chi-Square
	VG	G	OK	Rough Spots		
HWY + FWY	30	63	30	12	135	0.029
URBAN ROADS	36	158	60	13	267	
Total	66	221	90	25	402	

Most of the events occurred under medium (189 cases) and heavy (116 cases) traffic conditions. Table 4.4 shows that events on highways/freeways were significantly more affected by heavy (high density) traffic flow conditions.

Table 4.4 Type of Road and Traffic Conditions Chi-Square Correlation

Count							
Group	Traffic Conditions					Total	Pearson Chi-Square
	Very Light	Light	Medium	Heavy	Very Heavy		
HWY + FWY	1	20	59	49	6	135	0.040
URBAN ROADS	0	62	130	67	8	267	
Total	1	82	189	116	14	402	

Table 4.5 shows that the frequency distribution of the seven types of traffic control on the two types of road is significantly different; this is an expected result given the sparse application of traffic signal and stop signs on some highways and their complete absence from freeways.

Table 4.5 Type of Road and Traffic Control Chi-Square Correlation

Count									
Group	Traffic Control							Total	Pearson Chi-Square
	None	Free	Stop Sign	Ped Xing	Signal	Roundabout	Speedbump		
HWY + FWY	32	57	2	1	43	0	0	135	0.000
URBAN ROADS	94	6	4	42	119	1	1	267	

Count									
Group	Traffic Control							Total	Pearson Chi-Square
	None	Free	Stop Sign	Ped Xing	Signal	Roundabout	Speedbump		
Total	126	63	6	43	162	1	1	402	

The distribution of the number of pedestrians on urban roads and highway/freeways shown in Table 4.6. The presence of pedestrians is almost nonexistent on expressways; therefore, this result is expected.

Table 4.6 Type of Road and Pedestrian Chi-Square Correlation

Count				
Group	PEDESTRIAN		Total	Pearson Chi-Square
	NO	YES		
HWY + FWY	134	1	135	0.000
URBAN ROADS	171	96	267	
Total	305	97	402	

Table 4.7 shows that events and crashes on city streets were significantly more severe than on freeways were lower impact near-rear end events dominated the sample.

Table 4.7 Type of Road and Level of Perceived Impact Chi-Square Correlation

Count								
Group	Level of Perceived Impact						Total	Pearson Chi-Square
	1	2	3	4	5	6		
HWY + FWY	14	67	42	6	5	1	135	0.001
URBAN ROADS	26	90	117	30	1	3	267	
Total	40	157	159	36	6	4	402	

We also included more independent variables and we performed Chi-square correlation tests. Involving more variables in the tests also produced significant results but the interpretation of distributions is difficult, as in the example below. These correlations were better addressed with the development regression models. Table 4.8. Traffic Conditions and lighting conditions Chi-square correlation per type of road.

Table 4.8 Traffic Conditions and Lighting Conditions Chi-Square Correlation Per Type of Road

Count										Pearson Chi-Square
Group			Lighting Conditions					Total		
			Cloudy	Dark	Sunlight Clear	Sunset/ Sunrise/ Glare	Tunnel			
HWY+ FWY	Traffic Conditions	Heavy	7	10	27	5	0	49		0.001
		Light	1	4	11	1	3	20		
		Medium	9	3	43	4	0	59		
		Very Heavy	1	0	2	3	0	6		
		Very Light	0	0	1	0	0	1		
	Total		18	17	84	13	3	135		
URBAN ROADS	Traffic Conditions	Heavy	11	9	45	2	0	67		0.002
		Light	13	21	26	1	1	62		
		Medium	11	16	92	9	2	130		
		Very Heavy	1	0	7	0	0	8		
	Total		36	46	170	12	3	267		

4.2.2 Perceived Impact Versus Various Conditions

In Table 4.7 we observed that the type of road is related to the perceptual severity level. Various independent variables were tested against the variable Perceived Impact, and the significant correlations are presented below. Table 4.9 shows that construction and/or blocked lane is statistically significant because it dominated low impact events, that is, it affected high impact events (4,5 or 6) less.

Table 4.9 Level of Perceived Impact and Construction and/or Blocked Lane Chi-Square Correlation

Count								
Construction or other work zone or blocked lane	Level of Perceived Impact						Total	Pearson Chi-Square
	1	2	3	4	5	6		
NO	37	149	145	34	3	4	372	0.012
YES	3	8	14	2	3	0	30	
Total	40	157	159	36	6	4	402	

Table 4.10 shows that the involvement of pedestrian(s) is strongly statistically significant because it is more prevalent in high impact events, that is, it affected high impact events (4,5 or 6) more.

Table 4.10 Level of Perceived Impact and Pedestrian Chi-Square Correlation

Count								
Pedestrian	Level of Perceived Impact						Total	Pearson Chi-Square
	1	2	3	4	5	6		
NO	34	136	110	15	5	4	304	0.012
YES	6	21	48	21	1	0	97	
Total	40	157	158	36	6	4	401	

CHAPTER 5. MODEL DEVELOPMENT

The main goal of this part of the research was to identify the variables contributing to the most frequent types of events of each category. For this analysis, only code 1 and 2 were considered as event type, which means that for the events containing multiple codes, the type of event 99 in code 1 which indicates multiple events, was replaced by code 2. For this investigation, the SPSS software was used to develop stepwise linear regression models. The models presented in this chapter seek to find variables associated with statistically significant riskier or safer contributions to specific near crash event types. A <u>riskier</u> contribution has a statistically significant <u>positive</u> effect on an event type. A <u>safer</u> contribution has a statistically significant <u>negative</u> effect on an event type. In order to do this, all the variables were converted into (0,1) dummy variables to make the linear regression analysis possible. The Stepwise Linear Regression in SPSS builds a model in a sequential manner adding the most significant variables and eliminating the less significant ones while controlling for autocorrelation effects (i.e., independent variables that are correlated to each other are removed.) "The stepwise selection model allows for the analysis of the collection models that might not otherwise have been examined" by allowing SPSS to examine all or a large subset of the independent variables available [36].

