Potential Crash Exposure Measures Based on GPS-Observed Driving Behavior Activity Metrics

A Dissertation
Presented to
The Academic Faculty

By

Jungwook Jun

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in Civil Engineering

The Georgia Institute of Technology
December 2006

Potential Crash Exposure Measures Based on GPS-Observed Driving Behavior Activity Metrics

Approved by:

Dr. Randall Guensler, Advisor
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Michael Meyer
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Michael Hunter
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Michael Rodgers
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Catherine Ross
School of City Planning
Georgia Institute of Technology

Dr. Jennifer Ogle
Department of Civil Engineering
Clemson University

Date Approved: November 20, 2006

ACKNOWLEDGMENTS

I would like to express my sincere gratitude and appreciation to the members of my doctoral committee, Dr. Randall Guensler, Dr. Jennifer Ogle, Dr. Michael Meyer, Dr. Michael Rodgers, Dr. Michael Hunter, and Dr. Catherine Ross, for their invaluable advice and direction, which enabled me to successfully finish my doctoral thesis. First of all, my deepest appreciation must go to my advisor, Dr. Randall Guensler, for his encouragement and advice. To me, not only has he been an intellectual scholar, but also a supportive mentor throughout my study. His insightful advice and encouragement have always inspired me to improve and develop my research potential and to move on toward the completion of my thesis. I also would like to thank Dr. Jennifer Ogle, for she has helped and guided me as both an academic committee member and a colleague with her excellent experience and abundant knowledge. She has taught me how to solve problems and how to progress research, which helped me improve my work. Finally, I am very thankful to the other members of my doctoral committee, Dr. Michael Meyer, Dr. Michael Rodgers, Dr. Michael Hunter, and Dr. Catherine Ross, for giving me invaluable advice and insightful comments.

I am also very grateful to current members of my research group, Dr. Hainan Li, Vetri Elango, Jaesup Lee, and Kai Zuehlke as well as previous colleagues, Dr. Seungju Yoon, Dr. Joonho Ko, Hongbiao Sun, Chunxia Feng, and Cristina Pastore, for their assistance and encouragement.

Finally, I am much indebted to my family. To my wife, Soyoung, for her encouragement, patience, love, and friendship. Without her, I could not have done this.

To my lovely children, Claire and Aiden, for their unconditional love and wonderful smiles. To my brother, Jangho, for his constant support. And last but not least, my father, mother, and mother-in-law, for their love and support, which have made me today.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ARRB	Australian Road Research Board
ABS	Anti-Lock Brake System
C/A	Clear/Acquisition
CSD	Circuit Switched Data
CART	Classification and Regression Tree Analysis
DGPS	Differential Global Positioning System
DTC	Diagnostics Trouble Codes
DLC	Data Link Connector
DMI	Distance Measuring Instrument
DMVS	The Department of Motor Vehicle Safety
ECT	Engine Coolant Temperature
ESN	Electronic Serial Number
ECM	Electronic Control Module
EEPROM	Electronically Erasable Programmable Read Only Memory
EGR	Exhaust Gas Recirculation
FTP	File Transfer Protocol
FHWA	The Federal Highway Administration
FAIR	Fast and Interwined Regular
GPS	Global Positioning System
GDOT	Georgia Department of Transportation
GT-TDC	Georgia Tech Trip Data Collector
HOT	High Occupancy Toll
IP	Internet Protocol
I/M program	Inspection and Maintenance Program
ICC	Intelligent Cruise Control System
KS Test	Kolmogorov-Smirnov Test
LDA	Linear Discriminant Analysis
LDV	Light Duty Vehicle

IAT	Intake Air Temperature
MAP	Manifold Air Pressure
MAE	Mean of the Absolute Errors
ME	Mean of the Errors
MAF	Mass Air Flow
MIL	Malfunction Indicator Light
MARTA	Metropolitan Atlanta Rapid Transit Authority
MANOVA	Multivariate Analysis of Variance
NHTSA	National Highway Traffic Safety Administration
OBD	Onboard Diagnostics
PCM	Powertrain Control Modules
PC	Pass Counter
PAYD	Pay-As-You-Drive Insurance Program
QDA	Quadratic Discriminant Analysis
RTK	Real Time Kinematic
RPM	Resolutions per Minute
SA	Selected Availability
SMS	Short Message System
SAE	Society of Automotive Engineers
SVM	Support Vector Machine
TDMA	Time Division Multiple Access
TDM	Transportation Demand Management
TPS	Throttle Position Sensor
UGA	The University of Georgia
VSS	Vehicle Speed Sensor

SUMMARY

Identifying and understanding the relationships between observed driving behavior for long-term period and corresponding crash involvement rates is paramount to enhancing safety improvement programs and providing useful insights for transportation safety engineers, policy makers, insurance industries, and the public.

Unlike previous data collection methods, recent advancement in mobile technology and accuracy of global positioning systems (GPS) allow researchers to monitor driving activities of large fleets of vehicles, for long-time study periods, at great detail. The GPS-measured travel data provide abundant reliable information, which can help identify relationships between various driving behavior activities and crash involvement rates (crash risks) under varying conditions of facility type and time of day. Coupling the detailed travel information with known individual driver, household, and vehicle characteristics, activities (operations) can then be tied back to a wide variety of socio-demographic parameters. Furthermore, GPS-measured data can evaluate how driving behavior patterns change during a trip in response to changes in roadway operating conditions. In this respect, to identify and substantiate driving behaviors linked with crash involvements is the challenge.

For this research, this study investigates the driving patterns of drivers who have (141 drivers) and who have not experienced crashes (26 drivers) during a 14-month study period using the longitudinally collected GPS data during a six-month Commute Atlanta study. This study allows an empirical investigation to assess whether drivers with recent crash experiences have exhibited different driving behavior activity patterns (travel

mileage, travel duration, speed, acceleration, speed stability pattern, frequency of unfamiliar roadway activities, frequency of turn movement activities, and previous crash location exposures). To analyze driving behavior activity patterns, this study also discuss various techniques of implementing GPS data streams in safety analyses.

As a result, this study found that drivers within the same demographic characteristics such as age and gender might not produce the same behavioral patterns and that driving behavior activities were more strongly related to crash involvement rates, albeit with small sample drivers. Of the numerous potential behavior activity exposure measures that were examined in this study, linear discriminant analysis (LDA) and logistic regression model provided that travel mileage, travel duration, speeding pattern, hard acceleration/deceleration activity, and previous crash location exposures were the most important parameters for classifying potential crash involvement rates (crash risks) of individual drivers. Finally, this study provides useful guidance for researchers who plan to evaluate the relationships between driving behavior activity patterns and crash risk with larger sample data and proposes driving behavior activity exposure metrics of individual drivers for possible safety surrogate measures as well as for driver training and education programs.

Chapter One

INTRODUCTION

Background

Motor vehicle traffic crash is one of the leading causes of death in the United States and in the world. Based upon a recent NHTSA report [1], deaths caused from motor vehicle crashes were ranked 3rd, behind only cancer and heart diseases, and about one out of every 50 deaths was caused by motor vehicle traffic crashes in 2002. The State of Georgia had found that on average, 30 people lost their life in crashes each week and about 2,555 people were injured by crashes [2]. Totally 1,610 people were killed (5 per day) and 132,879 people were injured (364 per day) on roadways [2]. Thus, safety in the field of transportation is one of the biggest issues that raises attention and awareness [3].

There are three major components of transportation environment: the roadway, the vehicle, and the driver [4] (Figure 1), which are strongly related with each other. However, efforts to decrease the number of motor vehicle crashes and fatality rate have been intensively made only to roadway and vehicle designs. For example, engineers have mainly focused on improving roadway designs and implementing safety devices on roadways [5]. Vehicle manufacturers have also developed various safety devices such as anti-lock brake system (ABS), intelligent cruise control (ICC) system, and front-side air bag systems [5]. However, relatively few efforts have been made for the aspect of drivers.

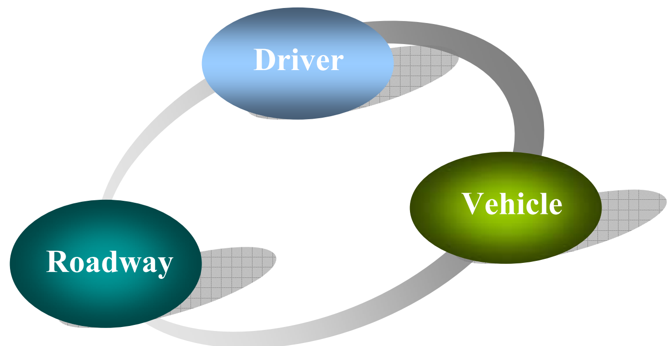


Figure 1: Three Major Components of Transportation Environment

In fact, those efforts regarding improvements of roadway designs and development of vehicle safety devices definitely help increase transportation safety, but parallel changes in driver behavior should be met since effectiveness of those improvements depends on how drivers travel those improved roadways and how they utilize vehicles equipped with better safety devices. In other words, the success of safety strategies focusing on roadways and vehicles is determined by drivers since the increases in the frequency of unsafe behavior potentially lead higher crash rates [5]. Thus, to maximize the effects of the current safety strategies, driver education and re-training program should be conducted at the same time and should be implemented with detailed information on knowledge about driving characteristics and activity patterns of crash-involved drivers.

However, driver behavior activity patterns are difficult to be evaluated due to the lack of reliable behavioral data and the absence of effective data collection techniques. Previous studies used only simple measures such as the total travel mileage or speeds observed at fixed roadway segments, so they did little explain about detailed relationships between crash involvement rate and various driver behavior activities such as disaggregated mileage, speed, and acceleration behavior representing when, where, and how drivers traveled. Depending on travel time (day or night), roadway types (freeway, arterial, or local roadway), regional area (urban or rural), or traffic conditions (congested or un-congested and sunny or rainy), drivers can be exposed to different crash involvement rates [4, 6].

Some previous studies, in fact, tried to identify relationships between driving behavior and crash risk and to examine which of driving behavior significantly differ between drivers who were involved and not involved in motor crashes with temporally or spatially limited sample data. As a result, those studies generally found simple relationships indicating that drivers traveling at higher speeds and longer distance had higher crash rates and those drivers also frequently produced speed changes such as acceleration and deceleration patterns [7-11]. However, they did not provide any detailed relationships (when, where, and how) between observed driving behavior activities and crash involvement rates [4] and did not explain that those behavioral activities could differ among drivers, even within the same driver groups in term of gender and age. Furthermore, due to the difficulty of data collection process and poor data quality, those previous studies did not provide better and more equitable classification measures that can effectively cluster drivers with potentially high crash involvement rates to improve current automobile insurance programs.

A few academic research groups and insurance companies are currently studying the feasibility of insurance program using behavioral exposure metrics, such as a Pay-As-You-Drive (PAYD) insurance pilot program [12, 13]. The PAYD insurance pilot program is one of the first attempts to integrate individual driver's behavioral exposures such as total travel mileage and speed profile during the specific period such as nighttime into the insurance classification decision rules where higher crash exposure means higher chances of being involved in a crash and yields higher premium costs. Although those current PAYD programs seem reasonable and equitable, this program also has limitations in classifying drivers who have high crash involvement rates since this program adopted the simple relationship resulted from previous studies that did not use detailed and accurate behavior activity data. Drivers having the same total mileage may have different potential crash involvement rates since probability of crash involvements may be different based on when, where, how, and under what conditions drivers traveled [4].

Young or older drivers were likely to travel much less (especially on freeways and major arterials) than middle-age drivers[14]. Thus, other redundant behavioral exposure measures (where drivers drive, when they drive, and under what environmental conditions) that effectively assess crash involvement rates of individual drivers need to be evaluated.

Research Objectives

Identifying and understanding the relationships between observed driving behavior activities over long-term periods and corresponding crash involvement rates during the same period will enhance safety improvement programs and provide useful insights for transportation safety engineers, policy makers, insurance industries, and the public.

- Transportation engineers and policy makers can identify relationships between driver behavior activities and crash involvement rates to improve driver training and education programs designed to modify risky driving habits.
- Speeding (or hard acceleration/deceleration activities) is a well-known risk-related driving behavior as it is highly correlated with the crash involvements as well as the severity of crashes. Roadway designs or supplement safety devices can be implemented on roadways where high speeding events (or hard acceleration/deceleration activities) frequently occur to reduce the number and the severity of crashes as a surrogate safety measure.
- Insurance companies may improve their insurance premium structure by enhancing existing pricing frameworks with more refined information on crash-related driving behavior (potential crash involvement rates are not efficiently priced in current insurance premium structures).
- Drivers can also choose to reduce their crash-related driving behavior activities such as high speeding and hard deceleration events and change their travel patterns.

Unlike previous data collection methods, recent advances in mobile technology and accuracy of global positioning systems (GPS) allow researchers to monitor driving activities of large fleets of vehicles, over long-time study periods, at great detail. The GPS-measured travel data provide abundant reliable information that help identify relationships between various driving behavior activities and crash involvement rates under varying conditions of facility type and time of day. Coupling the detailed travel information with known individual drivers, household, and vehicle characteristics, activities (operations) can then be tied back to a wide variety of socio-demographic parameters. Furthermore, GPS-measured data can be used to evaluate how driving behaviors change during a trip in response to changes in roadway operating conditions. In this respect, to identify and substantiate driving behaviors that linked with crash involvements is the challenge.

Of various crash-related driving behavior activities, speeding is currently considered the most significant factor in crash involvement [5, 8, 15]. However, traffic speeds depend not only on the driver, but also on the road conditions, traffic density, and time of day. During peak travel conditions, drivers may be in congested traffic and therefore forced to travel at slower speeds due to congestion. Only in free flow conditions can the driver choose his or her own speed. Thus, differences in speeding behavior between drivers with and without crash involvements may not be reliable using only data collected during the peak travel conditions. Therefore, the type of roadway and time of day must be investigated in regards to the speeds at which drivers travel. In addition to speeding behavior, other potential driving behavior activity measures need to be examined. Although speeding behavior may not provide any differences between drivers with and without recent crash experiences during congested periods, acceleration patterns may be different. Aggressive driving and tailgating behavior may be indicated by hard acceleration/deceleration activities.

This study investigates the driving patterns of drivers who have and who have not experienced crashes during a 14-month study period using the longitudinally collected GPS data during a six-month Commute Atlanta study. This study allows an empirical investigation to assess whether drivers with recent crash experiences exhibit different driving behavior activity patterns (travel mileage, travel duration, speed, acceleration, speed stability patterns, frequency of unfamiliar roadway activities, frequency of turn movement activities, and previous crash location exposures).

To utilize those driving behavior activity patterns, this study discusses various techniques of implementing GPS data streams in safety analyses. Through the investigation of various driving behavior activity metrics, this study finally provides useful guidance for researchers who plan to evaluate the relationships between driving behavior activity patterns and crash involvement with larger sample data and proposes behavior activity exposure metrics of individual drivers for possible safety surrogate measures as well as for driver training and education programs.

Research Process and Analytical Methodology

Data Collection

As one component of the Commuter Choice and Value Pricing Insurance Incentive Program (the Commute Atlanta study), a safety-related survey[1] was conducted in November 2004 to obtain information on crash involvement status during the 14-month study period (from September 2003 to November 2004) from program participants [5].

From the self-reported survey, this study was able to categorize drivers into two groups based on their crash involvement status over the study period. Among the 234 drivers of all ages who had instrumented-vehicles and returned the survey, drivers that shared a vehicle with another household member more than 10% of the time were

[1] This study discusses the safety-related survey in Chapter 3.

excluded because their personal driving trip data could not be adequately distinguished from those of other household members.

In addition, some of participants had replaced their vehicles during the study period and several Georgia Tech Trip Data Recorders (GT-TDCs) had to be reinstalled, and those drivers could not be used for this study. After the data cleaning process, this study obtained 167 drivers of all ages who had been continuously monitored through a whole 6-month period (January through June 2004) for which survey data were available. The GIS-based Georgia crash database containing crash locations occurred between 2000 and 2002 were also obtained from the Georgia Department of Transportation (GDOT) and were used for assessing previous crash location exposures of individual drivers.

<u>The Map-Matching Process</u>

To allow analyses of driving patterns of individual drivers such as speeding behavior based on speed limit and facility-specific behavior activity exposure, driving activity records observed by the GPS technology must be matched to roadway characteristics (facility type, posted speed limit, roadway design inputs, and rural/urban area type). The Commute Atlanta study research team at Georgia Tech developed series of automated map-matching algorithms to combine GPS-based trip data with roadway characteristics (RC) information in the Geographic Information System (GIS) [5, 16]. The research team determined that the use of two map-matching methods, route method and buffer method, in combination provided the most complete and accurate data for driving activity analysis [5, 16].

Data Mining and Filtering

Similar to the map-matching process, to reduce data processing time and to estimate reliable vehicle mileage and speed estimates, this study used an automated GPS data filtering algorithm by modifying the conventional Kalman filter. As will be discussed in Chapter 4, this was done because even though the GPS technology provided highly accurate data (with 95% of the data falling within 3 meters of the actual vehicle location), random errors were noted in the data stream [17]. While the conventional Kalman filter smoothes all GPS data points with the constant rate (one measurement error), the modified Kalman filter selects two GPS measurement errors based on the quality of GPS data points such as number of satellites and Position Dilution of Precision (PDOP) value.

In addition, unlike speed data, which are typically measured directly by onboard monitoring systems, acceleration is usually a derived parameter, calculated from consecutive speed measurements. Three approaches are often used to derive acceleration rates for a given speed vs. time profile: forward difference, backward difference, and central difference methods [18]. Thus, the different combinations of speed and acceleration depending on adopted methods may result in different results. Finally, this study utilized the central difference method for estimating acceleration behavior since this method not only has strong mathematical background, but also tends to provide moderate acceleration impact on the classification analysis.

Statistical Methods

To assess whether drivers who were involved in crashes have different mileage exposure, speeding behavior, acceleration patterns, and other driving behavior activity patterns from drivers who were not involved in any crashes, this study utilized various statistical methods. For testing differences in means of each driving behavior activity metric, this study performed the Wilks' lambda test (parametric technique) and the bootstrap technique (non-parametric technique) due to the small sample size and non-normality of some variables. To test normality of each variable, this study utilized three nonparametric methods; Jarque-Bera test, Lilliefors test, and Kolmogorov-Smirnov (KS) test [19].

This study also used Chi-square test, Kruskal Wallis test, histogram method, and Gaussian kernel density estimation method to evaluate distributions of behavior activity exposure measure. To filter random errors in GPS data streams, this study utilized the modified Kalman filter algorithm after comparing with least squares spline approximation method, the kernel-based smoothing technique, and conventional Kalman filter algorithm.

For the classification modeling process, the study mainly performed two different techniques, linear discriminant analysis (LDA) and logistic regression model. Classification and regression tree (CART) analysis was also utilized to visualize the result of classification. From the three classification techniques, this study provided insights about which driving behavior activity exposure metrics can effectively classify drivers into difference crash involvement groups.

Research Contributions

This study is one of the first attempts to evaluate crash involvement rates of individual drivers using driving behavior activity data longitudinally collected from GPS-instrumented vehicles. Thus, this study discusses limitations of existing methods used in various transportation research and provides useful techniques of implementing GPS data streams in safety research, especially in large-scale data collection processes.

Second, unlike the previous research efforts that employed aggregate exposure measures, this study proposes numerous driving behavior activity exposure metrics to evaluate the probability of being involved in a crash of individual drivers. Based on the proposed exposure metrics and the developed models, the cause-effect relationships between driver behavior activities and crash involvements can be evaluated in detail.

Third, this study examines differences in driving behavior activities of drivers who were involved and not involved in crashes and discusses that drivers within the same demographic characteristics such as age and gender might not exhibit the same driving behavior activity patterns. As a result, driving behavior activities may be much more strongly related to crash involvement rates than the general demographic characteristics. Thus, this study potentially identifies a better framework employed by insurance industries to estimate insurance premium since the results of this study shows much reasonable and equitable premium structure to customers.

Fourth, this study expects that driving behavior activity metrics of individual drivers may be utilized as one of safety surrogate measures to identify potential hazardous roadways where hard deceleration events or high speeding patterns frequently occur.

Fifth, this study provides more detailed and effective techniques that evaluate potential crash involvement rates of individual drivers and expects that current driver education programs and other safety campaigns may be improved.

Finally, it is expected that this study provides useful guidance and serves as a model for assessment of the relationships between driving behavior activity patterns and crash involvement rate to researchers who plan to evaluate with larger sample data in future.