For the non-vehicle category, the events with pedestrians were the most frequent. Therefore, specific analysis of contributory factors was conducted for crashes and near-crashes with pedestrians involved. For this, the dummy variable "pedestrian=1" was considered the dependent variable, and all the other variables were independent variables. The best fit models created are discussed in Section 5.1.1.

For Vehicle 1 events, event type 26 "Vehicle 1 almost rear-ended Vehicle 2" was the most frequent event type, so additional specific analysis of contributory factors was conducted for it. For this analysis, the dummy variable "Event Type 26=1" was the dependent variable, and all the other variables were independent variables. For improving the results, three different analyses were performed for different type of roads. The first analysis included only the cases that occurred on major arterials, minor arterials and local roads, the variable that groups all the urban type of roads, "Group 2=1", was used as a selection variable. The best fit models are presented in Section 5.1.2.1.

The second analysis preformed in this section repeated the previous analysis, but in this case only the highways and freeways event were analyzed. The variable that groups the highways and freeways, "Group 1=1", was used for selecting the events of this model. The best fit models are presented in Section 5.1.2.2. The last analysis of this category selected only the events on freeways. This model was created to compare the results with previous studies that analyzed near rear-end and rear-end events on freeways. The models created and the comparison with the results found in the literature review are presented in section 5.1.2.3. For Vehicle 2 events, the event types "V2 cut in front of V1" and "V2 lane-changing or weaving" represented approximately 50% of the events. For this reason, new variable indicating lane-changing movements were created to select those two event types. The "lane changing=1" was set as the dependent variable and all the other variables were independent variables. The research team could observe that the riskier lane changing events occurred on freeway, for this reason this model was developed with events on freeways. The best fit models are presented in Section 5.1.3.

5.1 Regression Analysis

5.1.1 Non-Vehicle Events

This category includes all near miss events related to "non-vehicles" which are defined as pedestrians, police officers, animals, and 2-wheelers (bicycles, and motorcycles).

5.1.1.2. Types of events with pedestrians

Based on the basic analysis presented in Chapter 4, we observed that the types of events 15 (Pedestrian in mid-block crossing), 18 (Pedestrian crossing during walk), 16 (Pedestrian mid-block illegal), and 17 (Pedestrian crossing during don't walk) represent the greatest percentage of the category. For analyzing the event types related to pedestrians, the dummy variable that indicates the presence of a pedestrian (pedestrian=1) was defined as the dependent variable. Table 5.1 shows the best model specifications, the estimation results of Stepwise Linear Regressions: events with pedestrians.

Table 5.1 Estimation Results of Stepwise Linear Regressions: Events with Pedestrians

Dependent: Pedestrian=1	Model 1		Model 2	
Variables in the Model	Estimate	t-statistic	Estimate	t-statistic
Constant	0.054	2.370	0.044	1.921
Traffic Control: Pedestrian Crossing	0.672	12.167	0.645	11.638
Event Type: Bike on lane	0.226	6.062	0.226	6.097
Type of Road: Major Arterial	0.573	5.016	0.507	4.395
Type of Road: Narrow Local Street	0.366	4.033	0.336	3.710
Type of Road: Regular Local Street	0.219	3.229	0.218	3.250
Level of Perceived Factor=4			0.179	2.934
No. of Observations	402		402	
F	49.282		43.293	
Adjusted R^2	0.376		0.388	

Table 5.1 presents two models; one with five independent variables and one with six. Model 1 has an adjusted R^2 value of 0.376 and an F value of 49.282. As expected, the presence of pedestrian crossings significantly increases the crash and near-crash events with pedestrians. Three out of four types of city streets also have large effects on near-miss events that include pedestrians.

Model 2 has an adjusted R^2 value of 0.388, which is higher than the first model. This model includes an additional variable, the level of perceived impact represents the researcher's point of view. The level of perceived impact 4 was selected by the stepwise linear regression as a significant variable, which reflects those events with pedestrians also had a high severity impact.

5.1.2 Vehicle 1 Events

This category represents the events in which V1, the vehicle with the event monitoring system installed, is responsible for the crash/near-crash event. Event type 26 (V1 almost rear-end V2) was the most frequent type of event. Due to different operating characteristics near-rear end events may have different causalities on urban roads, and highway/freeways, therefore, three different model developments were done: urban roads, highways and freeways, and freeways only.

5.1.2.1. Near rear-end events on urban roads

For this analysis event type 26 "V1 almost rear-end V2 on urban roads" was considered the dependent variable. A total of 90 events were included by choosing this type of event. Table 5.2 shows the best model specifications, the estimation results of Stepwise Linear Regressions: Near Rear-end Events on Urban Roads.

Table 5.2 Estimation results of Stepwise Linear Regressions: Near Rear-end Events on Urban Roads

Dependent: Group 2 = URBAN ROADS	Model 1		Model 2		Model 3	
Variables in the Model	Estimate	t-statistic	Estimate	t-statistic	Estimate	t-statistic
Constant	0.650	4.625	0.650	4.588	0.649	4.555
Traffic Control: Free	-0.518	-6.036	-0.518	-6.000	-0.516	-5.850
Parking: No	-0.294	-2.272	-0.296	-2.252	-0.293	-2.176
Lane: #2	0.165	1.944	0.165	1.934	0.164	1.913
Lighting Conditions: Sunlight Clear	0.225	2.676	0.225	2.660	0.226	2.645
Pedestrian: Yes	0.363	2.371	0.362	2.344	0.362	2.331
Day of the week: Sunday	-0.307	-2.299	.-307	-2.284	-0.307	-2.272
Distraction: YES			0.008	0.085	0.008	0.081
Autobraking: YES					-0.011	-0.118
No. of Observations	90		90		90	
F	13.282		11.249		9.727	
Adjusted R^2	0.453		0.446		0.440	

Model 1 has the highest adjusted R^2 of 0.453 and an F-value of 13.282. Model 1 shows that the variable traffic control=free has a negative coefficient which means that uninterrupted flow facilities reduce the risk of near rear-end events. Roads without parking lower the risk for this type of event. While watching the videos, we observed that sudden parking maneuvers can cause abrupt braking and consequently increase the risk for rear-ends. The second lane from the edge of the road is typically the faster lane and it comes with a higher risk for near rear-end events. The model included the sunlight/clear lighting condition as a variable that increases the risk of this type of event. In the 100-car study, it was also found that most of rear-ends and near rear-ends events occurred under sunlight clear lighting conditions on all types of roads [32]. Presumably these conditions are more suitable for carefree and higher speed driving compared to driving in more adverse conditions which may evoke more caution. The presence of pedestrians also has a positive coefficient, indicating that it increases the risk of near rear-end events. Model 1 also includes the variable "day of the week: Sunday" as significant with a negative coefficient, which means that the risk of near rear-end event is lower on Sundays.

We decided to augment Model 1 by including the variables distraction and auto-braking system. Model 2 produced an adjusted R^2 of 0.446, which is smaller than Model 1; it includes the distraction variable, but its effects are minimal and not significant. Its positive coefficient denoting increased risk is intuitive. Model 3 added the variable that represents vehicles with auto-braking system available. The auto-braking system effects are minimal and not significant. Its negative coefficient denoting decreased risk is intuitive.

5.1.2.2 Near rear-end events on highways and freeways

Table 5.3 shows the three best model specifications: Model 1 includes 8 and 10 lane roads and has an adjusted R^2 of 0.248 and F-value of 15.705. Model 2 also includes 6 lane roads and has a better adjusted R^2 of 0.297. We developed Model 3 by adding mobile usage, with little statistical improvement.

Table 5.3 Estimation Results of Stepwise Linear Regressions: Near Rear-End Events on Highways and Freeways

Dependent: Group 1 = HIGHWAYS AND FREEWAYS	Model 1		Model 2		Model 3	
Variables in the Model	Estimate	t-statistic	Estimate	t-statistic	Estimate	t-statistic
Constant	0.426	7.794	0.325	4.975	0.325	4.972
Total Number of Lanes=6	-	-	0.294	2.641	0.275	2.429
Total Number of Lanes=8	0.526	4.869	0.627	5.635	0.675	4.216
Total Number of Lanes=10	0.574	3.573	0.675	4.218	0.627	5.631
Distraction: Mobile Usage					0.400	0.944
No. of Observations	90		90		90	
F	15.705		13.514		10.346	
Adjusted R^2	0.248		0.297		0.296	

We observe that the wider the road, the higher the risk for near rear-end events is on highways and freeways. The use of mobile devices comes with an intuitive positive coefficient that represents an increasing risk for rear-end events on highways and freeways, however, its effect is not statistically significant. More on this in the next section.

5.1.2.3. Near rear-end events on freeways

This part includes only near rear end events on freeways. The best model specification is shown in Table 5.4, the estimation results of stepwise linear regressions: Near rear-end events on freeways.

Table 5.4 Estimation Results of Stepwise Linear Regressions: Near Rear-End Events on Freeways

Dependent: TYPE OF ROAD =FREEWAYS	Model 1	
Variables in the Model	Estimate	t-statistic
Constant	0.511	7.505
Traffic Conditions: Light	-0.511	-2.525
Distraction: Mobile Usage	0.364	2.042
No. of Observations	60	
F	6.305	
Adjusted R^2	0.144	

The model includes only two independent variables as the most significant: light traffic condition as the main variable, with a negative coefficient, which represents that light traffic density significantly reduces the risk of rear-end events on freeways. It also shows that mobile usage has a positive and significant coefficient increasing the risk of highway rear-end events.

It is interesting to compare our results with a study by Davis et al., [37] who used an empirical model combined with a mechanistic model to understand the relationship between traffic density and rear-end crash risk. Both their study and ours, while based on completely different methodologies, found a significant statistical relationship between traffic conditions and rear-end events on freeways.

5.1.3 Vehicle 2 Events

In this category, most of the Vehicle 2 types of events were related to lane changing maneuvers, as “V2 cut in front of V1” and “V2 lane changing or weaving”.

5.1.3.1 Lane changing events on freeways

For this analysis, a new variable was created to group all lane-changing events on freeways. A total of 60 events were analyzed by this selection and Table 5.5 shows the best model specification, the estimation results of stepwise linear regressions: Lane-changing events on freeways.

Table 5.5. Estimation Results of Stepwise Linear Regressions: Lane-Changing Events on Freeways

Dependent: LANE CHANGING=1	Model 1	
Variables in the Model	Estimate	t-statistic
Constant	0.420	0.544
Distraction: Mobile Usage	0.958	2.500
Auto braking System = NO	0.236	2.386
No. of Observations	60	
F	5.171	
Adjusted R^2	0.122	

Model 1 presented in the Table 5.5 has an adjusted R^2 value of 0.122 and an F-value of 5.171. The stepwise linear regression selected two variables that appear to impact the risk of lane-changing events the most, and both have intuitive contributions. The model shows that distraction related to the use of mobile devices increases the risk of lane-chancing near-crash events. Vehicles without an auto-braking system installed have a higher risk of lane-changing near-crash events on a freeway. Note that auto braking systems also include warnings of occupied adjacent lanes to help the driver avoid erroneous lane changing maneuvers.

CHAPTER 6. CONCLUSIONS AND FUTURE DIRECTIONS

This study collected naturalistic driving data derived from a collaboration between the University of Hawaii of Manoa (UHM) and Charley's Taxi and Limousine (CTL). Dashboard cameras and sensors were installed in 233 taxi vans on Oahu, Hawaii which produced several hours of events classified as naturalistic driving data (NDD) in a period of seven months between fall 2019 and spring 2020. The data collection was halted by the shutdown due to the Covid-19 pandemic. The main goals of this study were to develop a statistical database from the NDD by coding near-crash events, and then identify factors that relate to the near-crash/crash events.