Dissertation Outline

Following this introductory chapter, Chapter 2 summarizes previous and existing studies related to relationships between driving behavior activities and crash involvement rates and discusses their research limitations. Chapter 3 illustrates the capability of the data collection device employed in this study and explains the characteristics of the collected data streams. Chapter 4 discusses data quality issues and techniques for minimizing GPS random errors and provides the review on acceleration calculation methods. Chapter 5 evaluates relationships between disaggregated travel mileage-related metrics and crash involvement rates. Chapter 6 examines relationships between travel duration-related metrics and crash involvement rates. Chapter 7 discusses relationships between speed activities and crash involvement rates. Chapter 8 shows relationships between potential acceleration-related behavioral metrics and crash involvement rates. Chapter 9 assesses relationships between speed stability patterns and crash involvement rates. Chapter 10 investigates relationships between crash involvement rates and other potential crash-related behavioral metrics such as unfamiliar roadway exposure, frequency of turn movement exposure, and previous crash location exposure. Chapter 11 utilizes three statistical classification methods, the linear discriminant analysis (LDA), classification and regression tree analysis (CART), and logistic regression model and performs modeling process. Chapter 12 suggests possible applications of driving behavior activity metrics to assist current safety surrogate measures and driver education programs. Finally, Chapter 13 discusses the findings from this study and provides limitations of this study and future research suggestions.

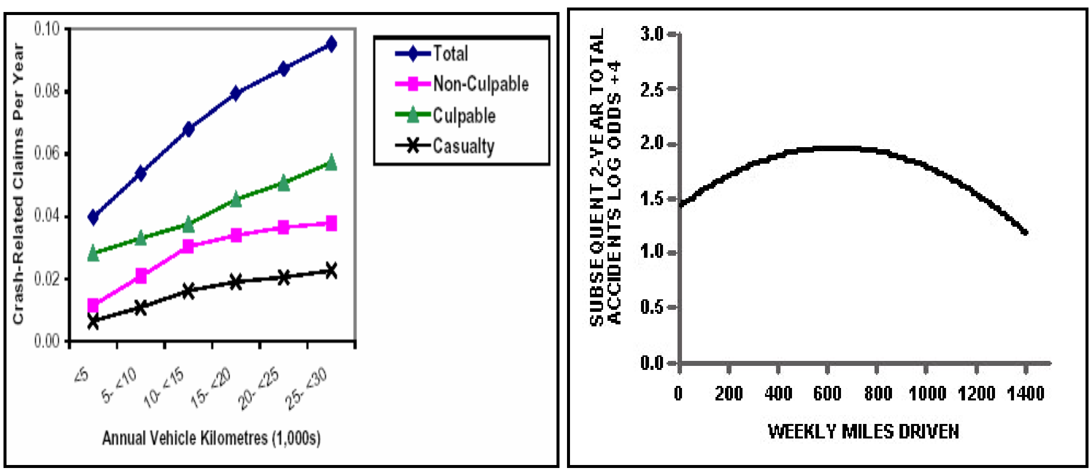
Chapter Two

LITERATURE REVIEW

As briefly discussed in the previous chapter, knowledge about characteristics of crash-involved drivers and their driving behavior activity patterns can enhance current countermeasure programs designed to improve safety on our roadways. However, much of the safety research efforts have mainly focused on the vehicle or the roadway design aspects, relatively few studies have been performed on the driving behavior activity components due to the lack of reliable behavior activity data and the absence of effective data collection techniques [5]. There are various driving behavior activities related with crash involvement rates such as travel mileage exposure, speeding behavior, hard accelerations and decelerations, etc. Travel mileage is the most commonly used behavioral exposure measure since crash risk is generally considered the product of travel mileage and per-mile crash frequency [20, 21]. The other reason why travel mileage is popularly used is that it is relatively easy to be collected than other driving behavior activity data such as speed, acceleration, speed stability patterns, unfamiliar roadway exposure, frequency of turn movement exposure, and previous crash location exposure.

Wolfe [20] defined the exposure as: "*A measure of the frequency of being in a given traffic situation, which number can be used as the denominator in a fraction with the number of accidents which take place in that situation as the numerator, thus producing an accident rate or risk of being in an accident when in that situation*".

Drummond [21] also considered exposure "*the opportunity being involved in a crash*" and suggested that various surrogate measures such as population, license held, registered vehicles, and distance or duration of travel could be used. Based on the definition of exposure, a positive relationship between travel mileage exposure and crash risk has been established in numerous previous studies [9, 22] (Figure 2), and travel mileage became a useful measure being used for classifying potentially high crash risk drivers in both safety programs (driver education programs) and automobile insurance programs (pay-as-you-drive or mileage-based insurance programs).



Liman [7] **Gebers [22]**

Figure 2: Relationships between Mileage and Crash Risk

In fact, while travel mileage is the popular exposure measure being used for classifying high crash risk drivers, there are some debate about the relationships between travel mileage and crash risk. For example, Litman [7] showed the crash-related claims are linearly related with the annual vehicle kilometers (VKT) (Figure 2), but Janke [9] claimed that higher-mileage vehicles tended to have relatively low per-mile crash rates since high VMT drivers usually have more recent modeled-vehicle and better safety devices and have much more experiences on driving. Similarly, Gebers [22] also provided that crash risk had a “inversed U-shaped curve”, which meant that crash risk generally increased by the some point of travel mileage but decreased again, especially for high mileage drivers such as commercial drivers.

However, except extremely high mileage drivers, crash risk of normal drivers has an increasing trend as travel mileage increases. Thus, most previous crash-exposure studies and the structure frame of mileage-based insurance program have used the mileage exposure measure to predict or estimate crash risk of drivers since they assumed that exposure had a positive relationship with crash involvement rates.

Background of Pay-As-You-Drive Insurance Program

Since modern societies are currently confronted with various transportation-related problems such as traffic congestion, high energy consumption, incident or accident, aggressive driving behavior, and poor air qualities caused from rapidly increasing vehicle ownerships and vehicle usage rates, transportation engineers and researchers are seeking the ways to solve those traffic-related problems though the various methodologies and techniques.

Until 1980s, to reduce traffic congestions or delays, most countries including the U.S.A. had tried to construct new infrastructure (mainly freeways) or increase the existing facility dimension (lane width or the number of lanes). However, these solutions could not clearly solve problems and degraded traffic conditions later since the construction of new facilities (or infrastructures) or the extension of existing roadways attracted more potential vehicle demands. In addition, the rapid growth of population in cities and the limited land capacity prevented this investment strategy from solving problems. Thus, transportation engineers started seeking alternative strategies to effectively solve transportation-related problems without the construction and the extension of facilities.

Transportation Demand Management (TDM) is a good example of recent approaches for overcoming those problems. Of the various TDM strategies, value pricing strategies are currently obtained much attention from traffic engineers or transportation policy decision makers. The value pricing strategies exactly focus on how to efficiently use the existing roadways instead of constructing new facilities. These value pricing strategies generally include various methodologies that can be classified into two approaches as follows.

- Provide incentive to users who pay for facilities
- Provide incentive to users who do not use facilities (or personal vehicles).

The first approaches that provide incentives to users who pay for using some types of roadway facilities are High Occupancy Toll (HOT) Lanes, Fast and Interwined Regular (FAIR), and cordon pricing scheme program. On the other hand, the second approaches that provide incentives to users who abandon private vehicle usages are commuter options, transit and rideshare, parking cash out, and value pricing of insurance program (PAYD).

Although those approaches have the same goal to reduce traffic congestions and mobile source emissions, due to the different approaches they used, the expectancy of users for these approaches is different. The first approach can provide some negative impacts on low-income drivers or minorities since this approach requires that users must pay when they use facilities. However, the latter approach does not provide the inequality issues since this approach do not require monetary cost to users when they need to use those facilities. Instead, this approach provides incentives to drivers who give up using private vehicles. Thus, the second approach is much satisfied with the transportation equity issue recognized as the important requirement for implementing congestion mitigation strategies.

Among various incentive programs, the pay-as-you-drive (PAYD) program (or mileage-based insurance program) has been receiving increased attention from planners and transportation policy makers since this PAYD program is focusing on the current problems of insurance premium estimates as well as of congestion and delay on roadways. The PAYD program can change the insurance premium from a fixed vehicle cost to a flexible cost. Figure 3 shows that the automobile insurance cost has 21 % of total vehicle operation cost [7].

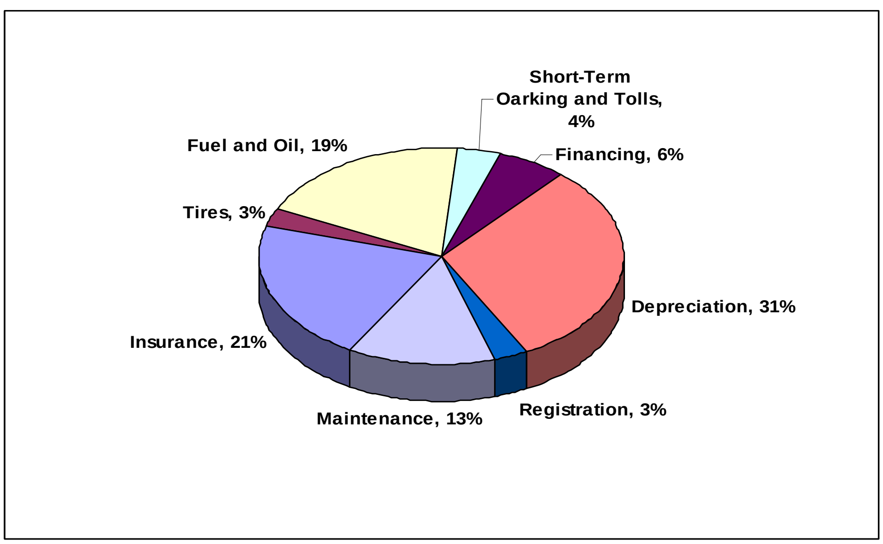


Figure 3: Vehicle Operation Costs (Litman [7])

Up to now, an automobile insurance premium was mostly determined by driver history, driver characteristics (age and gender), vehicle type, and registered home location and did not consider any vehicle usage rates. Thus, the current insurance premium structure does not properly estimate the crash risk (crash involvement rates) of individual drivers. If insurance companies implement the PAYD program using vehicle usage rates, people can reduce their economic burden since drivers can modify or reduce their vehicle usage rates such as travel mileage. Litman [7] found that drivers participated in the PAYD program in England obtained 30~50 pounds of the economic benefit on average.

The Current Automobile Insurance Premium Structure

The crash exposure measures based on individual driving behavior activities are seldom used to evaluate the level of safety of individual drivers due to the difficulty of data collection. Previous studies had focused on only aggregated or regional exposure measures such as total travel mileage (VMT), drivers licensed, vehicles registered, and age/gender-based population, which can be obtained from the census data. Thus, the current risk clusters are determined by only aggregated exposure information, meaning that once drivers are clustered the relative probability of being involved in crashes does differ significantly across the clusters. General information used by insurance industries is as follows;

- Driver characteristics
  - gender, age, marital status, crash and citation history, education level, and income level
- Vehicle types
  - vehicle model, vehicle type, model year, auxiliary safety devices, and repair history
- Residential locations
  - residential county, ZIP code, urban, rural, downtown, and commercial district

These classification factors using by insurance industry were developed with a relatively large crash data set for long time period, and numerous research and studies have verified the relationships between those classification factors and crash risk (involvement rates). However, those aggregated (or generalized) clusters cannot distinguish from differences in individual crash risks. Within each cluster, individual driver crash risk may vary significantly. For example, some young drivers may have significantly low crash risk than middle age drivers or older drivers, and some of female drivers may also have higher crash risk than some of male drivers. Due to the limitation of the current classification decision rule, these potentially misclassified drivers have disadvantages from the automobile insurance premium. In other words, the current insurance industries has a cross subsidies, which the misclassified low risk drivers should support other drivers and the misclassified high crash drivers pay relatively low insurance premiums. This current issue degrades the efficiency level of insurance program and increases drivers' (or customers) arguments. More efficient risk clusters (or alternatives to clustering) can be developed once more detailed behavioral data become available.

In addition, automobile insurance premiums are also determined by the registered home (or vehicle) locations that are just fixed points. However, vehicles are always moving based on driver's purpose and destination. Although a vehicle was registered in low crash area, this vehicle may frequently travel at congested or downtown area where crash rates are relatively high, depending on a driver's work location or the purpose of the trips. Without knowing travel or activity patterns, crash risks of individual drivers cannot properly estimated.

The last important limitation of current insurance premium structures is that there is little that drivers can do to reduce their insurance premiums (drivers cannot change their gender or age), other than to avoid receiving tickets and avoid being involved in crashes. Yet, drivers never receive reinforcement on driving behaviors that can help keep them from receiving tickets of being involved in crashes. If risk-related driving behavior measures are integrated into the insurance premium structure, and drivers receive feedback on their onroad performance, drivers can try to reduce their risk-related behavior activities to decrease their insurance cost. Drivers may avoid making certain trips, avoid nighttime travel, and reduce high-speed activity or hard acceleration events. Such changes may also produce positive impacts on various transportation-related problems such as fuel economy, emissions, and even traffic congestion.

The Commuter Choice and Value Pricing Insurance Incentive Program

A few academic research groups including Georgia Tech and insurance companies such as progressive insurance company are currently studying the feasibility of a value pricing of insurance program using a driving behavioral exposure. The PAYD insurance pilot program is the first attempt to integrate individual driver's driving behavioral exposure such as travel mileage (main component), speed, and time of travel (especially, young drivers at nighttime) into the insurance classification decision rule where a higher crash-related behavioral exposure means higher chances of being involved in crashes and yields higher premium costs.

The basic idea of the PAYD program is that drivers having higher dependency (high travel mileage and large number of trips) on vehicles must pay higher insurance premium than other drivers who travel low mileage with less numbers of trips. The actual concept of this PAYD insurance program is originated from a rational and statistical point of views, which means that the opportunity of motor vehicle crashes (or crash rates) generally increases as drivers travel longer or use vehicles more. This concept is expected as a more reasonable approach for estimating automobile insurance premium since the conventional structure of automobile insurance premiums is employing only common criteria to specific driver groups. As discussed earlier, all young male drivers are similarly adjusted by high insurance premiums with small variations depending on vehicle types, home locations, and crash or citation history since current research and insurance societies consider young male drivers high crash-risk drivers.

Although current PAYD programs seem reasonable, few studies have been able to analyze detailed individual exposure data including travel by time of day, day of week, facility class, and area types. Drivers traveling the same distance may face significantly different crash risks depending on time of day and facility types that they travel and based upon how they operate their vehicles. In addition, the travel mileage can not describe whether they drove at high speeds or they frequently changed their driving speeds, which are also highly related to the probability of crash involvements. New classification method can use driving behavior activity metrics that would describe major travel conditions: where drivers drive, when they drive, how they drive, and under what environmental conditions they drive [4, 6].

Potential variables affecting crash risk

1. where a driver is driven
    a. Freeways, arterials, and local roads
    b. Congested roadways and uncongested roadways
    c. Hazardous roadways (roadway design flaws)
2. When a driver is driven (time of day)
    a. Relative risk of encountering drunk drivers (e.g. nighttime)
    b. Relative risk of exposure to other vehicles (e.g. traffic volumes)
3. How a driver is driven
    a. Travel mileage and duration
    b. Speed profile (vehicle overspeed and underspeed)
    c. Acceleration profile (hard acceleration/deceleration events)

    d. Speed change rates

    e. Turning movements (left or right turns)

4. Environmental conditions

    a. Congested conditions (AM peak hour and PM peak hour)

    b. Transitions from uncongested to congested conditions

    c. Rain (pavement conditions and visibility)[2]

Driving Behavior Activity Exposure Metrics

Before GPS and mobile computing technology were introduced, information on a driver's exposure by facility type and time of travel was difficult or impossible to be collected. In the absence of better methods, travel diary surveys and telephone interviews were a popular means for collecting driving behavior activity data, especially to obtain travel mileage and travel duration. One limitation of diaries and interviews is the lack of reliability and validity information. Ogle et al. [5] recently conducted an analysis of the accuracy of household trip reporting by comparing simultaneous GPS-measured trip data in the Commute Atlanta program with trips reported in a standard two-day travel diary survey. A total of 2,292 trips were found from the GPS-measured trip files, but only 1,622 trips were reported by the corresponding two days of travel diaries, with an under-reporting rate of 29.2%.

In addition, after comparing trip duration and mileage estimates from the both methods, the researchers concluded that the trips reported by the survey diaries produced only 90% of total travel duration and 78% of total mileage. When comparing only

[2] The weather variables are not possible to use in this study since only one weather center (Hartsfield International Airport) in 13 counties is operating and its data is not précised.

reported survey trips with corresponding GPS-measure trips, respondents overestimated their travel duration and mileage by 15% and 2%, respectively. While the under-reporting and overestimation of trip duration nearly cancel each other out, the total travel mileage obtained from the surveys is off by 20% from actual.

Although travel mileage and duration could be obtained from the previously described survey methods, a driver's mileage and speed on the certain types of roadways (specific travel routes such as freeways, arterials, or local roads) could not be obtained. Kirk et al. [23] tried a new method for travel data collection. They distributed maps and asked participants to mark their travel routes directly on the maps. However, they found that participants were unwilling to follow the requirements due to safety issues and the additional burden. Unfortunately, they did not implement an alternative data collection method and therefore, the accuracy of this method was unknown.

On the other hand, one possible source of speed data corresponding with the crash data is the official crash reports, but speed information is usually biased since this information is a driver's self-reported value and a driver tends to not invoke his (or her) true driving speed. Witness can also provide over- or under-estimated driving speeds. Solomon [8] and Kim et al. [24] claimed that there was the obvious possibility that drivers might tend to under-estimate their speeds and maintained it was inconsequential. White and Nelson [25] also suggested that under-estimated speeds by drivers could contribute to biased relationships between crash rate and speed. O'Day [26] and Shinar et al. [27] mentioned that reported data about driver condition such as speeds or alcohol are often unreliable while driver information such as gender and age are generally correct. Although the speed information on the crash report is not highly accurate, most transportation engineers and researchers are using this information to verify the relationships between speeding and crash involvement rate.

Another technique for collecting speed and acceleration data of drivers is to directly measure vehicle speeds at the pre-determined roadways using a speed gun. A few studies recorded vehicle license plates and took pictures of drivers while measuring speeds of traffic. Later, those recorded license plates were used for obtaining driver's crash history data and matching a photo being taken at the site with driver information on crash data. Although they examined relationships between driving speeds and crash rates, since the speed (actually spot speed) data were measured at the only pre-determined sites, longitudinal speeding and acceleration behavior of individual drivers could not be investigated.

Other studies pre-assumed that young or male drivers were high crash risk groups and then examined the relationships between speeding behavior and demographic information such as age and gender. Those studies confirmed that young or male drivers have much speeding behavior than old or female drivers and analogized that speeding behaviors had a strong relationship with crash involvement rates based on the pre-determined assumption. Thus, since those studies did not investigated in detail about the speeding behavior of drivers within the same demographic driver groups, the results of them can be interpreted as all drivers within each demographic group have the same speeding behavior and imply the same crash involvement rates. In addition, it was not clearly provided when and where high risky drivers show high speeding patterns and what kinds of different speeding patterns existed between normal drivers and high crash risk drivers.

Few research efforts on evaluating acceleration behavior among drivers who have different crash involvement rates have been performed due to the difficulty of data collection. Greenshields at al. [10] evaluated the hypothesis that drivers with different crash histories exhibit different driving characteristics such as delay time, running time, amount of speed changes, amount of direction changes, acceleration reversals, steering-wheel reversal, and brake applications. They selected one test route, 17 mile distance containing urban and rural areas, recruited about 140 drivers (40 drivers in control group, 40 driver in high accident group, 20 drivers in high violation group, and 40 drivers in just beginning driver group) having different crash histories, and let them drive one instrumented vehicle through the months[3]. As a result, they found that significantly different driving behavior metrics between the control group and the high accident group were running time, accident reversal, and steer-wheel reversal. The amount of speed changes was not significantly different in this study.

Wahlberg [28] recently examined the relationship between acceleration behavior and accident frequency for local buses. Accident data were obtained from the local bus company in Uppsala, Sweden and acceleration data were gathered by a researcher traveling on a bus with the accelerometer equipment that measures speed changes with 10 Hz. Based on the measure of celebration, which is the absolute mean of speed changes, this study did not find any strong relationships between them but suggested that celebration behavior has a higher predictive power than speed.

Since only a few studies had analyzed acceleration behavior for evaluating crash risk and those studies used only means of accelerations as a metric, it is hard to determine whether acceleration behavior has a relationship with the crash involvement rates or not.

[3] The numbers of test runs of individual drivers were unknown.

However, acceleration behavior is still a possible candidate of driving behavior activity exposure measures because with the basic knowledge of aggressive drivers and their crash risk, a drivers having frequent hard accelerations or decelerations implies that they have many chances of being involved in crashes.

With data from GPS instrumented vehicles, recent studies have begun investigating the relationships between driving speeds and acceleration patterns and crash involvement rates in more detail (2-4). In addition to the means of accelerations that were traditionally employed, various metrics can be applied such as the amount of accelerations per trip (or mile) and the frequency of hard accelerations. On the other hands, Klauer et al. [29] recently conducted an analysis of impact of inattention on near-crash and crash risk from naturalistic driving data from 100 GPS-instrumented vehicles. However, they focused on driver distractions such as fatigue, drowsiness, eye glance, and secondary task (eating or cell phone use), so they did not investigate driving behavior activity patterns such as travel mileage, speeding, and acceleration patterns.