Several studies done previously in different parts of the world have shown that through NDD, it is possible to analyze near-crash events to understand crash risk factors. This is important because near-crash events are numerous, whereas crash numbers are low. This study developed a database with a total of 402 harsh events, of which were 398 near-crashes and four were crashes. Several variables such as road characteristics, environmental characteristics, driver characteristics and vehicle characteristics were coded for each event. The following objectives were set to achieve the study purpose: (1) collect data from NDD events where driving maneuvers caused an acceleration of 0.5g or higher; (2) develop a database suitable for statistical analysis; (3) derive basic statistics for all variables; (4) investigate correlations between variables; and (5) further investigate correlations (which may represent causality effects) for the most frequent types of events, using stepwise linear regression models.

The basic analysis, presented on Chapter 4, shows that several variables that were meaningful in previous studies were not considered relevant in our study for several reasons. The main findings of this study may be summarized as follows:

- Nearly 18% of events occurred on Thursdays and only 12% occurred on Sundays.
- Only 17.2% of the events occurred on a curvy road segment, while 82.8% occurred on a straight road segment.
- About 7.5% of the total events occurred on a construction zone or a blocked lane.
- About 31% of the events occurred on a road segment with no traffic control and about 16% occurred on a freeway.
- Signal traffic control was present in about 40% of the events.
- About 47% of the events occurred on roads with medium traffic congestion, and 29% under heavy traffic congestion.
- Pavement surface was wet in 10% of the events.
- Lighting conditions were as follows: 13% cloudy, 16% dark, 63% sunlight/clear, 6% glare/sunset/sunrise, and 1% tunnel.
- Pedestrians were present in 24% of the events.
- The auto braking did not seem to have a significant impact in event occurrence.
- We tested whether there was a correlation between the human perception of severity represented by the variable level of perceived impact and the machine severity level of severity represented by the G-force recorded by the sensors installed in the vehicles. A positive correlation with R2 of about 0.4 was found for near rear-end events on freeways.
- The most frequent type of event was V1 almost rear-ended V2, followed by lane changing related events, and events with pedestrians.
- Near rear-end events on freeways are common. The most influencing factors are light traffic (negative effect) and mobile phone use (positive effect).

- ❖ Lane changing events on freeways are strongly affected by mobile phone use and absence of automated driver aids in the vehicle.

A more detailed look on pedestrian, rear end and lane changing types of events using stepwise linear regressions provided additional detailed insights as follows:

- Uninterrupted flow facilities reduce the risk of near rear-end events.
- Wider expressways come with a higher risk for near rear-end events.
- Roads without parking lower the risk of near rear-end events.
- The risk of near rear-end event is lower on Sundays.
- Light traffic density significantly reduces the risk of rear-end events on freeways.
- Cellphone usage has a positive and significant coefficient increasing the risk of highway rear-end events.
- Cellphone usage increases the risk of lane-chancing near-crash events.
- Vehicles without an auto-braking system installed have a higher risk of lane-changing near-crash events on a freeway. Note that auto braking systems also include warnings of occupied adjacent lanes to help the driver avoid erroneous lane changing maneuvers.
- The presence of pedestrian crossings significantly increases the crash and near-crash events with pedestrians.

The installation of Samsara by the CTL company proved to be a successful tool for coaching drivers and the company proceeded with the installation of a different, and more advanced system in 2022.

This study was interrupted by the Covid-19 pandemic presenting some time limitations, there were also limitations related to the data sample, as this study recorded drivers of a taxicab company. For future research, it is recommended a longer data collection period and partnership with different types of transportation companies. It is also recommended to study non-professional drivers, as they represent most of the drivers in the real world.

It is also recommended for future studies to create a pre-event analysis focusing in the driver’s behavior prior the harsh event. This analysis should include the reaction time and speed data related to the moment when the driver realized the risk. It is also suggested a more detailed analysis of driver’s distraction inputted by human, as in this study the distraction was automatically recorded by the facial recognition system and did not provide clear information about activities distracting the driver.

REFERENCES

[1] National Center for Statistics and Analysis. Preview of motor vehicle traffic fatalities in 2019 (Research Note. Report No. DOT HS 813 021). National Highway Traffic Safety Administration, 2020.

[2] Bärgman, J. Methods for Analysis of Naturalistic Driving Data in Driver Behavior Research, 2016.

[3] Guo, Feng & Klauer, Sheila & Hankey, Jon & Dingus, Thomas. Near Crashes as Crash Surrogate for Naturalistic Driving Studies. Transportation Research Record: Journal of the Transportation Research Board. 2147. 66-74. 10.3141/2147-09, 2010.

[4] Fleet Dashcam: Dual-Facing GPS Dashcams for Drivers. (n.d.). Retrieved May 08, 2021, from https://www.azuga.com/dashcam, 2021.

[5] Commercial Vehicle Dash Cam - Fleet Camera Systems. Video Telematics and Fleet Management Solutions. Lytx. https://www.lytx.com/en-us/fleet-management/dash-cam, 2021.

[6] Blind Spot Detection - Protecting Vulnerable Road Users | Blog. https://www.mobileye.com/us/fleets/blog/blind-spot-detection-protect-vulnerable-road-users/, 2018.

[7] Bhagavath, S. Netradyne Driver Fleet Safety and Coaching Platform. https://www.netradyne.com/driveri/, 2021.

[8] Samsara, The Next Leap in Fleet Safety: Announcing AI Dash Cams, https://www.samsara.com/blog/the-next-leap-in-fleet-safety-announcing-ai-dash-cams#, February 2019.

[9] Vehicle Tracking Systems. https://www.teletracnavman.com/solutions/solutions-by-role/fleet-manager/vehicle-tracking-systems, 2021.

[10] Tarko, Andrew, Measuring Road Safety Using Surrogate Evens, Elsevier, 250 pp., 2020.