The GPS-measured travel data also provide abundant reliable information which can help better identify the relationships between driving behavior and crash risk under varying conditions of facility type and time of day. Coupling the detailed travel information with known driver, household, and vehicle characteristics, operations can then be tied back to a wide variety of socio-demographic parameters. Furthermore, GPS-measured data can be used to identify how driving behaviors change during a trip in response to changes in roadway operating conditions.

Using the GPS-observed data from the Commute Atlanta program, this study examines relationships between crash involvement rates and behavioral exposure measures such as travel distance, travel duration, speed, acceleration, speed stability patterns, unfamiliar roadway exposures, left/right turn exposures, and previous crash location exposures that previous studies could not perform.

Chapter Three

DATA COLLECTION

Description for Trip Data Collectors

Researchers from the DRIVE Atlanta Laboratory at the Georgia Institute of Technology developed a wireless data collection system known as the "Georgia Tech Trip Data Collector (GT-TDC)" and installed in approximately 460 light-duty vehicles in the metro Atlanta (about 13-county area) (Figure 4) through the Commuter Choice and Value Pricing Insurance Incentive Program (Commuter Atlanta program) funded by the Federal Highway Administration (FHWA) and the Georgia Department of Transportation (GDOT) and in four transit buses operated by the Metropolitan Atlanta Rapid Transit Authority (MARTA).

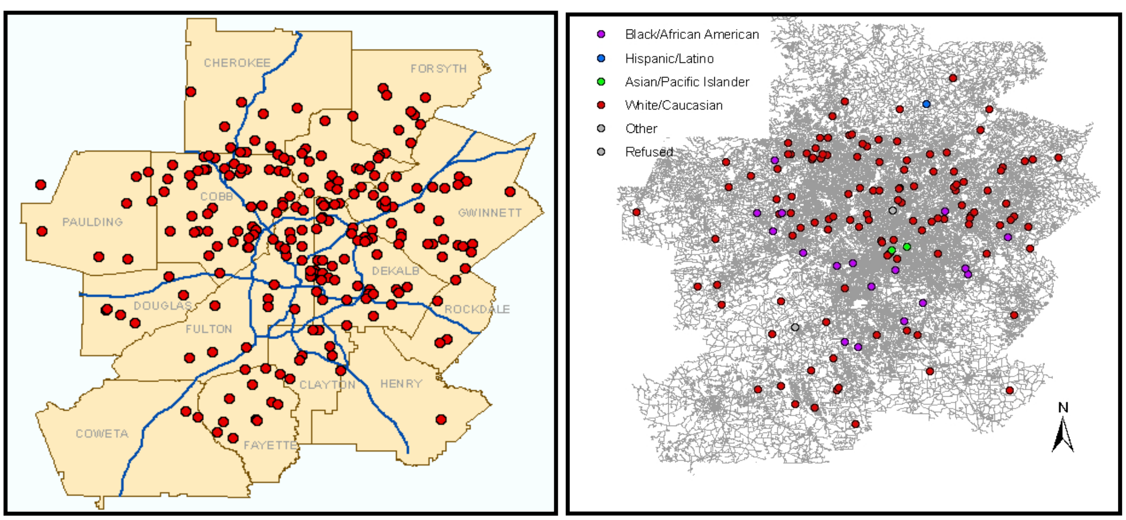
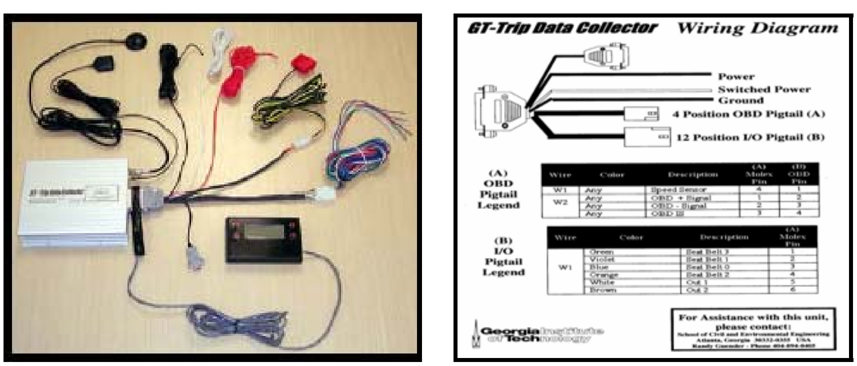


Figure 4: Participant Residential Locations and Ethnic Groups in Commute Atlanta Program (N = 261)

The GT-TDC (Figure 5) embeds the Linux-based 386 computer and operates a system with 12 V-vehicle power with an extremely low 3 mA power draw when a vehicle is not in operation mode. The GT-TDC is also very small size (8” by 6” by 2”) and light weight, so it is designed not to provide additional weight to a vehicle.

Figure 5: GT Trip Data Collector (GT-TDC)

The global positioning system (GPS) receiver (SiRF Star II), Time Division Multiple Access (TDMA) cellular transceiver (Ericsson DM 15), the onboard diagnostics (OBD) connection, and the vehicle speed sensor (VSS) connection are embedded in the GT-TDC in order to collect and transmit various vehicle operation data. The GT-TDC also requires the cellular antenna, GPS antenna, and OBD connector (Figure 6).



(GPS and Cellular Antenna, OBD connector, and Power cables)

Figure 6: Wiring Harnesses and Wiring Diagram

The GT-TDC is usually placed under the front passenger seat to protect the unit during a crash or from any unexpected damages, provide better connections and air flow, and not to inference with driving (Figure 7).

Figure 7: Placement inside a Vehicle

After collecting comprehensive second-by-second vehicle and engine operating parameters using the GPS, OBD, and VSS systems, the GT-TDC can transmit the combined data streams using the wireless data transmit system via cellular connection (the TDMA cellular transceiver) during off-peak periods (10 pm to 6 am during weekdays and anytime on weekends between 10 pm Friday and 6 am Monday) to a secured central server at Georgia Tech. Figure 8 illustrates how the GT-TDC collects data and transmits them to Georgia Tech servers via wireless network system.

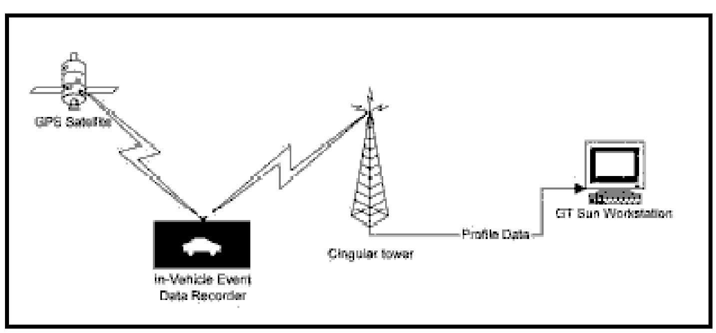


Figure 8: Data Transferring Process

Capability of the GPS Receiver implemented in the GT-TDC

The GT-TDC integrates the 12-channel SiRF Star II GPS receiver, which is designed for in-car navigation systems. This receiver was selected for the Commute Atlanta program in 2002 when a previous study conducted by Ogle et al. [30] found that this GPS receiver provided similar performance for collecting vehicle speeds and accelerations as did the Differential Global Positioning System (DGPS) receiver once selected availability (SA) was eliminated in 2000.

The SiRF Star II GPS receiver calculates the vehicle location based on Clear/Acquisition (C/A) code communication between satellites and the receiver and separately estimates vehicle speed using the Doppler effect (vehicle speed is independent of vehicle location). While Real Time Kinematic (RTK) GPS systems can resolve uncertainty in vehicle location and speed estimates, there are four reasons why researcher team could not implement the PTK-GPS system in the GT-TDC:

- RTK-GPS equipment is too costly for use in large deployments (the Commute Atlanta program recruited about 500 vehicles)
- RTK-GPS systems require sub-devices (two or more GPS antenna, a rover radio, a base station radio, base station GPS antenna, and rover receiver as well as additional base stations), and equipment packages needed to be small and self-contained
- RTK-GPS systems typically require the onboard GPS receiver be within a boundary of 6 miles (10 km) from the base station with line-of-sight between the

reference receiver and the rover receiver [31] (which would not be possible for vehicles roaming throughout the entire Atlanta 22,000 km^2 metropolitan area).

- Even though high-end RTK-GPS systems are very accurate, the loss of satellite signal lock due to the overhead obstructions will still affect position and speed data [32] and statistical smoothing techniques may still be required.

Bench Test and Programming Process

Before installing the GT-TDC into the participating vehicles, an initial configuration and testing processes (bench test) are required to insure whether the GT-TDC is fully functional or not in terms of data collection and data transmission. During this process, the GT-TDC is assigned with a specific serial number (a unit ID), cellular phone number matched with Electronic Serial Number (ESN) provided by Cingular, File Transfer Protocol (FTP) connection username, FTP password, FTP Internet Protocol (IP) address, and OBD mode (see Appendix A). After finishing this initial configuration process, a verification process should be performed to check data transmission capability. Figure 9 illustrates that the GT-TDC is collecting locations, speed, altitude, heading, and time data from the GPS receiver.

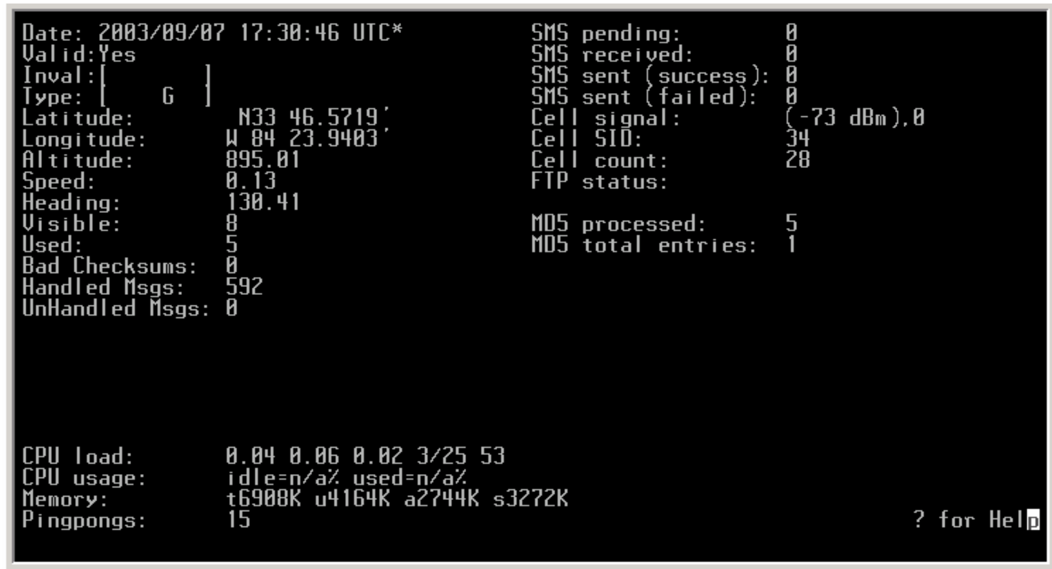


Figure 9: Statue of GPS Data collection in Bench Test

After checking the capability of GPS data collection, the cellular communication capability should be tested before GT-TDCs are deployed in the field. The communication between the bench-test platform and each GT-TDC through the cellular network can be established via a Short Message System (SMS), which is the similar with text message service. Through the SMS, researchers can retrieve GPS location data, speed, heading information, engine-related various data (see Appendix A), VSS sensor information, and cellular signal strength.

In addition, the embedded-software and configuration parameters of the GT-TDC can be upgraded via the SMS command. Figure 10 shows that the GT-TDC received two SMS commands (including one pending command) and sent corresponding SMS messages to the Georgia Tech server.

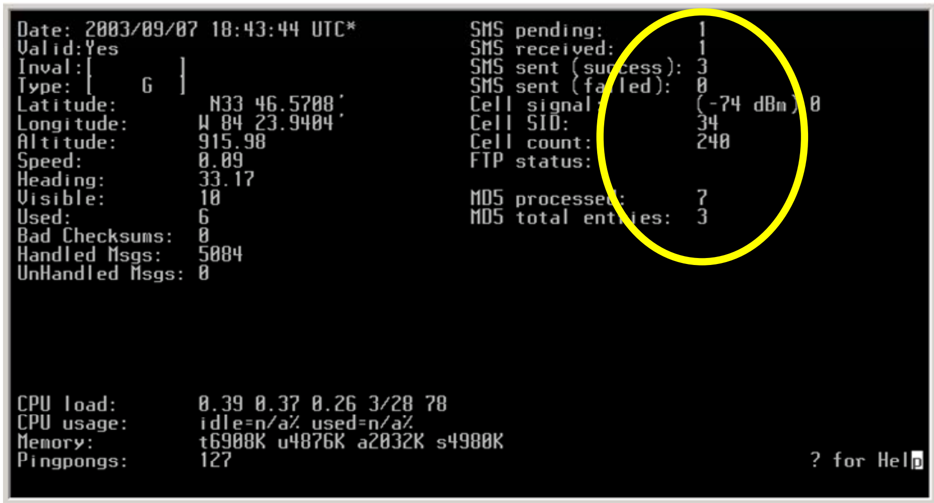


Figure 10: Communication Process with the GT-TDC in Bench Test

Furthermore, after testing the SMS capability, the capability of a Circuit Switched Data (CSD) transmission should be tested since the CSD system is used for the transmission of collected trip data. The GT-TDCs record second-by-second vehicle activities during vehicle operation, make a trip file after engine-off, and upload trip files via cellular CSD system into the Georgia Tech research server. Figure 11 illustrates trip files that were transmitted in the GT server from a GT-TDC installed on the vehicle of participants via a cellular network.

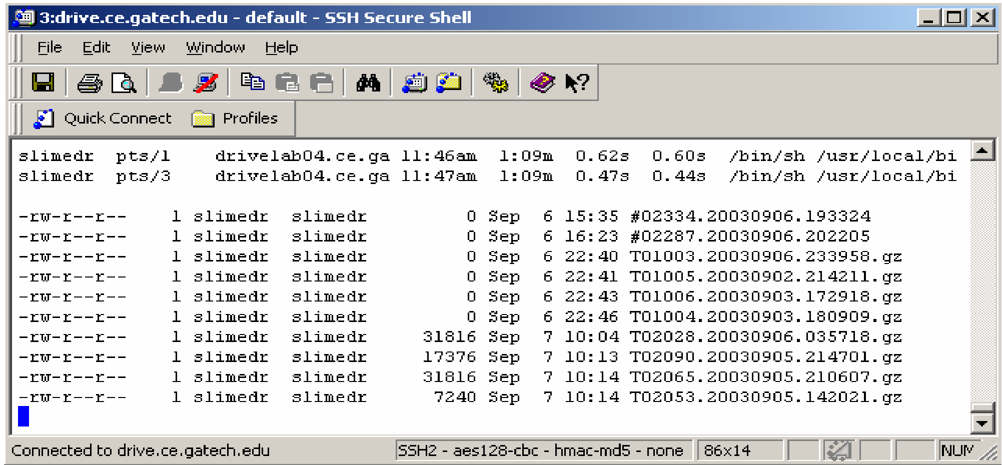


Figure 11: Status of Trip File Upload

Due to the large size of a trip file containing second-by-second vehicle activities and a privacy issue on trip files, all trip files inside the GT-TDC are initially encrypted in the small size and deciphered only within the secured GT server. After deciphering trip files (binary files) and converting them to ‘csv’ (comma separated values) files, researchers can utilize each trip file for additional data cleaning processes. Figure 12 illustrates the flow of data transmission.

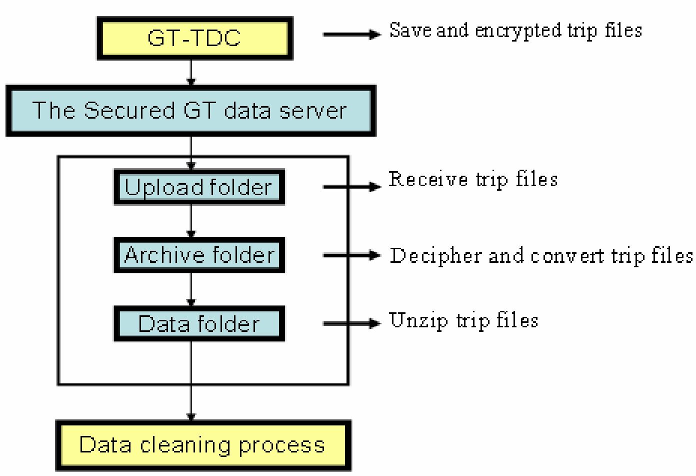


Figure 12: Trip Data Process in the GT Server

Data Characteristics of Trip Files

The GT-TDCs in the field monitor second-by-second (1 Hz) vehicle activity data including vehicle location and speed from a GPS receiver and capture redundant speed data (4 Hz) using the vehicle speed senor (VSS) system. The GT-TDC also collects ten engine-operating parameters including speeds from the onboard diagnostics (OBD-II) system for most post-1996 model year vehicles. These VSS- and OBD-measured speed profiles can be used to calibrate the GPS-measured speed data and modify the speed errors from the GPS.

<u>Onboard Diagnostics (OBD) System</u>

Among various OBD parameters, the GT-TDC collects ten engine operating parameters such as engine resolutions per minute (RPM), engine load, speed, engine coolant temperature (ECT), throttle position sensor (TPS), intake air temperature (IAT), manifold air pressure (MAP), mass air flow (MAF), ignition advance data, and oxygen sensors data during a vehicle's operation. Other parameters such as fuel trim data currently not available in the GT-TDC can be collected when the corresponding standard hex numbers of parameters are updated or changed by reporting the SMS command via the air.

To utilize the OBD-collected data, it should be understood that the OBD system does not have a standardized data polling time while the GPS provides location coordinates and speeds at 1Hz (second-by-second) basis and the wheel-tick sensor collects wheel-tick data at 4Hz (every 0.25 second). In fact, some of transportation engineers or researchers are using OBD-measured data without well understanding about the characteristics of the OBD data stream. The actual data transmit rate (data polling frequency) of the OBD system strongly depends on the algorithms of vehicle manufacturers and conditions of engine operations since the Society of Automotive Engineers (SAE) standardized only the maximum polling time of each engine parameter.

Thus, each vehicle model has the different data polling frequency, and even the same vehicles, an engine computer can change the data polling frequency of engine parameters depending on the current operating conditions such as hard acceleration or high load conditions, for example, a Dodge Durango stops transmitting real engine data for seconds at a time under hard accelerations. The OBD system is not a time-based data collection system, but an opportunity-based data collection system.

If the OBD data collector, so-called a data reader, provides the same speeds during the certain period, it can be either the truly same speeds during that period of time or just inaccurate speeds caused by the unavailable of OBD data polling. Thus, correct understanding on the OBD data streams is critical to analyze various types of transportation studies using the OBD-measured data.

To assure that the values of engine operating data are real-time data, the GT-TDC implemented an additional control parameter, a Pass Counter (PC), into the trip files. Thus, each OBD data stream contains the values of pass counter, which changes from zero to one or one to zero whenever real data are collected. The continuous OBD data streams having the same pass count values indicate that the values of parameters do not represent the real engine-related value at that moment (Figure 13).

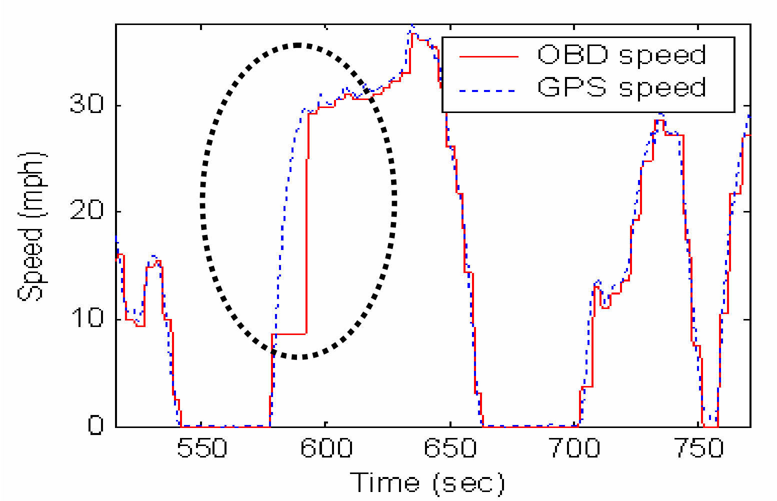


Figure 13: Example of Unreliable OBD Speed Profiles

Figure 14 illustrates examples of OBD parameters using the GPS location data in the GIS platform.