[11] Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J., Perez, M. A., Hankey, J., Ramsey, D., Gupta, S., Bucher, C., Doerzaph, Z. R., Jermeland, J., and Knipling, R.R. The 100-Car Naturalistic Driving Study: Phase II – Results of the 100-Car Field Experiment. (Interim Project Report for DTNH22-00-C-07007, Task Order 6; Report No. TBD). Washington, D.C. National Highway Traffic Safety Administration, 2006.

[12] Ingrid van Schagen, Fridulv Sagberg, The Potential Benefits of Naturalistic Driving for Road Safety Research: Theoretical and Empirical Considerations and Challenges for the Future, Procedia - Social and Behavioral Sciences, Volume 48, Pages 692-701, ISSN 1877-0428, https://doi.org/10.1016/j.sbspro.2012.06.1047, 2012.

[13] Sheila G. Klauer, Miguel Perez, Julie McClafferty, Chapter 6 - Naturalistic Driving Studies and Data Coding and Analysis Techniques, Editor(s): Bryan E. Porter, Handbook of Traffic Psychology, Academic Press, Pages 73-85, ISBN 9780123819840,https://doi.org/10.1016/B978-0-12-381984-0.10006-2, 2011.

[14] Backer-Grøndahl, A., Phillips, R., Sagberg, F., Touliou, K., Gatscha, M. Naturalistic Driving Observation: Topics and Applications of Previous and Current Naturalistic Studies. PROLOGUE Deliverable D1.1. Institute of Transport Economics, Oslo, Norway, 2009.

[15] Jeffrey S. Hickman, Richard J. Hanowski. Use of a Video Monitoring Approach To Reduce At-Risk Driving Behaviors In Commercial Vehicle Operations, Transportation Research Part F: Traffic Psychology and Behaviour, Volume 14, Issue 3, 2011, Pages 189-198, ISSN 1369-8478, https://doi.org/10.1016/j.trf.2010.11.010, 2011.

[16] Jennifer L. Bell, Matthew A. Taylor, Guang-Xiang Chen, Rachel D. Kirk, Erin R. Leatherman, Evaluation Of An In-Vehicle Monitoring System (IVMS) to Reduce Risky Driving Behaviors in Commercial Drivers: Comparison of In-Cab Warning Lights and Supervisory Coaching with Videos of Driving Behavior, Journal of

Safety Research, Volume 60, Pages 125-136, ISSN 0022-4375, https://doi.org/10.1016/j.jsr.2016.12.008, 2017.

[17] Jimiama Mafeni Mase, Shazmin Majid, Mohammad Mesgarpour, Mercedes Torres Torres, Grazziela P. Figueredo, Peter Chapman, Evaluating the Impact of Heavy Goods Vehicle Driver Monitoring and Coaching to Reduce Risky Behaviour, Accident Analysis & Prevention, Volume 146, 105754, ISSN 0001-4575, https://doi.org/10.1016/j.aap.2020.105754, 2020.

[18] Cohen, E., More Than Just Cameras: Video-Based Onboard Monitoring Systems for Fleet Safety, TR News 326, http://www.trb.org/Main/Blurbs/180625.aspx, TRB, Washington, D.C., 2020.

[19] Dingus, T. DRIVE-SMART Driver Monitoring and Crash Risk Mitigation System, Final Report for Safety IDEA Project 18, http://onlinepubs.trb.org/Onlinepubs/IDEA/FinalReports/Safety/S18FinalReport.pdf, TRB, Washington, D.C., 2013.

[20] Fitch G.M., Hanowski R.J. Using Naturalistic Driving Research to Design, Test and Evaluate Driver Assistance Systems. In: Eskandarian A. (eds) Handbook of Intelligent Vehicles. Springer, London. https://doi.org/10.1007/978-0-85729-085-4_21, 2012.

[21] Hankey, J., Perez, M.A., & McClafferty, J.A. Description of the SHRP 2 Naturalistic Database and the Crash, Near-Crash, and Baseline Data Sets, 2016.

[22] Blatt, Alan & Pierowicz, John & Flanigan, Marie & Lin, Pei-Sung & Kourtellis, Achilleas & Lee, Chanyoung & Jovanis, Paul & Jenness, James & Wilaby, Martha & Campbell, John & Richard, Christian & Good, David & Czar, Nora & Hoover, Michelle. Naturalistic Driving Study: Field Data Collection, 2015.

[23] Haiyan Hao, Yingfeng (Eric) Li, Alejandra Medina, Ronald B. Gibbons, Linbing Wang. Understanding Crashes Involving Roadway Objects with SHRP 2 Naturalistic Driving Study Data, Journal of Safety Research, Volume 73, Pages 199-209, ISSN 0022-4375, https://doi.org/10.1016/j.jsr.2020.03.005, 2020.

[24] Arai, Y., Nishimoto, T., Ezaka, Y., & Yoshimoto, K. Accidents and Near-Misses Analysis by Using Video Drive-Recorders in a Fleet Test, 2001.

[25] Cheng, Bo, Q. Lin, T. Song, Y. Cui, L. Wang, and S. Kuzumaki. Analysis of Driver Brake Operation in Near-Crash Situation Using Naturalistic Driving Data. International Journal of Automotive Engineering. Volume 2, Issue 4, pp. 87-94, https://doi.org/10.20485/jsaeijae.2.4_87, 2011.

[26] Lin, Qingfeng, Cheng, Bo and Lu, Guang. Analysis of Characteristics of Vehicle-Bicycle/Pedestrian Conflicts Using Video Drive Recorder. Advanced Materials Research. 243-249. 4413-4417. 10.4028/www.scientific.net/AMR.243-249.4413, 2011.

[27] Abbas Sheykhfard, Farshidreza Haghighi, Eleonora Papadimitriou, Pieter Van Gelder, Analysis of the Occurrence and Severity of Vehicle-Pedestrian Conflicts in Marked and Unmarked Crosswalks Through Naturalistic Driving Study, Transportation Research Part F: Traffic Psychology and Behaviour, Volume 76, Pages 178-192, ISSN 1369-8478, https://doi.org/10.1016/j.trf.2020.11.008, 2021.

[28] Lerner N, Jenness J, Singer J, Klauer S, Lee S, Donath M, Manser M, Ward N. An Exploration of Vehicle-Based Monitoring of Novice Teen Drivers: Final Report. Washington, DC. NHTSA, U.S. DOT. DOT HS 811 333; 2010.

[29] Lv, Bin, Yue, Rui and Zhang, Yongsheng. The Influence of Different Factors on Right-Turn Distracted Driving Behavior at Intersections Using Naturalistic Driving Study Data. IEEE Access. PP. 1-1. 10.1109/ACCESS.2019.2942841, 2019.

[30] Simons-Morton, B. G., Klauer, S.G., Ouimet, M.C., Guo, F., Albert, P.S., Lee, S.E., Ehsani, J.P., Pradhan, A.K, & Dingus T.A. Naturalistic Teenage Driving Study: Findings and Lessons Learned. Journal of Safety Research.

Vol. 54, pp. 29-44, 2015.

[31] Klauer, S.G., Dingus, T., Neale, V., Sudweeks, J., and Ramsey, D.J. The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data, 2006.

[32] Lee, S. E., Llaneras, E., Klauer, S.G., and Sudweeks, J. Analyses of Rear-End Crashes and Near-Crashes in the 100-Car Naturalistic Driving Study to Support Rear-Signaling Countermeasure Development. (Report number DOT HS 810 846). Washington, DC: U.S. Department of Transportation, 2007.

[33] McLaughlin, S.B., Hankey, J.M., Klauer, S.G., and Dingus, T. A. Contributing Factors to Run-Off Road Crashes and Near-Crashes. Report No. DOT, HS 811 079. Washington, DC. National Highway Traffic Safety Administration, 2009.

[34] Fitch, G. Lee, S.E. Klauer, S.G., Hankey, J.M., Sudweeks, J., and Dingus, T.A. Analysis of Lane Change Crash and Near-Crash. (Report No. DOT HS 811 147) Washington, DC: National Highway Traffic Safety Administration, 2009.

[35] Das, Anik and Ghasemzadeh, Ali. Analyzing The Effect of Fog Weather Conditions on Driver Lane-Keeping Performance Using the SHRP2 Naturalistic Driving Study Data. Journal of Safety Research. 68. 10.1016/j.jsr.2018.12.015, 2018.

[36] Alshatti, D. Examining Driver Risk Factors in Road Departure Conflicts Using SHRP2 Data. (Electronic Thesis or Dissertation). https://etd.ohiolink.edu/, 2018.

[37] Davis, G. A., Chatterjee, I., Gao, J., and Hourdos, J. Traffic Density versus Rear-End Crash Risk on Freeways: Empirical Model, Mechanism Model, and Transfer to Automated Vehicles. Journal of Transportation Engineering Part A: Systems, 147(4), [04021007]. https://doi.org/10.1061/JTEPBS.0000501, 2021.

APPENDIX A. CHI-SQUARE TESTS

LIGHTING CONDITIONS

Chi-Square Significance		
Tested Variables	Lighting Conditions	
Variables	Hwy+Fwy	Urban Roads
Day of Week	0.691	0.761
Time	0.171	0.105
Driver's Gender	0.050	0.345
Vehicle 1 Movement	0.000	0.412
Involves Pedestrian	0.962	0.026
V1 has Auto Braking System	0.748	0.267
Perceived Impact or Wow Factor	0.612	0.763
Construction, Work Zone or Blocked Lane	0.431	0.410
Type of Road	0.267	0.497
Road Alignment	0.114	0.866
Pavement Quality	0.517	0.038
Parking	0.761	0.293
Traffic Control	0.690	0.693
Traffic Conditions	0.001	0.002
Rural road	0.748	0.794
Lighting Conditions	-	-
Weather	0.194	0.045
Pavement Surface	0.001	0.000

Lighting Conditions by Driver's Gender

Count									
Group			Lighting Conditions					Total	Pearson Chi-Square
			CLOUDY	DARK	SUNLIGHT CLEAR	SUNSET/ SUNRISE/ GLARE	TUNNEL		
HWY+ FWY	Gender	F	2	1	9	2	2	16	0.05
		M	16	16	75	11	1	119	
	Total		18	17	84	13	3	135	
URBAN ROADS	Gender	F	3	2	20	2	1	28	0.345
		M	33	44	150	10	2	239	
	Total		36	46	170	12	3	267	

Lighting Conditions by Vehicle's 1 Movement

Group			CLOUDY	DARK	SUNLIGHT CLEAR	SUNSET/ SUNRISE/ GLARE	TUNNEL	Total	Pearson Chi-Square
			Count						
			Lighting Conditions						
HWY+ FWY	Vehicle 1 Movement	LT	1	0	3	0	2	6	0.000173
		RT	0	1	3	0	0	4	
		TH	17	16	78	13	1	125	
	Total		18	17	84	13	3	135	
URBAN ROADS	Vehicle 1 Movement	LT	3	7	13	0	1	24	0.411802
		RT	4	4	9	1	0	18	
		TH	29	35	148	11	2	225	
	Total		36	46	170	12	3	267	

Lighting Conditions by Event Involves Pedestrian

Group			CLOUDY	DARK	SUNLIGHT CLEAR	SUNSET/ SUNRISE/ GLARE	TUNNEL	Total	Pearson Chi-Square
			Count						
			Lighting Conditions						
HWY+ FWY	PEDESTRIAN	NO	18	17	83	13	3	134	0.962
		YES	0	0	1	0	0	1	
	Total		18	17	84	13	3	135	
URBAN ROADS	PEDESTRIAN	NO	25	23	117	4	2	171	0.026
		YES	11	23	53	8	1	96	
	Total		36	46	170	12	3	267	