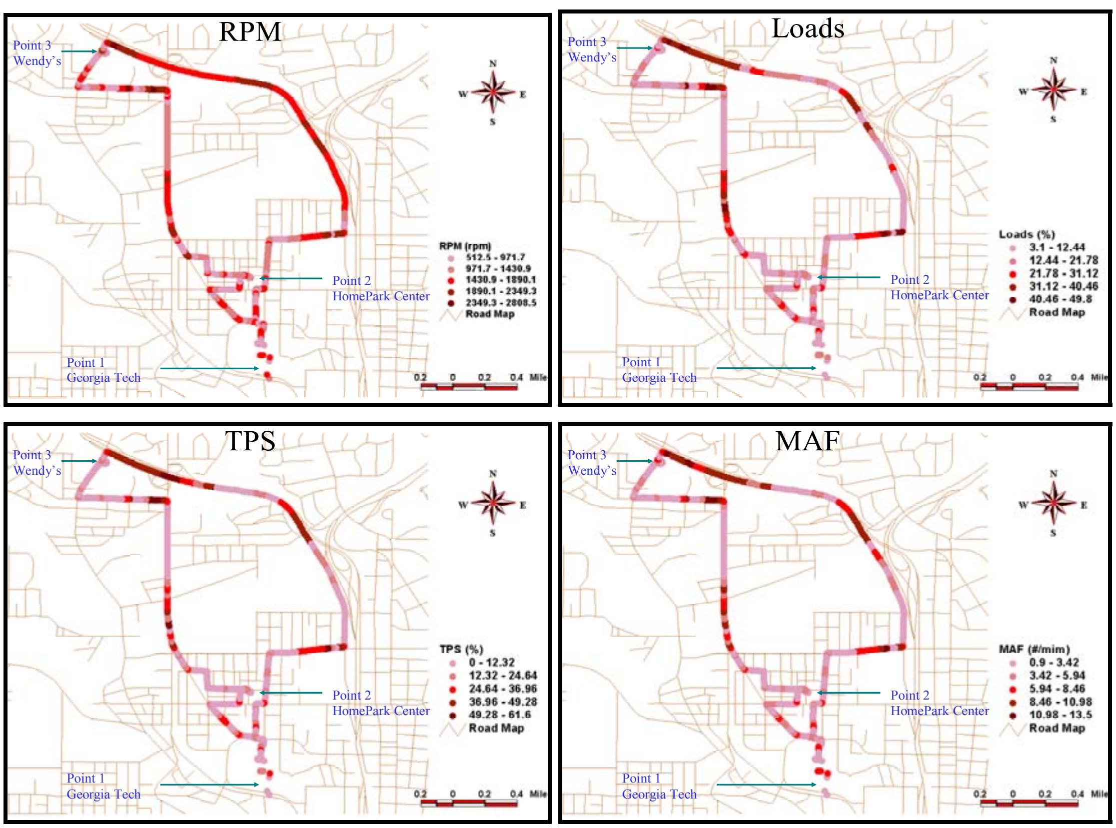


Figure 14: Displays of OBD Parameters in GIS Platform

The Vehicle Speed Sensor (VSS) Technology

Simply speaking, the vehicle speed sensor (VSS) system is the same as the OBD system in terms of speed measurement since the VSS system provides signal information regarding speed measures and the change of speed component to the engine computer. However, The GT-TDC provides different VSS speed measure from the OBD system since the GT-TDC directly monitors speed information from the VSS system measuring transmission or transaxle output shaft, not from the engine computer while the GT-TDC collects OBD speed data from the engine computer. Thus, the VSS-measured speed profile can help researchers identify the quality of GPS and OBD-measured speed profiles.

The VSS monitors the number of electronic pulses which counts the number of driveshaft revolutions [30], and the computer module inside the engine computer converts the number of electronic pulses into the speed profile based on a function of the standard tire diameter of vehicles. To derive speed profiles from the number of electronic pulses measured by the VSS, the calibration number which is the number of pulses for traveling a distance of one mile is required. In general, vehicle manufactures use their own standard calibration numbers such as 2,000 pulses/mile (two-pulse at one revolution), 4,000 pulses/mile (four-pulse at one revolution), or 8,000 pulses/mile (eight-pulse at one revolution), based on vehicle and VSS types. The more detailed description of the VSS system will be provided in Chapter 4.

Roadway and Network Characteristics

To allow analysis of driving behavior activity exposure metrics with the specific facility types, driving activity records must be matched to roadway characteristics (facility type, posted speed limit, and rural/urban area type). The Commute Atlanta program research team at Georgia Tech developed an automated map-matching algorithm to combine GPS-based trip data with roadway characteristics (RC) information in the GIS system. The research team determined that the use of two map-matching methods, route method and buffer method, in combination provided the most complete and accurate data for driving activity analysis [5].

The University of Georgia (UGA) GIS Laboratory provided the most recent roadway maps of the 13 counties in the metro Atlanta region [5]. The UGA GIS Laboratory is continuously updating and managing the roadway network maps and roadway characteristics while under contract to Georgia Department of Transportation [5]. After the map-matching process, each GPS data point is associated with the corresponding roadway characteristics such as facility types, number of lanes, lane width, and posted speed limit. Finally, those map-matched GPS data profiles are used to examine driving behavior activities differentiating drivers who were not involved and involved in crashes.

Crash Data Collection

As one component of the Commute Atlanta Program, a safety-related survey was conducted in November 2004 to obtain the information on crash involvements during the study period (about 14 months between September 2003 and November 2004) [5]. The relevant questions in regards to the current project from the survey include the number of crashes during the study period. The questions in the self-reported crash survey were as follows:

1. Driver license status and duration licensed
2. Speeding tickets received
    a. In the last 5 years
    b. In the lifetime
3. Number of crashes they were involved in
    a. During the study period
    b. In the last 5 years
    c. In the lifetime
4. Number of at-fault crashes in their lifetime
5. Number of injury crashes in their lifetime
6. Speeding behavior
    a. Faster than the posted speed limit (yes or no)
7. Feeling about the speed limit
    a. Too high, about right, and too low
8. Seat belt usage
    a. Always, most of time, sometimes, rarely, and never

It is possible that the self-reported number of crashes during the study period could be underestimated since some drivers might not report their crash or simply forget to report minor crashes [4]. Since the official crash report database is well organized and managed by the federal or state agencies and can include all fatal crashes, the official crash database can be more popularly used. However, the self-reported crash data may include minor crashes that cannot be obtained from the official crash report database since motor crashes resulting in minor property damage and occurring at non-public roadways are usually not reported [33]. The self-reported crash data may also include other crashes that occurred at other states. Furthermore, Hauer et al [33] found that under-reported crashes in the official crash report also varied with accident type, location, and time of day.

Another merit of the self-reported crash data is that various information (socio-demographic information, speeding tendency, driver distraction, fatigue level, seat usage rate, average travel mileage, and passing behavior) and causes of crash involvements (speeding or fail to yield) can be additionally obtained. However, from the official crash report data, the only crash risk regarding age or gender can be evaluated since the other information on high crash-risk driving behavior is difficult to obtain.

This study compared the crash rates per licensed driver in the 13 county study area in 2002 (11.24%) [5] with the crash rate based on the self-reported crash survey (13.6%), indicating that the crash survey probably did not significantly underestimate the actual results [4] (Table 1).

---

[4] The researchers in Georgia Tech will request crash reports to DMVS in Georgia in order to verify the accuracy of the self-reported crash survey data under contract to participants in future.

Table 1: 2002 Georgia Crash Rates per Licensed Driver in 13 Counties, Atlanta [2]

County	Licensed Drivers	Drivers in Crashes	Rate (%)
Fulton	632,636	97,304	15.38
Clayton	186,789	23,921	12.81
Dekalb	522,929	64,798	12.39
Gwinnett	520,884	53,728	10.31
Douglas	82,583	8,439	10.22
Cobb	529,761	53,503	10.10
Henry	113,111	11,291	9.98
Rockdale	63,782	6,095	9.56
Forsyth	86,502	7,136	8.25
Coweta	76,062	5,967	7.84
Fayette	85,708	5,505	6.42
Cherokee	130,507	7,969	6.11
Paulding	71,967	2,986	4.15
Total	3,101,221	348,642	**11.24**

Crash Location Information in the State of Georgia

This study obtained crash location data from the Georgia Department of Transportation (GDOT) although the crash locations of participants could not be identified. The Department of Motor Vehicle Safety (DMVS) collects crash data and transfers them to the GDOT since the GDOT has responsibility for managing the safety level of roadways [5]. After receiving crash data, the GDOT develops crash location with the coded roadway network coordinates including roadway characteristic (RC) information. Figure 15 shows an example of crash locations in 13 counties occurred from 2000 to 2002 with crash and roadway information.

Figure 15: Crash Location and Related Information

Driver Selection

From this self-reported survey, this study was able to categorize drivers into two groups based on their crash involvements over the 14-months study period. Among the 234 drivers of all ages[5] who had instrumented-vehicles and returned the survey, drivers that shared a vehicle with another household member more than 10 % of the time were excluded because their personal driving trip data could not be adequately distinguished from that of other household members (Figure 16).

[5] The total number of participant who did the survey was 316, but some of them did not have their vehicles.

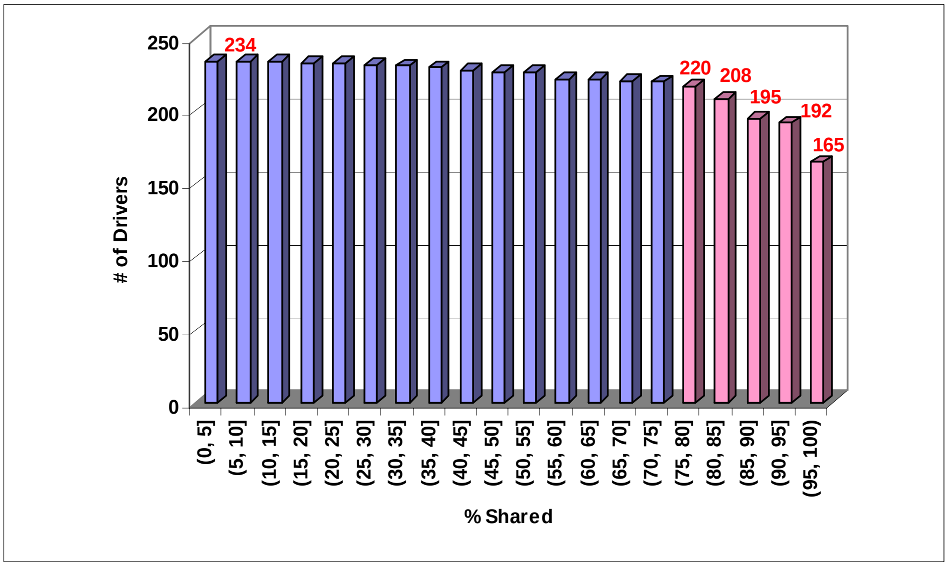


Figure 16: Number of Drivers based on % Shared

While this study collected crash involvement status during the 14-month period, the same period may not be used for collecting driving behavior activity data because drivers may change their driving behavior after being involved in a crash. Thus, driving behavior activity data longitudinally observed before the actual crash involvement date need to be evaluated for estimating potential crash risk of individual drivers. However, the self-reported crash survey did not include actual crash date and time, so it was difficult to decide the study period for evaluating driving behavior activity patterns. Thus, this study decided the study period for the driving behavior activity data based on the instrumentation status of GT-TDCs.

Figure 17 shows that the number of installed GT-TDCs had stabilized between January and June 2004. Activity data in December 2003 were excluded because drivers might have different patterns from their normal driving patterns during a holiday season. The 6-months study period possibly includes driving behavior activity data after crash involvements but may show more unbiased behavior activity patterns than shorter study period such as two- or three-month period. However, this study still agrees that the actual crash date should be employed and suggests that difference in driving behavior before and after crash involvements also need to be evaluated in future research.

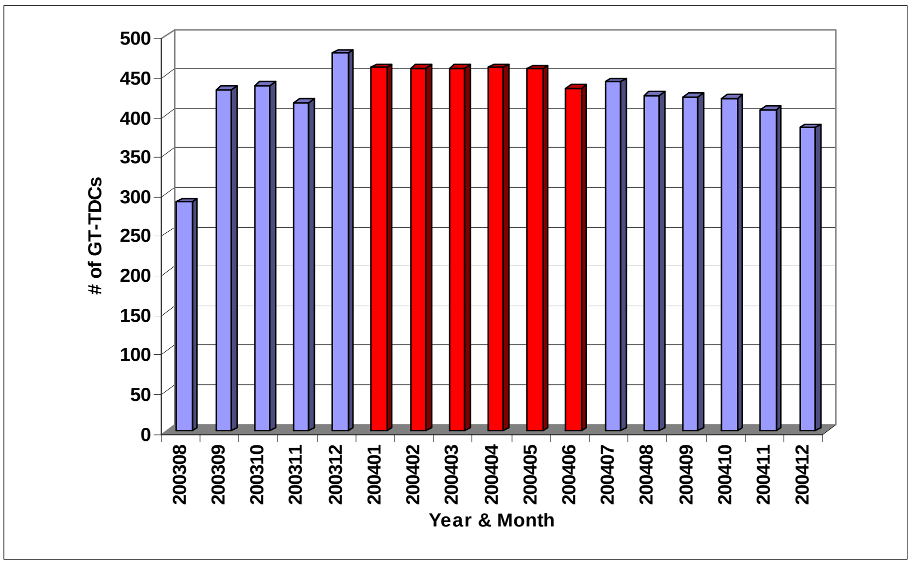


Figure 17: Numbers of GT-TDC in the Filed by Every Month

During the 6-months study period for collecting driving behavior activity data, some participants had purchased new vehicles, and several GT-TDCs had malfunctioned. Those GT-TDCs had to be replaced, so those drivers could not be used for this analysis because travel mileage and duration data could not be correctly obtained. After the data cleaning process, the study found 167 drivers of all ages who had been continuously monitored through a whole 6-months period (January through June 2004) for which survey data were available.

Since the self-reported crash survey used in this study included only crash involvement status of individual drivers during the 14-months study period and did not contain the detailed crash information regarding crash type, crash date, cause of crash, and severity, this study classified the driver data into two sets based on with and without crash involvements. Second-by-second travel data for the 6-months period (January through June 2004) were extracted from the Commute Atlanta program database to estimates disaggregate behavior exposures such as mileage, speeding, acceleration, unfamiliar roadway exposure, turn movement exposure, and previous crash location exposure and to identify the differences across the two driver-groups.

Chapter Four

DATA PROCESS AND QUALITY ASSURANCE

As discussed in previous chapters, among various incentive programs, pay-as-you-drive (PAYD) insurance and variable congestion tolls have been receiving increased attention from planners and transportation policy makers because the program will likely reduce vehicle usage rates and improve driving behavior to achieve safety benefits. Plus, on the average, such pricing programs should provide significant benefits to consumers through reduced insurance premiums. In implementing future programs, tracking of mileage, location of travel, and driving behavior activity will be an important variable [13].

As such, future use of GPS data beyond current freight logistics applications is likely to be instrumental to implementation of the most refined pricing programs. The accuracy of estimated mileage, speeds by road classification, and even acceleration rates based upon GPS data become paramount. As PAYD insurance programs are expected to assess insurance premiums based on travel mileage and driving speed, the goal of this study is to evaluate an individual driver's potential crash risk and to help set premiums that are proportional to such risk. Hence, this study needs (even insurance companies and customers in future) to ensure that reliable data are used in such analysis (or programs).

Evaluation on Acceleration Calculation Methods [6]

Of the various vehicle activity and driving behavioral metrics required for estimating crash risk of individual drivers, the most important components are the vehicle speed and acceleration profiles. To collect speed profiles, numerous data collection devices such as the distance measuring instrument (DMI), laser guns, onboard diagnostics systems (OBD), and satellite-based global positioning systems (GPS) can be employed. Many studies have also proven those data collection devices provide reliable and accurate speed data that can be useful in transportation research, although each type of device has its own set of unique limitations [5, 17, 30, 34-37]. Among these available devices, the GPS has been widely used in transportation and air quality research since it provides information as to individual vehicle activities as well as a complete trajectory of X-Y coordinates that identify where the vehicle activities occurred.

Acceleration is typically a derived value, calculated from consecutive speed measurement data. Systems that tap into the vehicle speed sensor can monitor speed at 2-10 Hz, providing high-resolution calculations of acceleration values based upon consecutive speed readings taken faster than once per second. GPS-based systems record satellite-signal-derived speed readings once per second. Because acceleration is calculated, acceleration values can differ depending upon the acceleration calculation method used.

The most common methods include the forward, backward, and central difference methods [38]. The definitions and concepts of these methods have been discussed in numerous mathematical references. However, the literature does not clearly identify

[6] Issues on the acceleration calculation method are in detail discussed in the paper written by Jun et al., "Impacts *of Acceleration Calculation Methods on Onroad Vehicle Engine Power Estimation*".

which acceleration calculation method is best-applied to estimate acceleration profiles in transportation research. In fact, researchers rarely even identify the method employed in their research and such research does not usually provide tables containing speed and acceleration data that help readers identify the method they used in their studies.

In addition, although many studies have used instantaneous GPS speed data to estimate acceleration profiles, this study found little literature that dealt with in detail acceleration calculation methods that compute accelerations from speed data and their effect on analytical results. Only one paper written by Sin [39] briefly discussed the impact of acceleration calculation methods on mobile emissions.

This study reviews acceleration calculation methods, discusses the characteristics of the acceleration profile derived from each method, and finally examines each distribution of accelerations based on the specific speed bin.

Reviews on Acceleration Calculation Methods

The basic idea of the forward difference method is that the acceleration at time (t_0) is the difference between the current (t_0) and the next ($t_0 + h$) speeds if the time interval is one second (Equation 1).

$$v'(t_0) = \frac{v(t_0 + h) - v(t_0)}{h} = a_F(t_0),\, h > 0, \tag{1}$$

where, a_F = acceleration derived by the forward difference method,

v = speed,

h = time interval.

Rakha et al. [40] used the forward difference method to calculate acceleration profiles for developing a vehicle kinetics model regarding maximum truck acceleration levels.

On the other hand, to calculate vehicle acceleration profiles, Rakha et al. [41], Sin [39], and Hallmark et al. [37] used the backward difference method (Equation 2). In contrast to the forward difference method, the acceleration at time (t_0) is the difference between the current (t_0) and the previous (t_0 - h) speeds when the time interval is one second.

$$v'(t_0) = \frac{v(t_0) - v(t_0 - h)}{h} = a_B(t_0)\,,\, h > 0, \tag{2}$$

where, a_B = acceleration derived by the backward difference method.

The last method, the central difference method, usually employs the three-point (Equation 3) or five-point methods (Equation 4). Although the five-point method provided smaller errors than the three-point method [38], this study discusses only the three-point method since the five-point method could decrease the sensitivity of driving behavior monitoring due to the larger smoothing effects. Recently, Yoon et al. [42] have used the three-point central difference method to estimate acceleration profiles for developing speed-acceleration matrices in a load-based mobile source emissions model.

Three-Point central difference method:

$$v'(t_0) = \frac{v(t_0 + h) - v(t_0 - h)}{2h} = a_C(t_0)\,,\; h > 0, \tag{3}$$

Where, a_C = acceleration derived by the central difference method

Five-Point central difference method:

$$v'(t_0) = \frac{v(t_0 - 2h) - 8v(t_0 - h) + 8v(t_0 + h) - v(t_0 + 2h)}{12h}\,,\; h > 0, \tag{4}$$

In addition, when a constant time interval such as second-by-second GPS speeds is used, the central difference method produces the average acceleration of the forward and backward difference methods as shown in Equation 5.

$$\begin{aligned} a_C(t_0) &= \frac{\{a_F(t_0) + a_B(t_0)\}}{2} \\ &= \frac{1}{2} \times \left\{ \frac{v(t_0 + h) - v(t_0)}{h} + \frac{v(t_0) - v(t_0 - h)}{h} \right\}, \\ &= \frac{v(t_0 + h) - v(t_0 - h)}{2h} \end{aligned} \tag{5}$$

Figure 18 illustrates how the acceleration at time (t_0) has different values (slope of each line) depending on the acceleration calculation method employed.

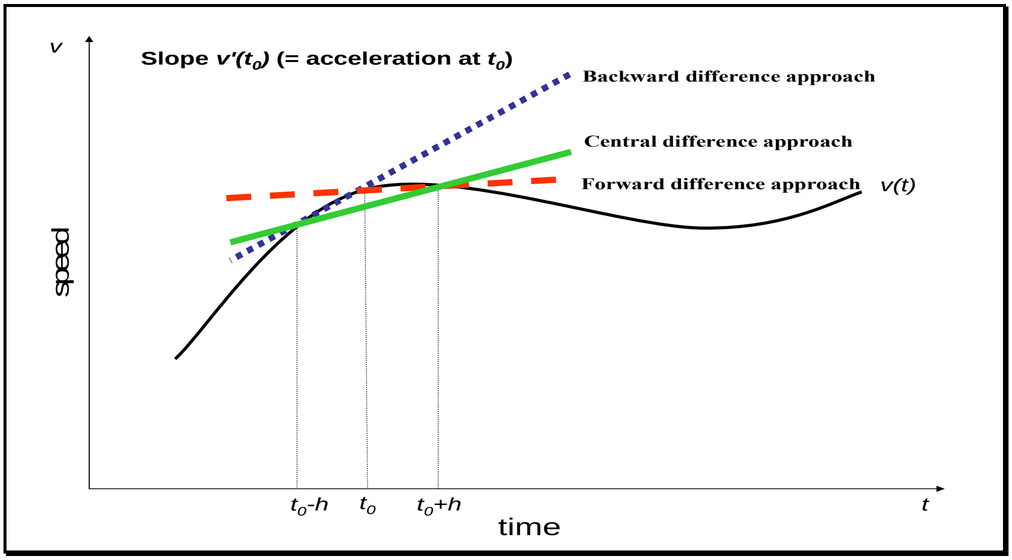


Figure 18: Graphical Interpretation of Three Acceleration Calculation Methods (Numerical Difference Methods)

In addition to about 460 light duty vehicles, the Commute Atlanta program installed four GT-TDCs in transit buses operated by the Metropolitan Atlanta Rapid Transit Authority (MARTA) from 2004 and 2005. Since transit buses usually produce more frequent acceleration activities than light-duty vehicles (LDV), this study used the second-by-second GPS speed data collected from Marta transit buses for the case study. As shown in Figure 19, the GPS antenna was installed on the roof of transit buses, the wireless cellular antenna was attached to the window, and the GT-TDS was installed inside the transit bus equipment cabinet.