Lighting Conditions by Pavement Quality

Group			CLOUDY	DARK	SUNLIGHT CLEAR	SUNSET/ SUNRISE/ GLARE	TUNNEL	Total	Pearson Chi-Square
			Count						
			Lighting Conditions						
HWY+ FWY	Pavement Quality	GOOD	9	9	36	7	2	63	0.521
		ROUGH	0	0	10	2	0	12	
		ROUGH SPOTS	4	2	22	1	1	30	
		VERY GOOD	5	6	16	3	0	30	
	Total		18	17	84	13	3	135	
URBAN ROADS	Pavement Quality	GOOD	22	30	100	5	1	158	0.038
		ROUGH	1	0	12	0	0	13	
		ROUGH SPOTS	6	5	42	6	1	60	
		VERY GOOD	7	11	16	1	1	36	
	Total		36	46	170	12	3	267	

Lighting Conditions by Traffic Conditions

Group			Count: Lighting Conditions: CLOUDY	DARK	SUNLIGHT CLEAR	SUNSET/ SUNRISE/ GLARE	TUNNEL	Total	Pearson Chi-Square
HWY+ FWY	Traffic Conditions	HEAVY	7	10	27	5	0	49	0.001
		LIGHT	1	4	11	1	3	20	
		MEDIUM	9	3	43	4	0	59	
		VERY HEAVY	1	0	2	3	0	6	
		VERY LIGHT	0	0	1	0	0	1	
	Total		18	17	84	13	3	135	
URBAN ROADS	Traffic Conditions	HEAVY	11	9	45	2	0	67	0.002
		LIGHT	13	21	26	1	1	62	
		MEDIUM	11	16	92	9	2	130	
		VERY HEAVY	1	0	7	0	0	8	
	Total		36	46	170	12	3	267	

Lighting Conditions by Weather

Group			Count: Lighting Conditions: CLOUDY	DARK	SUNLIGHT CLEAR	SUNSET/ SUNRISE/ GLARE	TUNNEL	Total	Pearson Chi-Square
HWY+ FWY	Weather	CLEAR	16	16	83	13	3	131	0.194
		LIGHT RAIN	2	1	1	0	0	4	
	Total		18	17	84	13	3	135	
URBAN ROADS	Weather	CLEAR	33	42	168	12	3	258	0.045
		LIGHT RAIN	3	4	2	0	0	9	
	Total		36	46	170	12	3	267	

Lighting Conditions by Pavement Surface

Group			Count: Lighting Conditions: CLOUDY	DARK	SUNLIGHT CLEAR	SUNSET/ SUNRISE/ GLARE	TUNNEL	Total	Pearson Chi-Square
HWY+ FWY	Pavement Surface	DRY	12	15	82	12	3	124	0.001
		WET	6	2	2	1	0	11	
	Total		18	17	84	13	3	135	
URBAN ROADS	Pavement Surface	DRY	21	39	163	12	3	238	0.000
		WET	15	7	7	0	0	29	
	Total		36	46	170	12	3	267	

PEDESTRIAN

All

Tested Variables	Pedestrian	
Variables	Hwy+Fwy	Urban Roads
Day of Week	-	0.293
Time	-	0.799
Driver's Gender	-	0.012
Vehicle 1 Movement	-	0.933
Involves Pedestrian	-	-
V1 has Auto Braking System	-	0.511
Perceived Impact or Wow! Factor	-	0.000
Construction, Work Zone or Blocked Lane	-	0.210
Type of Road	-	0.134
Road Alignment	-	0.712
Pavement Quality	-	0.517
Parking	-	0.936
Traffic Control	-	0.000
Traffic Conditions	-	0.070
Rural road	-	0.284
Lighting Conditions	-	0.026
Weather	-	0.382
Pavement	-	0.320
What Happened	-	0.000
G-Force	-	0.863

Pedestrian by V1 Driver Gender

Count						
Group			PEDESTRIAN		Total	Pearson Chi-Square
			NO	YES		
HWY+FWY	Gender	F	16	0	16	0.713
		M	118	1	119	
	Total		134	1	135	
URBAN ROADS	Gender	F	24	4	28	0.012
		M	147	92	239	
	Total		171	96	267	

Pedestrian by Traffic Control

Count						
Group			PEDESTRIAN		Total	Pearson Chi-Square
			NO	YES		
HWY+FWY	Traffic Control	FREE	56	1	57	0.848
		NONE	32	0	32	
		PED XING	1	0	1	
		SIGNAL	43	0	43	
		STOP SIGN	2	0	2	
	Total		134	1	135	
URBAN ROADS	Traffic Control	FREE	5	1	6	0.000
		NONE	71	23	94	
		PED XING	4	38	42	
		ROUNDABOUT	1	0	1	
		SIGNAL	85	34	119	
		SPEEDBUMP	1	0	1	
		STOP SIGN	4	0	4	
	Total		171	96	267	

AUTO BRAKING FEATURE

All

Tested Variable	AutoBraking	
Variables	Hwy+Fwy	Urban Roads
Day of Week	0.572	0.849
Time	0.458	0.502
Driver's Gender	0.317	0.383
Vehicle 1 Movement	0.700	0.047
Involves Pedestrian	0.184	0.511
V1 has Auto Braking System	-	-
Perceived Impact or Wow! Factor	0.260	0.091
Construction, Work Zone or Blocked Lane	0.136	0.553
Type of Road	0.304	0.075
Road Alignment	0.633	0.564
Pavement Quality	0.953	0.926
Parking	0.186	0.703
Traffic Control	0.184	0.013
Traffic Conditions	0.550	0.441
Rural road	0.996	0.410
Lighting Conditions	0.748	0.267
Weather	0.633	0.280
Pavement	0.996	0.326

Tested Variable	AutoBraking	
Variables	Hwy+Fwy	Urban Roads
What Happened	0.598	0.038
G-Force	0.411	0.293