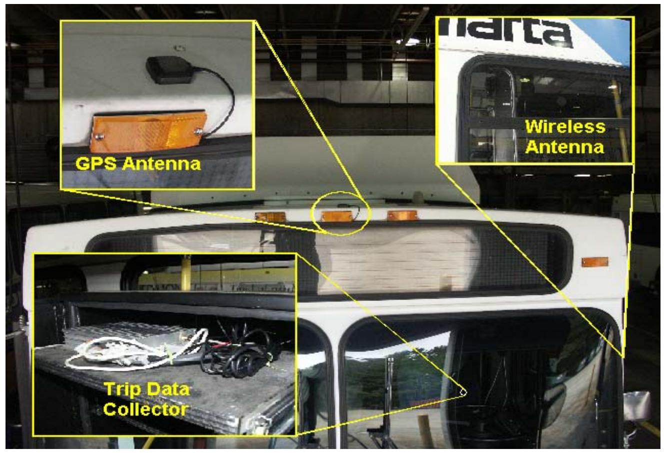


Figure 19: Installation of the TDC in a MARTA Transit Bus

To obtain reliable transit bus speed data, this study used only GPS speed data with a minimum of five satellites in view and where the position dilution of precision (PDOP) value was less than or equal to six [43]. In order to reduce the erroneous GPS data points, this study also adopted the acceleration threshold method using the maximum acceleration value of 10 mph/s (More detailed discussion on the automated filtering process about the GPS data will be provided in later in this chapter).

After performing the data cleaning process, this study selected 485,162 speed points for evaluating and analyzing the characteristics and the impact of accelerations estimated by each calculation method. With the selected GPS sample data, researchers estimated three acceleration profiles from the same consecutive GPS speed data using three different acceleration calculation methods.

Table 2 shows that acceleration profiles derived from the forward and backward methods are basically the same values since they estimate acceleration profiles based on the difference in continuous speeds at a given time interval. Thus, the general statistics such as the mean and the standard deviation of acceleration profiles derived from these two methods are theoretically equal when people consider only the characteristics of acceleration behavior. However, in case of speeds paired with the accelerations derived from these two methods, the binning of each second of vehicle activity into a speed/acceleration bin will differ depending upon the acceleration calculation method employed. Thus, the speed at time *t* has different acceleration depending on acceleration calculation method.

Table 2: Equations of numerical difference formulations

Time	Speed	Forward Method		Backward Method		Central Method	
		Acceleration	Pair	Acceleration	Pair	Acceleration	Pair
t_1	v_1	$a_1 = \frac{v_2 - v_1}{t_2 - t_1}$	(v_1, a_1)	-	-	-	-
t_2	v_2	$a_2 = \frac{v_3 - v_2}{t_3 - t_2}$	(v_2, a_2)	$a_1 = \frac{v_2 - v_1}{t_2 - t_1}$	(v_2, a_1)	$a_1^* = \frac{v_3 - v_1}{t_3 - t_1}$	(v2, a_1*)
t_3	v_3	$a_3 = \frac{v_4 - v_3}{t_4 - t_3}$	(v_3, a_3)	$a_2 = \frac{v_3 - v_2}{t_3 - t_2}$	(v_3, a_2)	$a_2^* = \frac{v_4 - v_2}{t_4 - t_2}$	(v3, a_2*)
t_4	v_4	$a_4 = \frac{v_5 - v_4}{t_5 - t_4}$	(v_4, a_4)	$a_3 = \frac{v_4 - v_3}{t_4 - t_3}$	(v_4, a_3)	$a_3^* = \frac{v_5 - v_3}{t_5 - t_3}$	(v4, a_3*)
t_5	v_5	-	-	$a_4 = \frac{v_5 - v_4}{t_5 - t_4}$	(v_5, a_4)	-	-

Table 3 shows an example indicating how each method produce different acceleration values based on consecutive speed values and indicates that vehicle operation mode (e.g., cruise, acceleration, or deceleration mode[7]) can differ based on which acceleration calculation method is adopted.

[7] Cruise Mode: -1 < Acceleration < 1
Acceleration Mode: 1 <= Acceleration
Deceleration Mode: -1 >= Acceleration

Table 3: Example of speed and acceleration combinations

Speed (mph)	Acceleration (mph/s)					
	Forward	Backward	Central	Forward	Backward	Central
26.65	0.83	1.32	1.08	**Cruise**	**Acceleration**	**Acceleration**
27.48	-0.67	0.83	0.08	Cruise	Cruise	Cruise
26.81	-1.32	-0.67	-0.99	**Deceleration**	**Cruise**	**Cruise**
25.49	-2.53	-1.32	-1.93	Deceleration	Deceleration	Deceleration
22.96	-0.63	-2.53	-1.58	**Cruise**	**Deceleration**	**Deceleration**
22.33	-1.40	-0.63	-1.01	**Deceleration**	**Cruise**	**Deceleration**
20.93	-1.70	-1.40	-1.55	Deceleration	Deceleration	Deceleration
19.23	-0.83	-1.70	-1.26	**Cruise**	**Deceleration**	**Deceleration**
18.40	-1.68	-0.83	-1.25	**Deceleration**	**Cruise**	**Deceleration**
16.71	-1.68	-1.68	-1.68	Deceleration	Deceleration	Deceleration
⋮	⋮	⋮	⋮	⋮	⋮	⋮
6.89	-6.89	-3.33	-5.11	Deceleration	Deceleration	Deceleration
⋮	⋮	⋮	⋮	⋮	⋮	⋮
17.84	-0.34	-0.04	-0.19	Cruise	Cruise	Cruise
17.50	-1.09	-0.34	-0.71	**Deceleration**	**Cruise**	**Cruise**
16.40	-1.48	-1.09	-1.28	Deceleration	Deceleration	Deceleration
⋮	⋮	⋮	⋮	⋮	⋮	⋮
5.68	-1.52	-3.85	-2.69	Deceleration	Deceleration	Deceleration
4.16	0.98	-1.52	-0.27	**Cruise**	**Deceleration**	**Cruise**
5.14	1.79	0.98	1.39	**Acceleration**	**Cruise**	**Acceleration**
6.93	1.97	1.79	1.88	Acceleration	Acceleration	Acceleration
⋮	⋮	⋮	⋮	⋮	⋮	⋮
16.33	1.90	2.93	2.42	Acceleration	Acceleration	Acceleration
18.24	0.61	1.90	1.25	**Cruise**	**Acceleration**	**Acceleration**
18.85	0.80	0.61	0.70	Cruise	Cruise	Cruise

Figure 20 visually shows the result of the combinations with speeds measured by GPS and accelerations derived by three different methods. The ellipses in Figure 20 illustrate how accelerations are paired with different speeds as the forward and backward difference methods are used. The vertical lines at each speed point show how each speed point has different acceleration values. High accelerations produce large difference between the paired speeds, and difference between the paired speeds decreases as accelerations become smaller. This result indicates that the amount of difference of accelerations between the forward and backward methods increases as a driver transitions from deceleration to acceleration, or vice-versa. Forward and backward difference methods couple the highest acceleration rates with different speed values, potentially placing the activity into a different speed/acceleration bin. The results indicate that researchers employing joint speeds and acceleration data, such as a vehicle emissions rate or a driver behavior analyses, need to pay careful attention to the potential impacts that acceleration calculation methods may have on their derived speed/acceleration profiles.

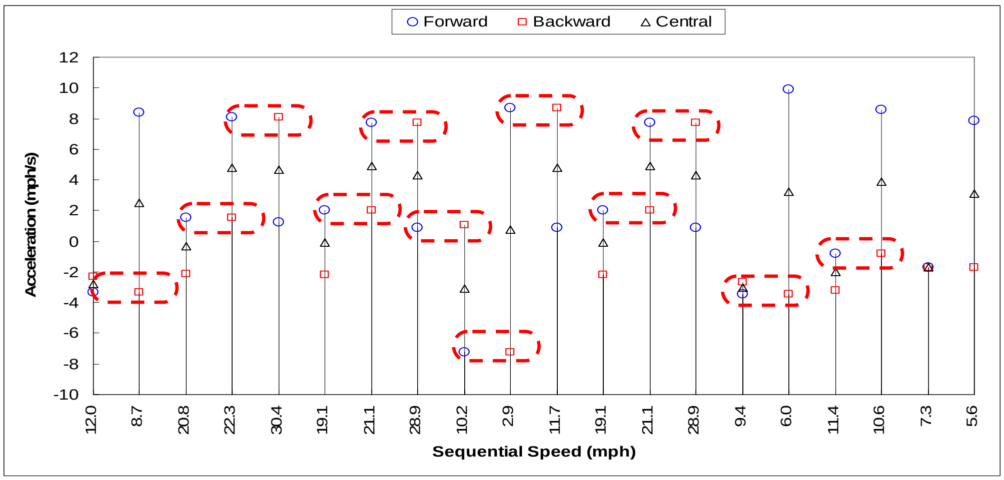


Figure 20: Comparison of Large Accelerations and Decelerations with Speeds

Based on these findings, four operating regions can be generated (Figure 21). Region 1 indicates that all methods indicate the same cruise driving mode having low acceleration values (-1 < acceleration < 1). Regions 2 and 3 indicate that all methods indicate the same acceleration and deceleration driving modes, respectively, although the magnitude of accelerations or decelerations can differ. Finally, region 4 indicates that all methods (or two of them) can produce different driving operation modes. For example, the forward difference method produces the deceleration mode, but the backward difference method can produce the acceleration mode or cruise mode.

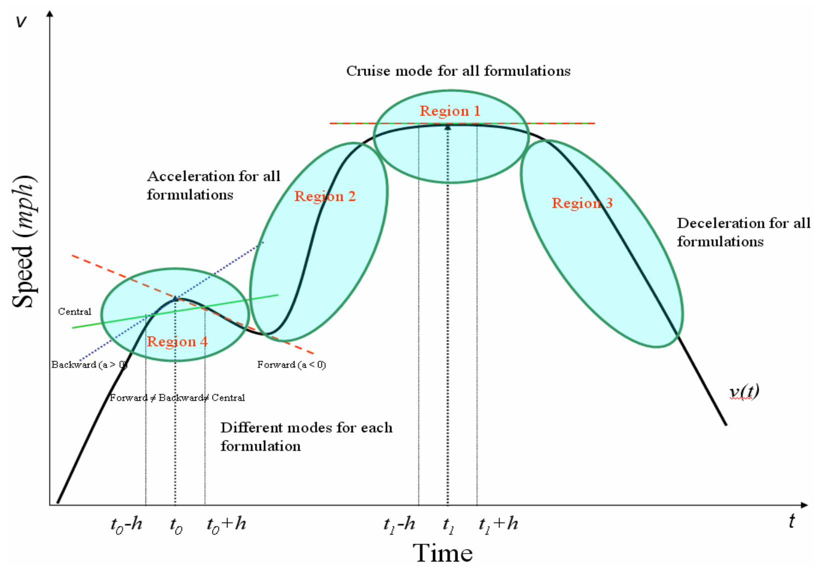


Figure 21: Four Regions of Different Acceleration Profiles

With a complete data stream (no speed values excluded through the data screening process based upon satellites in view and PDOP values), the acceleration distributions will be identical and are shifted by one second. However, if data screening is employed, maximum and minimum acceleration rates derived by the forward, backward, and central difference methods will differ. In this case of the MARTA bus data, the forward, backward, and central difference methods yielded 9.89, 8.72, and 5.05 mph/s as maximum acceleration dates and -8.89, -8.89, and -8.63 mph/s as maximum deceleration rates, respectively. While this is not much important in most applications, if a large amount of data were missing from the data stream, different results could arise.

Although this study found that each acceleration calculation method produce different acceleration profiles through the visual inspections and mathematical values, it is still required to verify whether they truly produce different acceleration profiles in terms of their distributions. The study created scatter plots on the speed profile. Figure

22 shows the backward and the central acceleration method provide the negative acceleration (or deceleration) at the zero speed due to their theoretical backgrounds.

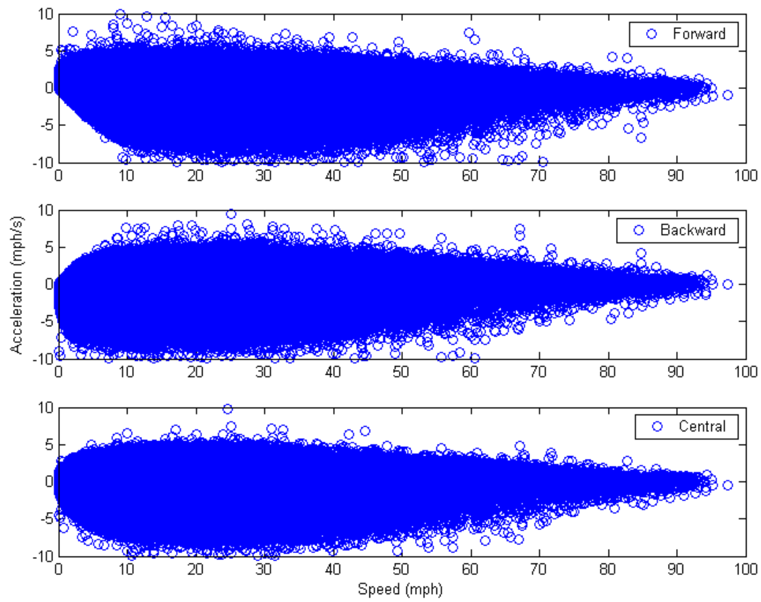


Figure 22: Scatter Plots of Acceleration and Speed Values

This study also performed the chi-square test to verify their heterogeneous distributions. Using the speed profile measured by GPS and acceleration profiles derived from three methods, a two-dimensional contingency table with a 5 mph speed increment and 0.5 mph/s acceleration increment bin widths was created. The hypothesis for testing the homogeneity is formulated as follows:

$$H_0:\ F(x) = F(y)$$

$$H_1:\ F(x) \neq F(y)$$

Table 4 shows the result of the test of the homogeneity of acceleration estimates in each speed interval and indicates that all chi-square statistics are significantly greater than the critical value except for the high speed greater than 60 mph. In addition, the difference between the chi-square test value and the critical value becomes smaller as a speed increases because high speeds usually produce low acceleration values. However, this study can conclude that each acceleration calculation method provides a different distribution of acceleration estimates.

Table 4: The Result of Chi-Square Test (α = 0.05) (Acceleration Distribution)

	$0 \leq$ speed <5			$5 \leq$ speed <10			$10 \leq$ speed <15		
	F vs. B	B vs. C	C vs. F	F vs. B	B vs. C	C vs. F	F vs. B	B vs. C	C vs. F
Test value	5163.73	1672	1834	751	364.3	430.7	364.89	328.7	199.38
Number of groups	17			23			26		
Critical value	**26.3**			**33.92**			**37.65**		
	$15 \leq$ speed <20			$20 \leq$ speed <25			$25 \leq$ speed <30		
	F vs. B	B vs. C	C vs. F	F vs. B	B vs. C	C vs. F	F vs. B	B vs. C	C vs. F
Test value	360.9	232.7	157.3	507.54	232.8	190.2	431.18	267.7	162.7
Number of groups	24			27			21		
Critical value	**35.17**			**38.89**			**31.41**		
	$30 \leq$ speed <35			$35 \leq$ speed <40			$40 \leq$ speed <45		
	F vs. B	B vs. C	C vs. F	F vs. B	B vs. C	C vs. F	F vs. B	B vs. C	C vs. F
Test value	575.86	319.6	285.35	346.8	219.6	162.8	186.2	78.2	78.76
Number of groups	21			21			17		
Critical value	**31.41**			**31.41**			**26.3**		
	$45 \leq$ speed <50			$50 \leq$ speed <55			$55 \leq$ speed <60		
	F vs. B	B vs. C	C vs. F	F vs. B	B vs. C	C vs. F	F vs. B	B vs. C	C vs. F
Test value	77.94	60.39	86.19	31.24	44.72	62.34	22.23	34.87	40.2
Number of groups	16			10			9		
Critical value	**25**			**16.92**			**15.51**		
	$60 \leq$ speed <65			-			-		
	F vs. B	B vs. C	C vs. F						
Test value	9.06	18.1	4.44						
Number of groups	4								
Critical value	**7.81**								

F: Forward Difference Method, B: Backward Difference Method, C: Central Difference Method

Jun et al. [18] investigated impacts of these calculation methods on the estimation of engine power using a chi-square test showing that all three methods reported statistically different distributions of engine power estimates. Briefly speaking, engine power estimates with the backward difference method significantly differed from the forward difference method. Difference between forward and backward difference methods was 4.4% in the case of the MARTA bus power demand analyses. Meanwhile, the central difference method reported almost averages of the difference engine power estimates between forward and backward methods. This result implies that one can estimate about 4 % less emissions rates or 4 % higher emissions rates although they both use the same activity data set.

Summary on Acceleration Calculation Methods

This study examined the characteristics of three acceleration calculation methods and investigated how these different methods produced different results on engine power estimates as a case study using GPS data [18]. Because these methods employ different calculation approaches to estimate acceleration values, each acceleration value will be combined with a different speed point (conversely, each speed point will be assigned a different acceleration value). Based on the chi-square test, accelerations derived by each method have significantly different bin distributions. Thus, the results of this case study imply that researchers should note which acceleration computation method is implemented in their research, understand the differences of acceleration computation methods, and be careful in particular when the estimates are obtained from speed-acceleration combinations.

Quality Control of GPS-Measured Data [8]

The quality of GPS-measured data and the control method of some erroneous GPS data should be discussed and developed. Among various data collection devices, the GPS has been the most common choice in transportation research (including PAYD programs), because it provides more useful data, such as travel routes, start and stop points of a trip, travel time, speed, and acceleration rates. Although an accurate data measurement device, as shown in previous studies [30, 34], the GPS is still subject to various systematic and random errors:

- Systematic errors may be due to a low number of satellites, a relatively high Position Dilution of Precision (PDOP) value which relates to satellite orientation on the horizon and the impact on position precision, and other parameters (for example, antenna placement) that affect precision and accuracy of the device used [44].
- Random errors may result from satellite orbit, clock and receiver issues, atmospheric and ionospheric effects, multi-path signal reflection, and signal blockage [30, 34].

While systematic errors can be readily identified and removed, random errors are more difficult to address. Depending upon how the GPS data will be used, and upon the magnitude of the random error effect, it may be necessary to process the GPS data to minimize the effects of random error for some processes in which the data will be

[8] Issues on the automatic filtering methods are in detail discussed in the paper written by Jun et al., "*Smoothing Methods Designed to Minimize the Impacts of GPS Random Errors on Travel Distance, Speed, and Acceleration Profile Estimates*".

employed. Although in smaller research efforts, GPS errors can be identified through visual inspection of the data, in deployments that yield large GPS data sets, visual inspection is not practical. Due to significant data processing time, automated analysis techniques are required. Statistical smoothing techniques may be useful processing tools since they are designed not only decrease the impact of random errors on the results of the study but also require less time for detecting random errors than visual inspection.

Bench Test of GPS data

This study collected 114,014 GPS speed data points at a stationary location (Georgia Tech) for about 32 hours and estimated the speed errors since the actual speed in this bench test was known, which equaled to zero. Figure 23 shows the histogram and the cumulative density plot of speed errors. This bench test used the GPS speeds showing a good quality with at least four satellites and PDOP values less than eight. From the bench test of the GPS data, the average speed error was 0.25 mph and the maximum speed error was 2.15 mph. The probability of speed errors less than 1 mph was 0.99 (99%) of the time.

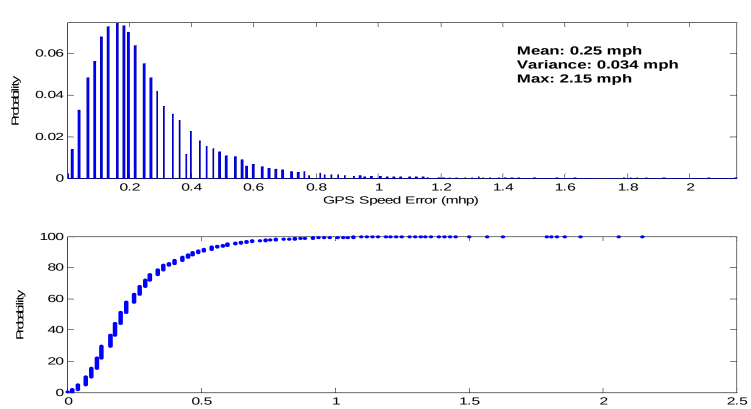


Figure 23: Bench Test Results for GPS Speed

However, it should be known that the GPS speed errors may be different in the real world condition. Using 1,510,105 GPS-measured second-by-second data points from ten vehicles, the distribution of number of satellites in the real travels was estimated (Figure 24). About 11.79 % of all GPS data points had less than four satellites, which could be considered unreliable data. In addition, since the GPS can provide random errors in data profiles even with good satellite signals, the potential range of GPS errors can be larger in the real world condition.

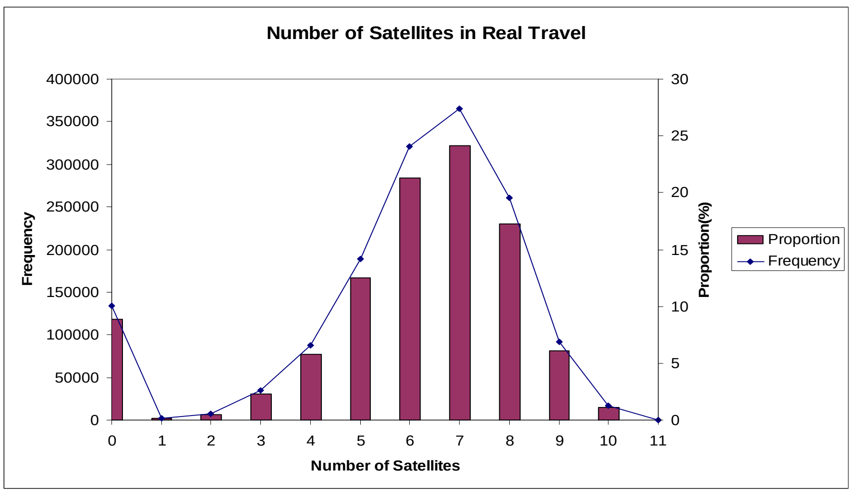


Figure 24: Distribution of Number of Satellites in Real World Condition

Thus, to minimize the impact of random errors on speed, acceleration, or travel distance estimate, researchers should add a supplemental correction process after collecting the GPS data using the various statistical smoothing methods since these methods decrease the impact of random errors on the results of the study and take less time to detect random errors than visual inspection.

Although in smaller research efforts, GPS errors can be identified through visual inspection of the data, in deployments that yield large GPS data sets, visual inspection is not practical. Due to significant data processing time, automated analysis techniques are required. Statistical smoothing techniques such as the Kalman filter may be useful processing tools since they are designed not only decrease the impact of random errors on the results of the study but also require less time for detecting random errors than visual inspection [44].

Another technique is to use a specific acceleration threshold. Using this method, researchers can filter out GPS data having extremely higher acceleration values [18, 39], and this technique was used for evaluating the acceleration calculation methods above.

Statistical smoothing techniques can be categorized by their statistical backgrounds into three types: the first is to minimize overall error terms, the second is to adjust the probability of occurrence, and the last type is to recursively perform feedback system. Although each approach is capable of detecting random errors in the GPS data profiles, given their different statistical backgrounds, each technique can result in different outputs. Thus, before adopting a specific smoothing technique for identifying random errors in the GPS data profile, researchers need to better understand their characteristics. This study describes the characteristics of three smoothing techniques that are popularly used in a variety of traffic-related research and also have different statistical algorithms or backgrounds: the least squares spline approximation, the kernel-based smoothing method, and the Kalman filter.

- The least squares spline approximation minimizes the residual sum of squared errors (RSS) and has a statistical background similar to regression-based smoothing techniques such as the local polynomial regression, cubic fits, robust exponential smoothing, and time series models

- The kernel-based smoothing method adjusts the probability of occurrences in the data stream to modify outliers and has the same statistical background as nearest neighbor smoothing and locally weighted regression models
- The Kalman filter, smoothes data points by recursively modifying error values

This study evaluates one smoothing method within each general category of smoothing techniques. Each smoothing technique is applied to a large GPS data set collected in Atlanta, GA and then comparatively evaluated for the impact on estimated speeds, accelerations, and travel distance profiles. While not exhaustive, the study believes that the three general smoothing approaches examined are representative of each general statistical approach.

In addition, the study believes that the evaluation of smoothing techniques for the GPS data and better understanding of their statistical performance is necessary for transportation researchers because inexpensive GPS receivers (non-RTK-GPS systems) will be employed in large-scale deployments.

Statistical Smoothing Techniques

The basic principle of smoothing techniques is to augment or reduce erratic data points by replacing the value of input variables [45]. Erratic location and speed data recorded from the GPS receiver can lead to erroneous determinations on acceleration values. Most GPS receivers, including the SiRF Star II, employ a proprietary filtering algorithm to compensate for data points beyond known variances [5, 30]. That is, the device software embedded within the receiver automatically provides some level of data correction. Additional measures of reliability are included in the data stream to help identify questionable data. Researchers have developed numerous techniques to filter the data based on these measures with some degree of success. However, regardless of these smoothing and filtering algorithms, the proprietary filtering algorithms cannot filter all outliers, as evidenced by random errors that are still present in the GPS output data stream.

To minimize the impact of random errors on speed, acceleration, and travel distance estimates, this study propose a supplemental smoothing process for post-collection GPS analysis. Without the full identification and correction of random GPS errors, researchers cannot reasonably evaluate driver acceleration and deceleration behaviors and travel mileage. This study evaluates three statistical smoothing techniques and compares their capabilities minimizing the GPS random errors in the data streams.

<u>Least squares spline approximation</u>

The least squares spline approximation, or the so-called "piecewise polynomial regression model," divides the data set (Y_i) into several pieces with a pre-determined width (or interval) and estimates predictors ($\hat{Y}_i$) using the residual sum of squared errors [45, 46]. The local polynomial regression model derives a regression function from each localized data set using Equations 6 and 7. Equation 7 measures the residual sum of squared errors (RSS) and estimates each parameter ($\beta_0, \dots \beta_{d-1}$) within each interval.

$$\hat{f}(X) = \beta_0 X^0 + \beta_1 X^1 + \beta_2 X^2 + \ldots + \beta_{d-1} X^{d-1} + \varepsilon , \tag{6}$$

$$RSS(\hat{f}) = \sum_{i=1}^{n} \{Y_i - \hat{f}(x_i)\}^2 , \tag{7}$$

where d is an order (or degree) of the function, and n is the sample size within the selected interval [45].

To evaluate the ability of the least squares spline approximation as a smoothing method, researchers must decide the bandwidth representing the interval of the local data set and the order (or degree) of the regression function. The one-second and two-second intervals have only one and two GPS data points, respectively. These intervals conceptually do not have sufficient data points for the polynomial model (one or two GPS data points cannot be smoothed by the smoothing algorithm). As bandwidths increase, they contain larger numbers of data points, and filtering may yield speed estimates for which some of the actual speed variability is smoothed away. Thus, this evaluation used a three-second interval to avoid rapid increases and rapid decreases in acceleration rates calculated via change in speed over two consecutive seconds. In the case of order selection, since this study selected three-second interval, the quadratic function (d = 3) is selected as the order of the regression function.

Kernel-Based Smoothing Method

The kernel-based smoothing method assigns a weight (or a smoothing parameter) using the kernel density estimator [45]. To obtain this estimator, the study uses the Gaussian kernel estimator in Equations 8 [45, 46] and estimates the smoothing curve using the Nadaraya-Watson kernel smoothing algorithm in Equations 9 [45, 46], as follows:

$$K_h(X_i, x) = K\left(\frac{|X_i - x|}{h}\right) = (2\pi h^2)^{-\frac{1}{2}} e^{\left\{-0.5\times\left(\frac{X_i - x)}{h}\right)^2\right\}}, \tag{8}$$

$$\hat{f}_{NW(x)} = \frac{\sum_{i=1}^{n} K_h(X_i - x)\hat{Y}_i}{\sum_{i=1}^{n} K_h(X_i - x)}, \tag{9}$$

where h is the kernel bandwidth that controls the width of the localized data set, and $K(t)$ is a kernel function that satisfies the following condition:

$$\int K(t)dt = 1, \tag{10}$$

Kernel-based smoothing method also requires bandwidth selection. Although the correct width (h) is not simply selected, and various references for selecting the kernel width exist, the normal reference rule in Equation 11 can be used in this study because of its relative simplicity [45]. Bandwidths from the normal reference rule are between two-second and four-second intervals based on the initial sample test.

This study uses a three-second bandwidth for the kernel-based smoothing for two reasons: 1) the four-second bandwidth significantly degrades the capability of the GPS data to be smoothed, and 2) the least squares spline approximation also uses the three-second interval. Sin (11) also used the three-second interval as the bandwidth parameter for evaluating the Epanechnikov kernel smoothing method and showed that this three-second interval produced the best overall results.

$$h = (\frac{4}{3})^{1/5} \sigma n^{-1/5} \approx 1.06 \sigma n^{-1/5}, \tag{11}$$

where h is the bandwidth, σ is the standard deviation, and n is the number of data points.

Discrete Kalman Filter

The final smoothing method in this study, the discrete Kalman filter, recursively estimates outputs using the feedback system in Figure 25 [47].

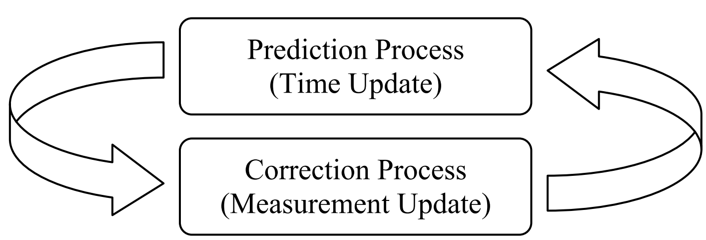


Figure 25: The Kalman Filter Cycle

To perform the feedback system, the Kalman filter uses two processes: the prediction process (or the time update) and the correction process (or the measurement update) and initially estimates a one-step predictor (a priori predictor) from the prediction process and obtains the correction (a posteriori predictor) from the correction process [47-49].

The time update equations are

$$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_k , \tag{12}$$

$$P_k^- = AP_{k-1}A^T + W , \tag{13}$$

where k is the time step, $\hat{x}_{k-1}$ and P_{k-1} are the initial predictor and the initial error noise, respectively, u_k is an additional known-input parameter, W is the prediction error variance, which is the Gaussian noise: $N(0,Q)$, and A and B are the time transition matrices for the prediction process [47-49].

Since this study uses only GPS unit as a measurement device and separately tests the Kalman filter for smoothing speed (and therefore acceleration) and trip location points (X and Y coordinates), u_k in Equation 12 becomes zero (the one-dimensional Kalman filter). In addition, this study uses the second-by-second GPS speed data, therefore, the time transition matrix, *A*, is one second. Thus, Equations 12 and 13 are reduced to the following form:

$$\hat{x}_k^- = \hat{x}_{k-1}, \tag{12}$$

$$P_k^- = P_{k-1} + W, \tag{13}$$

The measurement update equations are

$$K_k = P_k^- H^T (H P_k^- H^T + V)^{-1}, \tag{14}$$

$$\hat{x}_k = \hat{x}_k^- + K_k (z_k - H\hat{x}_k^-), \tag{15}$$

$$P_k = (I - K_k H) P_k^-, \tag{16}$$

where K is the Kalman gain matrix, H is the time transition matrix for the observation process, z is the observed data, P is the modified error variance in the Kalman filter, and V is the measurement error variance, which is the Gaussian noise: $N(0, R)$.

Similar to the above reduced equations, the measurement update equations can also be reduced:

$$K_k = P_k^- (P_k^- + V)^{-1}, \tag{14}$$

$$\hat{x}_k = \hat{x}_k^- + K_k (z_k - \hat{x}_k^-), \tag{15}$$

$$P_k = (I - K_k) P_k^-, \tag{16}$$

Just as the least squares spline approximation and the kernel based smoothing method required a bandwidth value and the order of the function prior to conducting the smoothing process, the Kalman filter requires values for the measurement noise (R) and the process noise (Q).

The Modified Kalman filter

Although the correct value of the measurement noise for the Kalman filter is not easily determined, previous studies [48, 49] suggested using the square of the mean error value from a manufacturer's technical specification. For smoothing vehicle location (X-Y coordinates), this study used 100 feet (10^2 feet) as the measurement noise (R) [48-50]. However, researchers should understand that this mean error in the manufacturer's technical specification was estimated in the perfect GPS condition, which means that this value does not truly indicate the mean of errors in real-world conditions.

In the case of speed profiles, this study compared 1,171,496 GPS-measured speeds and corresponding VSS-derived speeds over a two-month period and estimated the mean delta speed to be 0.5 mph. Thus, this study used 0.25 mph (0.5^2 mph) as the GPS speed measurement noise. Given a 1 Hz data capture rate, the process noise of locations was the same as the measurement noise (1^2 second × 10^2 feet) and the process noise of speeds was also same as the measurement noise of speeds (1^2 second × 0.5^2 mph).

Here, another critical problem occurs when researchers use the measurement noise associated with location and speed data. The quality of the GPS data strongly depends on the GPS signal condition, usually represented by the number of satellites and PDOP values. When the condition of the GPS signal does not reach the sufficient level of minimum requirement, such as at least four satellites in view and PDOP values less than or equal to eight, the measurement errors are much greater than the above estimates.

In addition, the most important component of the Kalman filter is the measurement error since the measurement error determines how much random GPS random error should be reduced. Thus, this study modified the conventional discrete Kalman filter by using two measurement errors based on the GPS quality criteria, the number of satellites and PDOP values. This study estimated the first measurement error in the conditions of at least four satellites in view and PDOP values less than or equal to eight and the second measurement error from the other GPS signal conditions.

Based on this approach, this study used 10^2 degree (690^2 mile) as the measurement error of X-Y coordinates based on the result of preliminary evaluations and also used 10^2 mph of the measurement error for the speed profiles in the bad GPS signal conditions such as the loss of GPS signal lock.

Analyses on Filtering Techniques

This study evaluated three smoothing techniques to discern their effect on minimizing random GPS errors prior to calculating speed, acceleration, and distance profiles. Since reliable acceleration profiles can be derived from reliable speed profiles, both speed and acceleration profiles were tested by each smoothing technique. For evaluating travel distance profiles, this study conducted smoothing techniques to second-by-second X-Y coordinates and estimated travel distances. To compare all outputs produced by each technique and to verify their effectiveness, this study used speeds, accelerations, and distance profiles derived by the vehicle speed sensor as supplemental measurements.

Most previous studies of smoothing techniques generally tended to compare the original GPS data with the filtered GPS data estimated by smoothing techniques, primarily because they did not have alternative source of data, or ground truth. This study compared speed profiles obtained by the GT-TDC from the GPS receiver, the vehicle speed sensor monitor, and the onboard diagnostics (OBD) system (note that speed values from the VSS and OBD originate from the same source [5], transaxle rotation sensors, but are monitored and processed at different frequencies).

Quality Control of VSS Speed Data

Numerous studies have compared vehicle speed profiles measured by various data measurement devices such as the global positioning system (GPS) and the distance measuring instrument (DMI). For example, Zito et al. [34] used an instrumented vehicle equipped with the Australian Road Research Board (ARRB) travel time data acquisition system (TTDAS) to compare the GPS speed and acceleration profiles with those obtained directly from the speedometer in that vehicle. Ogle et al. [30] compared the GPS speed profiles, measured by four Palm Pilot devices, with speed profiles measured by a Nu-Metrics Nitestar NS-60 distance measurement instrument (DMI). In addition, Jackson et al. [35] recently compared the GPS-measured speed profile with the speed profile obtained from the on-board diagnostics (OBD) system in a vehicle.

According to these previous studies, Zito et al. [34] found that the average speed error and the standard deviation for central business district (CBD) area was 0.6 kph and

4.2 kph, and the average speed error and standard deviation for rural area was 0.21 kph and 1.35 kph. Ogle at al. [30] also found that GPS units provided 1.1mph of speed difference from the DMI unit. From these comparative studies, researchers determined that the speed profiles measured by each device did not significantly differ and that all data measurement devices could be used in transportation-related research depending on the purpose of research to collect speed profiles.

Although the previous studies compared the accuracy of speed profiles measured by different data measurement devices, they did not investigate various error sources in each device that could systematically or randomly occur in speed profiles, nor did they discuss in detail the impact of these errors on the quality of the speed profiles measured by each measurement device.

For example, since the speed profile derived by the vehicle speed sensor (VSS) systematically depends on tire diameter, which fluctuates as a function of air pressure and temperature inside the tires, the VSS-derived speed profile can naturally contain a certain amount of bias caused by the fluctuation of the tire diameter if the tire diameter was not correctly modified as the GPS also provides random errors that often cause unrealistic accelerations in the speed profile.

In addition, these previous studies estimated the speed difference using the overall average or the standard deviation in all speed ranges. Since GPS receivers inherently produced random errors at low speeds (a manufacture's smoothing algorithm embedded in receivers are not efficiently operated due to the poor heading measurements at low speeds) [30], the reliability of the GPS speed should be analyzed based on specified speed intervals.

Therefore, these systematic and random errors in the speed profiles measured by both the VSS and GPS should be addressed and minimized before each speed profile is compared or when researchers plan to use them. This study discusses the systematic and random errors in detail in the GPS-measured and VSS-derived speed profiles and investigates the impact of these errors on the speed profiles. In addition, this study evaluates the speed difference between them based on specified speed intervals and also investigates the reliability of acceleration profiles.

As discussed in Chapter 3 (Data Collection), the VSS monitors the number of electronic pulses which counts the number of driveshaft revolutions [30] and converts the number of electronic pulses into the speed profile based on a function of the tire diameter. To derive speed profiles from the number of electronic pulses measured by the VSS, the calibration number which is the number of pulses for traveling a distance of one mile is required.

In general, vehicle manufactures use their own standard calibration numbers such as 2,000 pulses/mile, 4,000 pulses/mile, or 8,000 pulses/mile, based on vehicle and VSS types. However, after a person purchases a vehicle, he/she can change a tire having bigger or smaller size than the standard tire size. In addition, the tires of the vehicle can loose air pressure, resulting in different size from the standard ones. Figure 26 shows the examples of un-calibrated VSS speeds estimated by these standard calibration numbers from two vehicles.

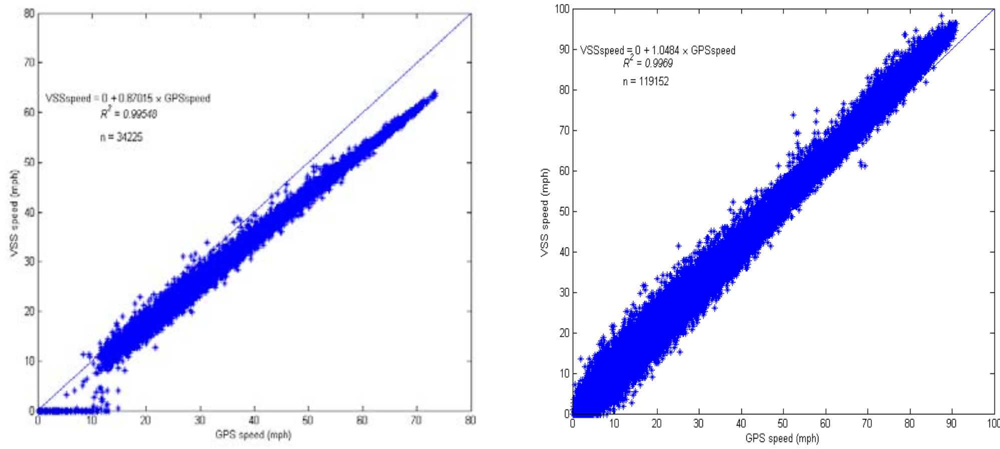


Figure 26: Un-calibrated VSS Speeds vs. GPS Speeds

As shown in Figure 26, variation in tire size or air pressure inside tires from the manufacturer specification changes these numbers and causes overestimated or underestimated speeds [5]. To obtain specific calibration numbers for each vehicle, researchers should directly measure the circumstance of tires of them, but it is practically hard to accomplish in this study since more than 460 vehicles are used in the Commute Atlanta program and they are also distributed in thirteen counties in Metropolitan of Atlanta.

Thus, this study estimated the specific calibration number producing the highest Pearson correlation coefficient between the GPS and the VSS speeds of each vehicle. In addition, regression analysis, $Y = \beta_1 X$, was used for each vehicle to validate this number since the coefficient of regression, i.e., β_1, should be close to 1 if the correct calibration number was selected. From this process, each vehicle had the correct calibration number better representing its tire conditions. Figure 27 shows examples of calibrated VSS speeds from the same vehicles.