Auto Braking by V1 Movement

Count						
Group			AUTO BRAKING SYSTEM		Total	Pearson Chi-Square
			NO	YES		
HWY+FWY	Vehicle 1 Movement	LT	3	3	6	0.700
		RT	3	1	4	
		TH	80	45	125	
	Total		86	49	135	
URBAN ROADS	Vehicle 1 Movement	LT	19	5	24	0.047
		RT	17	1	18	
		TH	155	70	225	
	Total		191	76	267	

Auto Braking by Traffic Control

Count						
Group			AUTO BRAKING SYSTEM		Total	Pearson Chi-Square
			NO	YES		
HWY+FWY	Traffic Control	FREE	36	21	57	0.184
		NONE	23	9	32	
		PED XING	0	1	1	
		SIGNAL	27	16	43	
		STOP SIGN	0	2	2	
	Total		86	49	135	
URBAN ROADS	Traffic Control	FREE	2	4	6	0.013
		NONE	58	36	94	
		PED XING	34	8	42	
		ROUNDABOUT	1	0	1	
		SIGNAL	93	26	119	
		SPEEDBUMP	0	1	1	
		STOP SIGN	3	1	4	
	Total		191	76	267	

TRAFFIC CONTROL

All

Tested Variables	Traffic Control	
Variables	Hwy+Fwy	Urban Roads
Day of Week	0.266	0.246
Time	0.154	0.477
Driver's Gender	0.702	0.802
Vehicle 1 Movement	0.100	0.719
Involves Pedestrian	0.848	0.000
V1 has Auto Braking System	0.184	0.013
Perceived Impact or Wow! Factor	0.827	0.491
Construction, Work Zone or Blocked Lane	0.458	0.746
Type of Road	0.000	0.014
Road Alignment	0.028	0.011
Pavement Quality	0.492	0.853
Parking	0.479	0.156
Traffic Control	-	-
Traffic Conditions	0.254	0.069
Rural road	0.869	0.000
Lighting Conditions	0.690	0.693
Weather	0.438	0.919
Pavement	0.973	0.921
What Happened	0.000	0.000
G-Force	0.542	0.967

Traffic Control by Road Alignment

Count												
Group			Traffic Control							Total	Pearson Chi-Square
			FREE	NONE	PED XING	ROUNDABOUT	SIGNAL	SPEEDBUMP	STOP SIGN		
HWY+FWY	ROAD ALIGMENT	CURVY	3	9	0	0	9	0	1	22	0.028
		STRAIGHT	54	23	1	0	34	0	1	113	
	Total		57	32	1	0	43	0	2	135	
URBAN ROADS	ROAD ALIGMENT	CURVY	4	14	6	0	22	1	0	47	0.011
		STRAIGHT	2	80	36	1	97	0	4	220	
	Total		6	94	42	1	119	1	4	267	

Traffic Control by Rural Road

Count												
Group			Traffic Control							Total	Pearson Chi-Square Significance
			FREE	NONE	PED XING	ROUNDABOUT	SIGNAL	SPEEDBUMP	STOP SIGN		
HWY+ FWY	RURAL	NO	53	28	1	0	40	0	2	124	0.869
		YES	4	4	0	0	3	0	0	11	
	Total		57	32	1	0	43	0	2	135	
URBAN ROADS	RURAL	NO	6	88	42	0	117	0	4	257	0.000
		YES	0	6	0	1	2	1	0	10	
	Total		6	94	42	1	119	1	4	267	

DAY OF THE WEEK

All

Tested Variables	Day of Week	
Variables	Hwy+Fwy	Urban Roads
Day of Week	-	-
Time	0.683	0.208
Driver's Gender	0.512	0.998
Vehicle 1 Movement	0.090	0.745
Involves Pedestrian	0.407	0.293
V1 has Auto Braking System	0.572	0.849
Perceived Impact or Wow! Factor	0.387	0.892
Construction, Work Zone or Blocked Lane	0.002	0.296
Type of Road	0.045	0.726
Road Alignment	0.688	0.519
Pavement Quality	0.839	0.591
Parking	0.363	0.549
Traffic Control	0.266	0.246
Traffic Conditions	0.155	0.386
Rural road	0.754	0.040
Lighting Conditions	0.691	0.761
Weather	0.083	0.000
Pavement	0.717	0.186
What Happened	0.354	0.179
G-Force	0.750	0.882

Day of the Week by Construction, Work Zone or Blocked Lane

Group			SUN	MON	TUE	WED	TH	FR	SAT	Total	Pearson Chi-Square
			Count								
			Day of Week								
HWY+ FWY	Construction, or other work zone or blocked lane	NO	14	15	13	23	15	11	25	116	0.002
		YES	0	0	6	3	2	7	1	19	
	Total		14	15	19	26	17	18	26	135	
URBAN ROADS	Construction, or other work zone or blocked lane	NO	33	34	35	30	52	37	35	256	0.296
		YES	1	2	1	4	2	0	1	11	
	Total		34	36	36	34	54	37	36	267	

Day of the Week by Rural Road

Group			SUN	MON	TUE	WED	TH	FRI	SAT	Total	Pearson Chi-Square
			Count								
			Day of Week								
HWY+ FWY	RURAL	NO	14	13	17	23	15	17	25	124	0.754
		YES	0	2	2	3	2	1	1	11	
	Total		14	15	19	26	17	18	26	135	
URBAN ROADS	RURAL	NO	33	32	34	31	54	37	36	257	0.040
		YES	1	4	2	3	0	0	0	10	
	Total		34	36	36	34	54	37	36	267	

Day of the Week by Weather

Group			SUN	MON	TUE	WED	THU	FRI	SAT	Total	Pearson Chi-Square
			Count								
			Day of Week								
HWY+ FWY	Weather	CLEAR	12	15	19	24	17	18	26	131	0.083
		LIGHT RAIN	2	0	0	2	0	0	0	4	
	Total		14	15	19	26	17	18	26	135	
URBAN ROADS	Weather	CLEAR	28	35	36	34	54	37	34	258	0.000
		LIGHT RAIN	6	1	0	0	0	0	2	9	
	Total		34	36	36	34	54	37	36	267	