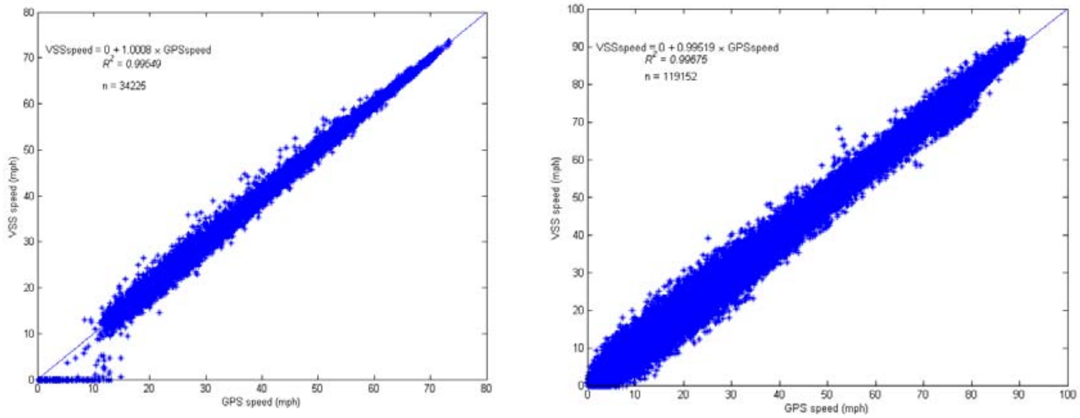


Figure 27: Calibrated VSS Speeds vs. GPS Speeds

Since the tire diameter also fluctuates as a function of air pressure inside the tires, the VSS can produce systematic errors caused by tire inflation if the tire diameter is not correctly modified. To decrease systematic errors in the VSS-derived speeds, Benz et al. [36] suggested a weekly calibration of both tire diameter and tire pressure. However, since the tire diameter can be still changed every day due to the ambient temperature [51], weekly calibration is not enough to decrease systematic speed errors. For example, the study conducted by Tire Tech [51] showed that tire pressure increased about 1 psi every 10º Fahrenheit that air temperature increased.

In addition, tire diameter can be changed even during a trip since the air pressure inside the tire increases as the driving speed (or travel distance) increases. Tire Tech [51] evaluated how the air pressure inside a tire related to the driving speed and found that the air pressure increased approximately 1.5 psi every 10 kph (6.2 mph) that speed increased until tire inflation reached the maximum standard pressure of the tire. Therefore, when the VSS is used to measure speed data, the fluctuation of the tire diameter during a trip also need to be considered. Incorrect tire diameter results in an erroneous calibration factor, which causes unreliable speed estimates containing systematic errors.

Although the study [51] indicated that driving speeds increase the air pressure inside a tire, this study did not demonstrate in detail how the difference in air pressure in the tire caused by driving speeds affected the changes of tire diameter. Moreover, most previous studies did not consider the fluctuation of tire inflation while estimating the speed profile since the tire diameter of a running vehicle could not directly be measured in real world conditions.

To examine the fluctuation of the tire diameter during a trip and to investigate this impact on the speed estimate, this study used an indirect approach with the following procedure:

- Estimate the travel distance based on the Kalman-filtered GPS coordinate data.
- Estimate the specific calibration number for the VSS speed estimates within each travel interval (a two-mile distance) per trip.
- Calculate the tire diameter using the specific calibration number per travel distance interval per trip.

<u>Investigation of Tire Inflation Trend</u>

This study investigated the fluctuation of specific calibration numbers, indicating tire inflation trends, based on sequential trips, and Figure 28 is the example showing that initial calibration numbers estimated during the first two-minute travel of every trip from one selected vehicle and indicates that these calibration numbers differ for every sequential trips. Although these calibration numbers had a little variation, this study observed that initial calibration numbers gradually increased with time, indicating the tire size progressively decreased.

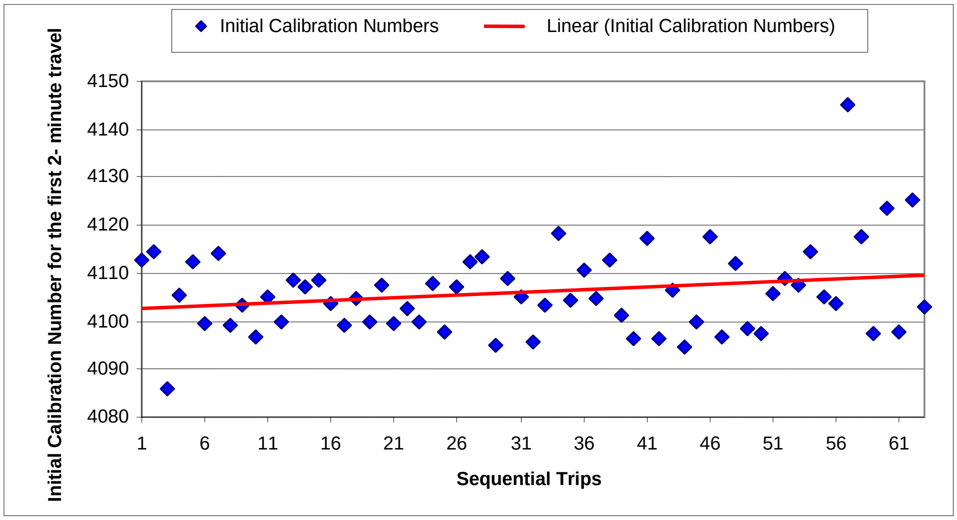


Figure 28: Example of Initial Calibration Numbers per Trip

In addition to the fluctuations of trip-based calibration numbers, this study examined the characteristics of tire inflation during a trip. The specific calibration numbers per the two-mile distance interval imply how tire diameter fluctuates during a trip. Figure 29 shows the scatter plot and the histogram of tire diameters of a vehicle estimated by specific calibration numbers.

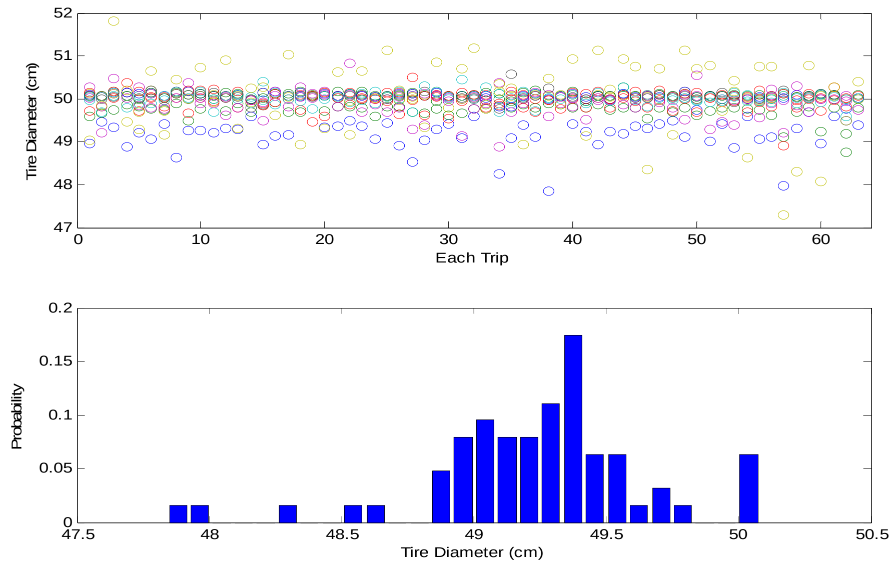


Figure 29: Scatter Plot and Histogram of Tire Sizes

Figure 30 depicts the average calibration numbers and the average driving speeds based on the two-mile travel interval. The result shows that average calibration numbers decreases as average speeds increase.

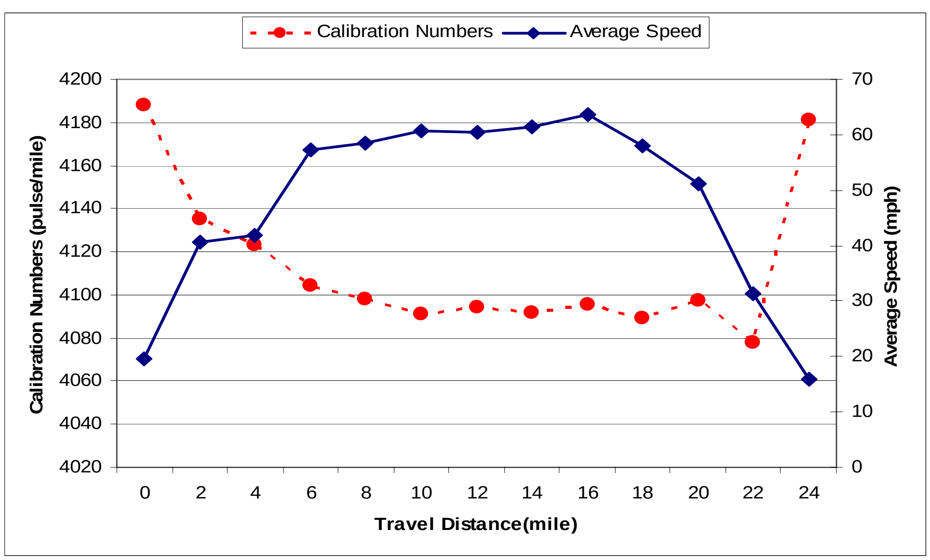


Figure 30: Relationships between Tire Size and Driving Speeds

The correction factors for the OBD speed data could be obtained by the same procedure for VSS speed estimates. After completing the calibration process, better VSS and OBD speeds were estimated and the example of results was shown in Figure 31. However, it should be known that regression analysis indicated how speed data measured by two different measurements were related; it did not explain in detail on the accuracy of speed data measured by GPS since there were some outliers having much difference of speed values between GPS and VSS.

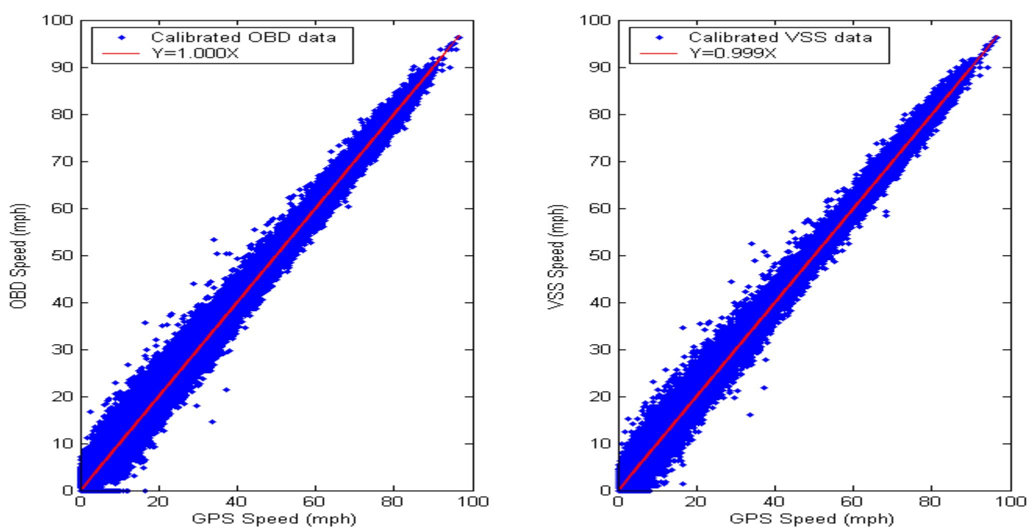


Figure 31: Results of Regression Analysis with All Calibrated Speed Data

<u>Accuracy of Filtered GPS Data</u>

With the main objective of eliminating or reducing unrealistic speed and acceleration data (or "outliers") from driving profiles, researchers carefully examined the results of the smoothing process to determine the effects of the smoothing. This study visually inspected the characteristics of speed and acceleration results from each smoothing technique with the original GPS-recorded speeds and accelerations, and statistically compared the speed and acceleration estimates with the VSS-derived speeds and accelerations. Further, this study also investigated how the smoothing algorithms actually dealt with these outliers. It is important to examine this effect because:

- Given that acceleration profiles are derived from sequential GPS speed data points, the impact of each smoothing technique on the original speed profile results in different acceleration profiles
- Given that random GPS errors in the speed profile provide unrealistic accelerations, extremely high acceleration or deceleration values must be eliminated by the smoothing technique
- The smoothing technique do not generally estimate much higher accelerations (or decelerations) than the original accelerations (or decelerations)

After running each smoothing technique with the original GPS-measured speed profile, this study estimated three statistics: the mean of the errors (ME), the variance of the errors (VE), and the mean of the absolute errors (MAE), using the following equations:

$$ME = Mean(Y_i - \hat{Y}_i)\,, \tag{17}$$

$$VE = Var(Y_i - \hat{Y}_i)\,, \tag{18}$$

$$MAE = \frac{\sum_i^n abs(Y_i - \hat{Y}_i)}{n}\,, \tag{19}$$

Speed and Acceleration Accuracy [17]

The results of the comparative analysis are presented in Table 5. For the impact of each smoothing technique on all GPS speed data, all techniques provided the similar mean of delta speeds, but the modified Kalman filter provided the smallest mean delta speed when the signal of GPS system indicated the poor quality such as less than four satellites. This result shows that the modified Kalman filtering method is superior to other smoothing techniques. In the case of accelerations, it also provided the smallest difference from the VSS-derived accelerations across all metrics.

Table 5: Speed and Acceleration Smoothing Results

Speed Comparison	Mean of Delta Speeds (mph)	
	From all GPS data	From GPS data with bad quality signal
The least squares spline approximation	-0.50	4.4
The kernel-based smoothing method	-0.49	4.4
The discrete Kalman filter	-0.49	4.4
The modified Kalman filter	-0.50	4.0

Acceleration Comparison	Mean (mph)	Variance (mph)	MAE (mph)
The least squares spline approximation	-0.00179	1.9669	0.77372
The kernel-based smoothing method	-0.00158	1.6287	0.69836
The discrete Kalman filter	-0.00133	1.4388	0.63735
The modified Kalman filter	-0.00047	1.4173	0.63222

To verify if the means of delta speed and delta acceleration between those estimates derived by each smoothing technique and the VSS-derived speed are significantly different, this study performed the *t*-test (α = 0.05). The hypothesis for testing the homogeneity is formulated as follows:

$$H_0: \ \mu(x) = \mu(y)$$

$$H_1: \ \mu(x) \neq \mu(y)$$

Table 6 shows that all delta speeds and delta accelerations did significantly differ, which indicated that each smoothing method except the conventional Kalman filter and the modified Kalman filter overall provided the different error distribution even though the means of delta speeds and accelerations were similar.

Table 6: Results of t-Test for the Means of Delta Speed

Delta Speed	(VSS – Spline)		(VSS – Kernel)		(VSS – Kalman)	
	Result	p value	Result	p value	Result	p value
(VSS – Spline)	-	-	-	-	-	-
(VSS – Kernel)	Reject	0	-	-	-	-
(VSS – Kalman)	Reject	1.68E-28	Reject	7.45E-156	-	-
(VSS – The modified Kalman)	Reject	2.08E-22	Reject	7.42E-172	Accept	0.22181

Delta Acceleration	(VSS – Spline)		(VSS – Kernel)		(VSS – Kalman)	
	Result	p value	Result	p value	Result	p value
(VSS – Spline)	-	-	-	-	-	-
(VSS – Kernel)	Reject	7.49E-15	-	-	-	-
(VSS – Kalman)	Reject	1.40E-13	Reject	1.69E-50	-	-
(VSS – The modified Kalman)	Reject	3.19E-18	Reject	1.22E-59	Accept	0.22397

In addition, because the statistical background of each smoothing method was different, they provided a unique output. For example, the kernel-based smoothing method often negatively impacted speed accuracy estimates while it did decrease outliers (large error-containing speeds). On the other hand, the least squares spline approximation, which minimized the residual sum of the squared errors (RSS) between the original data profile and the estimated output profile, also affected reliable speed points near suspected outliers. In contrast to these two methods, the Kalman filter did have as significant an impact on those GPS speed points with low fluctuations between the sequential points but instead affected those sequential speed points with the largest speed difference (Figure 32).

The least squares spline approximation provided higher and lower speed estimates than original speeds (sometimes, the least squares spline approximation provided negative speed estimates). The kernel-based smoothing method simultaneously smoothed the large range of speed data points around the outliers, which resulted in larger speed errors between the original and smoothed speed profiles.

Figure 32 also illustrates how each smoothing method produced different acceleration profiles. As expected, the least squares spline approximation frequently provided higher accelerations (or decelerations) than the original accelerations (or decelerations), which was not a desirable result in the smoothing process. Based on these results, the Kalman filter was the preferred smoothing method.

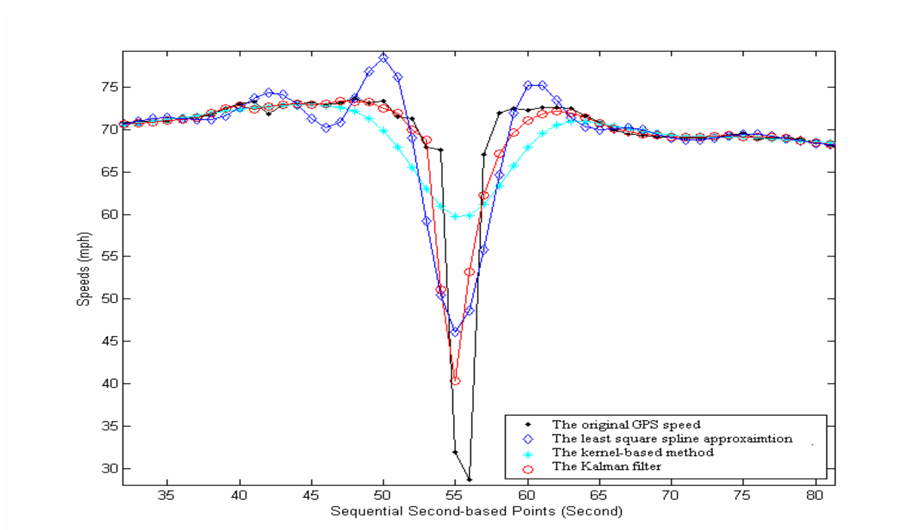


Figure 32: Smoothing Impacts of Outliers

Accuracy of Distance (or X-Y Coordinates) [17]

In addition to the speed and acceleration profiles, travel distance profiles were also compared. Travel distances could be estimated from either the GPS speed data or the GPS X-Y coordinates. This study used X-Y coordinates instead of GPS speed data, as the latter were already investigated in the previous section and because distance errors were expected to be larger when calculated using sequential position data. This study examined each smoothing technique for its ability for minimizing the impact of erroneous GPS data points on the estimates of travel distance.

Table 7 presents the results of the distance smoothing process. Similar to the speed and the acceleration, the groups of the Kalman filter provided the lowest delta distance. The modified Kalman filter provided almost same travel distances as the VSS-derived travel distances (Table 7 and Figure 33).

Table 7: Distance Smoothing Results Using the X-Y Coordinates

Distance Comparison	Mean of Errors in Travel Distance (mile)	MAE of Travel Distance (mile)
The least squares spline approximation	-97.414	97.904
The Kernel-based smoothing method	-56.604	57.127
The discrete Kalman filter	-52.919	53.537
The modified Kalman filter	0.179	0.192

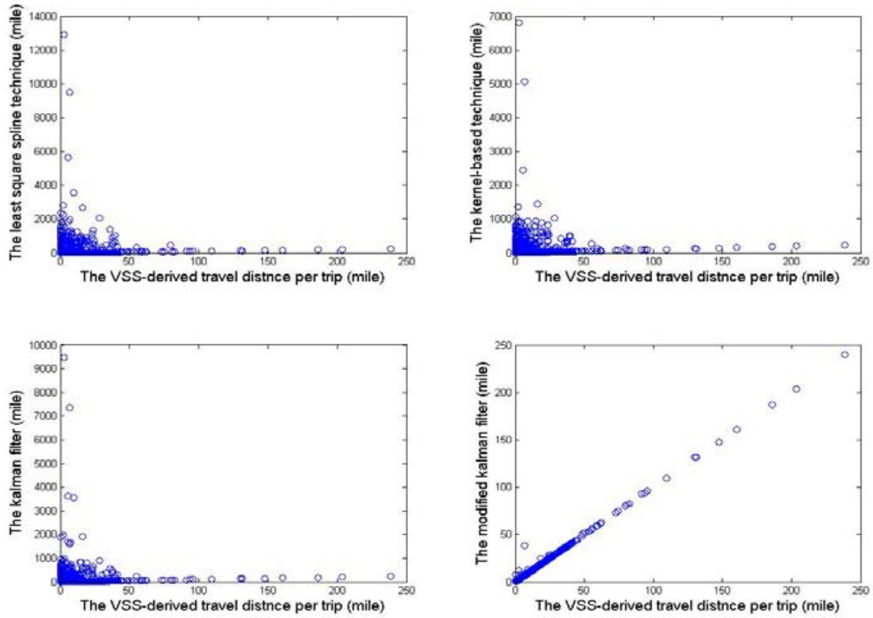


Figure 33: Travel Distance Comparisons

This study also performed the chi-square test to verify whether travel distance estimates were homogeneous with the VSS-derived distance. A contingency table (1 mile interval) with estimated chi-square statistics was created in Table 8, and the hypothesis for testing the homogeneity was formulated as follows:

$$H_0:\ F(x) = F(y)$$

$$H_1:\ F(x) \neq F(y)$$

For 40 degrees of freedom, the critical value is $\chi^2_{40,0.05} = 55.76$. Table 8 shows that all chi-square statistics were significantly greater than the critical value except that of the modified Kalman filter, which indicated that only travel distance estimate from the modified Kalman filter did not differ from the VSS-derived distance. The result of t-test also showed that travel distances filtered by the modified Kalman filter were not significantly different from those derived from the VSS data and that travel distances filtered by other techniques were significantly different.

Table 8: Contingency Table for Travel Distance

Distance Interval (mile)	VSS	Spline		Kernel		Kalman		The Modified Kalman	
	Freq.	Freq.	χ^2	Freq.	χ^2	Freq.	χ^2	Freq.	χ^2
0 ~ 1	302	270	1.79	286	0.44	279	0.91	284	0.55
1 ~ 2	162	162	0.00	159	0.03	163	0.00	169	0.15
2 ~ 3	126	94	4.65	93	4.97	93	4.97	124	0.02
3 ~ 4	80	59	3.17	60	2.86	59	3.17	84	0.10
4 ~ 5	62	67	0.19	60	0.03	64	0.03	55	0.42
5 ~ 6	75	58	2.17	64	0.87	65	0.71	76	0.01
6 ~ 7	53	59	0.32	59	0.32	56	0.08	51	0.04
7 ~ 8	104	110	0.17	110	0.17	112	0.30	102	0.02
8 ~ 9	131	67	20.69	78	13.44	67	20.69	135	0.06
9 ~ 10	59	53	0.32	42	2.86	53	0.32	60	0.01
10 ~ 11	33	44	1.57	42	1.08	42	1.08	41	0.86
11 ~ 12	33	34	0.01	33	0.00	34	0.01	28	0.41
12 ~ 13	21	22	0.02	23	0.09	23	0.09	23	0.09
13 ~ 14	63	33	9.38	31	10.89	34	8.67	66	0.07
14 ~ 15	37	27	1.56	28	1.25	29	0.97	41	0.21
15 ~ 16	20	13	1.48	12	2.00	13	1.48	15	0.71
16 ~ 17	10	8	0.22	10	0.00	6	1.00	14	0.67
17 ~ 18	5	11	2.25	9	1.14	11	2.25	5	0.00
18 ~ 19	14	3	7.12	2	9.00	2	9.00	15	0.03
19 ~ 20	8	9	0.06	8	0.00	8	0.00	6	0.29
20 ~ 21	10	9	0.05	10	0.00	10	0.00	10	0.00
21 ~ 22	7	8	0.07	7	0.00	8	0.07	9	0.25
22 ~ 23	9	10	0.05	16	1.96	10	0.05	9	0.00
23 ~ 24	34	37	0.13	44	1.28	39	0.34	34	0.00
24 ~ 25	65	38	7.08	35	9.00	39	6.50	61	0.13
25 ~ 26	6	19	6.76	15	3.86	21	8.33	9	0.60
26 ~ 27	8	9	0.06	9	0.06	7	0.07	7	0.07
27 ~ 28	4	10	2.57	11	3.27	10	2.57	5	0.11
28 ~ 29	31	19	2.88	22	1.53	20	2.37	25	0.64
29 ~ 30	14	15	0.03	9	1.09	13	0.04	20	1.06
30 ~ 31	6	5	0.09	7	0.08	5	0.09	5	0.09
31 ~ 32	1	4	1.80	1	0.00	4	1.80	4	1.80
32 ~ 33	9	8	0.06	8	0.06	9	0.00	7	0.25
33 ~ 34	7	8	0.07	7	0.00	7	0.00	7	0.00
34 ~ 35	6	4	0.40	3	1.00	3	1.00	6	0.00
35 ~ 36	12	4	4.00	6	2.00	6	2.00	12	0.00
36 ~ 37	3	10	3.77	11	4.57	8	2.27	4	0.14
37 ~ 38	13	10	0.39	11	0.17	11	0.17	13	0.00
38 ~ 39	8	9	0.06	5	0.69	9	0.06	8	0.00
39 ~ 40	13	8	1.19	9	0.73	9	0.73	14	0.04
40 ~	38	255	160.71	247	153.27	241	147.70	39	0.01
Total	1702	1702	249.38	1702	236.04	1702	231.91	1702	9.90

Summary of the Automatic Filtering Process

GPS data contain random errors that have the potential to affect speed, acceleration, and travel distance estimates based upon instrumented vehicle data. To use vehicle-based GPS data for insurance pricing, emissions analyses, and other modeling, GPS data smoothing may be required. This study evaluated three smoothing techniques that were popularly used in various traffic-related research and that were also characterized as different statistical background groups and evaluated their capabilities to minimize the impact of error-containing GPS data while estimating driving speeds, accelerations, and travel distances. In addition, this study modified the conventional discrete Kalman filter algorithm to apply better to GPS data smoothing process.

The study found that the modified Kalman filter provided the smallest differences from the VSS-derived speed, acceleration, and travel distance estimates across all statistical metrics [17]. In addition, through the visual inspection of impacts of each smoothing technique on the second-by-second data streams, the modified Kalman filter was superior to other smoothing techniques since this technique controlled for outliers in a more effective way. Furthermore, the Kalman filter required less computational time than the other method, which indicates that this technique can be applied for the real time smoothing algorithm.

Although only three smoothing methods were evaluated in this study, the study is currently recommending the use of the modified discrete Kalman filter for smoothing GPS speed and position data.

Analysis on Delta Speeds based on Speeds

After eliminating systematic and random errors in both GPS and VSS-derived speeds, the relationships between the delta speeds and the driving speeds were further investigated. Table 9 summarizes the statistics of the delta speeds based on speed intervals with 2.5 mph increments and shows that the largest delta speeds based on the mean of the errors (MEs) and the mean of the absolute errors (MAEs) occurred between speed intervals of 2.5 mph and 7.5 mph and that the standard deviations, indicating the dispersion of delta speeds, were also relatively high (Figure 34).

Table 9: Summary of Delta Speeds based on Driving Speeds

Speed Interval (mph)	ME (mph)	MAE (mph)	SE (mph)	Speed Interval (mph)	ME (mph)	MAE (mph)	SE (mph)
0 ~ 2.5	0.03	0.17	0.47	37.5 ~ 40	-0.04	0.34	0.81
2.5 ~ 5	-0.11	0.44	0.76	40 ~ 42.5	-0.03	0.28	0.60
5 ~ 7.5	-0.07	0.46	0.82	42.5 ~ 45	0.00	0.27	0.64
7.5 ~ 10	-0.09	0.44	0.76	45 ~ 47.5	0.01	0.25	0.57
10 ~ 12.5	-0.08	0.43	0.67	47.5 ~ 50	0.01	0.26	0.67
12.5 ~ 15	-0.04	0.41	0.74	50 ~ 52.5	0.01	0.26	0.69
15 ~ 17.5	-0.04	0.43	0.72	52.5 ~ 55	0.00	0.24	0.60
17.5 ~ 20	-0.02	0.40	0.67	55 ~ 57.5	0.01	0.25	0.62
20 ~ 22.5	0.01	0.37	0.60	57.5 ~ 60	-0.01	0.23	0.53
22.5 ~ 25	0.02	0.35	0.56	60 ~ 62.5	0.00	0.24	0.58
25 ~ 27.5	0.05	0.38	0.65	62.5 ~ 65	0.02	0.26	0.66
27.5 ~ 30	0.05	0.38	0.68	65 ~ 67.5	0.03	0.24	0.62
30 ~ 32.5	0.04	0.36	0.62	67.5 ~ 70	0.01	0.21	0.49
32.5 ~ 35	0.04	0.35	0.71	70 ~ 72.5	0.04	0.21	0.52
35 ~ 37.5	0.03	0.37	0.83	72.5 ~ 75	0.07	0.22	0.54
37.5 ~ 40	-0.04	0.34	0.81	-	-	-	-

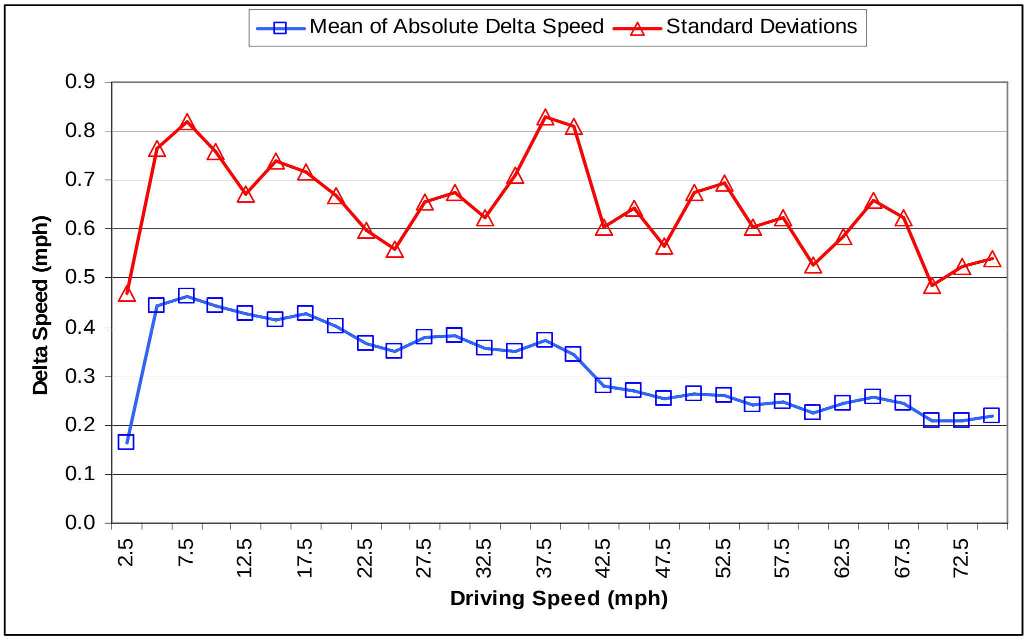


Figure 34: Mean of Absolute Delta Speeds and Standard Deviations

For the delta accelerations, Table 10 shows that speed range less than 10 mph also provides higher mean of the errors (ME) than other ranges. Based on the mean of absolute errors (MAE), speed range between 15 and 20 mph also provided higher delta accelerations.

Table 10: Summary of Delta Accelerations based on Driving Speeds

Speed Interval (mph)	0 ~ 5	5 ~ 10	10 ~ 15	15 ~ 20	20 ~ 25	25 ~ 30	30 ~ 35	35 ~ 40
ME (mph/s)	-0.1	-0.341	-0.178	0.176	0.151	0.158	0.138	0.098
MAE (mph/s)	0.193	0.866	0.982	0.85	0.609	0.529	0.469	0.404
SE (mph/s)	0.485	1.226	1.565	1.408	0.886	0.724	0.671	0.569
Speed Interval (mph)	40 ~ 45	45 ~ 50	50 ~ 55	55 ~ 60	60 ~ 65	65 ~ 70	70 ~ 75	-
ME (mph/s)	0.083	0.071	0.044	0.029	0.003	-0.002	0.018	-
MAE (mph/s)	0.358	0.332	0.309	0.319	0.333	0.293	0.308	-
SE (mph/s)	0.508	0.494	0.449	0.455	0.467	0.414	0.428	-

This study investigated the relationships between delta speeds and driving speeds and found that the speed range less than 7.5 mph provided a higher average and standard deviation of delta speeds than other speeds, indicating that the GPS speed profile in this speed range might be less reliable than in other speed ranges. This speed range also produced higher delta accelerations. Although the both GPS and VSS generally provide accurate speed estimates, particularly, when their systematic and random errors are corrected, this study concludes that they potentially provide higher discrepancy in their speed estimates in the low speed range, which also results in larger difference of acceleration rates. This study suggests that when researchers use either GPS-measured or VSS-derived speeds, systematic and random errors must be eliminated and need to exercise caution in using low speed data.

Chapter Five

POTENTIAL BEHAVIORAL EXPOSURE I: Travel Mileage

The goal of this study is to evaluate potential driving behavior activity exposure measures using GPS-observed metrics to identify drivers who have high possibility of potential crash involvements from the driver population. Ultimately, this knowledge will help drivers modify their risky driving or activity patterns. Thus, those driving behavior activity exposure measures should explain the differences in driving or activity patterns of drivers between who were involved and were not involved in crashes.

Relationships between Crash Involvements and Age/Gender

Before assessing the relationships between driving behavior activity exposures and crash risk, this study evaluated if any differences in crash involvement rates based on age and gender exist in the sample data using the chi-square test. Table 11 and Figure 35 show the relationships between crash involvements and gender. The result of the test of the homogeneity of crash involvement rates between male and female drivers indicated that all chi-square statistics (3.84 and 6.63 at the 0.05 and 0.1 significance levels, respectively) were significantly greater than the critical value (0.93). Thus, this study suggests that crash involvement rates are not significantly different between male and female drivers (at least within the study sample data).

Table 11: Gender vs. Crash Involvements

Gender	Drivers who *were not involved* in crashes	Drivers who *were involved* in crashes	Total
Male	125	20	145
Female	148	23	171
Total	273	43	316[9]

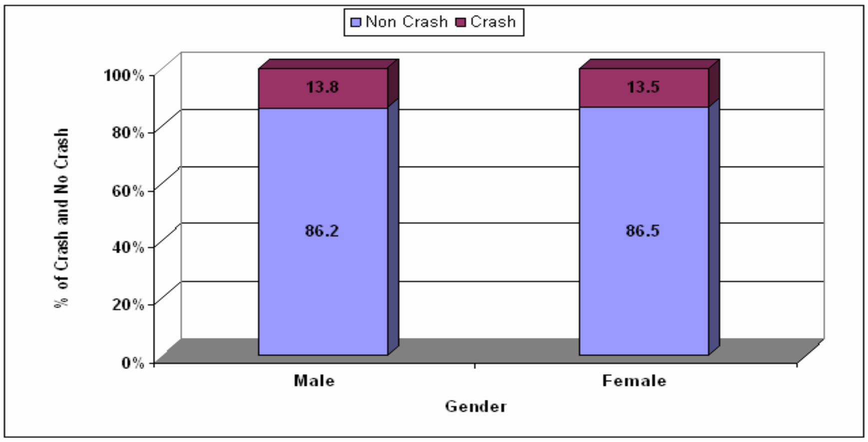


Figure 35: Crash Involvement Rates between Male and Female

Table 12 and Figure 36 show the relationships between crash involvements and driver age. The result of the test of the homogeneity of crash involvement rates between different age drivers indicated that all chi-square statistics (11.07 and 15.09 at the 0.05 and 0.1 significance levels, respectively) were significantly greater than the critical value (0.84). Thus, this study may also suggested that crash involvement rates were not different between different age drivers (again, at least within the study sample data). The relationships between gender/age and crash involvement rates will be analyzed again in Chapter 11 with the selected exposure variables.

[9] The total sample number (316 drivers) indicates number of all responders including all household members who do not have installed vehicles.

Table 12: Age vs. Crash Involvements

Age	Drivers who were *not* involved in crashes	Drivers who were involved in crashes	Total
(- , 20)	23	5	28
(21, 35)	30	6	36
(36, 45)	56	10	66
(46, 50)	34	5	39
(51, 60)	76	8	84
(61, -)	54	9	63
Total	273	43	316

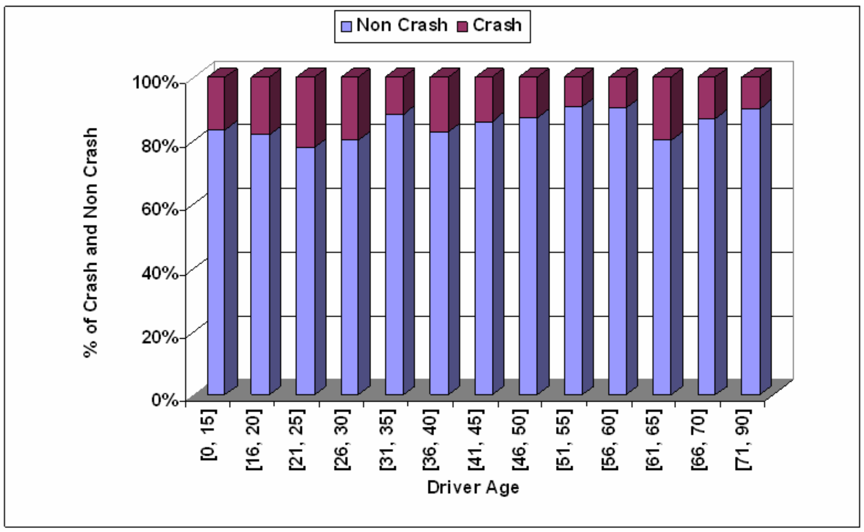


Figure 36: Crash Involvement Rates based on Driver Age

Test of Difference in Means Using the Nonparametric Bootstrap Technique

To test if drivers who were involved in crashes have higher travel mileage, engage in frequent and extreme speeding behavior, or produce other different behavior patterns such as hard accelerations compared to drivers who were not involved in crashes, this study examined differences in average values of behavioral exposure metrics derived from each group.

Based on the normality assumption using the central limit theorem, a parametric test such as t-test is popularly used in most research to evaluate the equality of the means, but the sample size of each variable need to be greater than 25 [52, 53]. However, other statistical references also show that the number of samples needs to be greater than 100 to obtain a satisfactory result if the sample is not normally distributed [19, 46, 53].

The sample size of drivers who were not involved in crashes is larger than 100 (141 drivers), but that of drivers who were involved in crashes is small (26 drivers). After performing a preliminary test about the distribution of each driving behavior activity exposure metric, this study found that most disaggregated metrics derived from the two driver-groups were not normally distributed (even after log transformation) utilizing three nonparametric methods; Jarque-Bera test, Lilliefors test, and Kolmogorov-Smirnov test (KS-test).

Due to the small sample size and non-normal distribution, this study used an alternative method for the confidence interval estimation of means as well as the parametric test method. A nonparametric bootstrap resampling method was utilized to estimate the confidence intervals of sample means. The bootstrap method treats the original sample set as a pseudo population in order to estimate the true population [46].

This bootstrap method employs uniform random sampling with replacement method to create new data sets from the original sample data $x = (x_1,...x_n)$. Uniform resampling means that each data point x_i has the same probability (i.e., 1/n), of being randomly selected. The bootstrap creates new sample data which is $x^{*b} = (x_1^{*b},...,x_n^{*b})$ where b represents the number of sampling processes [46]. From the each new bootstrap sample, the empirical distribution $\hat{F}$ can be estimated (Equation 20), and then $\hat{\theta}^{*b}$ is calculated from the empirical distribution $\hat{F}$ (Equation 21).

$$F(\hat{P}_n) = F(x_1^{*b}, ..., x_n^{*b}), \tag{20}$$

$$\hat{\theta}^{*b} = f(x^{*b}), \tag{21}$$

where, b represents the number of sampling process, $b = 1, 2, \ldots, B$ for the b^{th} bootstrap sample [46].

Because the bootstrap method samples randomly with replacement, any value x_i can appear more than once or not at all in a new bootstrap sample data set. For the number of bootstrap resampling B, Martinez et al. [46] recommended that B should be more than 1000 for the confidence interval of each statistic such as mean. Thus, this study conducted 1000 times for the bootstrap resampling process B and used nonparametric based model, so called the bootstrap percentile interval, to estimate confidence intervals since the nonparametric based method is more stable than other methods, especially when B is more than 1000 [46]. Equation 22 shows how the bootstrap method calculates the confidence interval.

$$(\hat{\theta}_B^{*(\alpha/2)}, \hat{\theta}_B^{*(1-\alpha/2)}), \tag{22}$$

When α is 0.05 and B is 1000, the $\hat{\theta}_B^{*(\alpha/2)}$ indicates that the 25th data point (0.025 %) of the ordered bootstrap samples and the $\theta_B^{*(1-\alpha/2)}$ represents the 975th data point (0.975 %).

Finally, this bootstrap technique creates a pseudo sample data set by resampling exposure metrics based on each driver to estimate the sample mean. Due to the replacement process, one driver's metrics may be shown several times in the pseudo sample data set, and some of drivers may not be selected in each bootstrap sample process. Thus, the bootstrap technique can reduce the impact of biased samples on the result of analysis (confidence interval in this study) by the repeated sampling and un-sampling process.

Test of Difference in Means Using the Wilks' Lambda Test

In addition to the nonparametric bootstrap technique, this study used the Wilks' lambda test to verify differences in means of driving behavior activity metrics between the two driver-groups since the sample size of drivers who were involved in crashes is greater than 25, but this sample size may not be large enough. (As mentioned earlier, the central limit theorem can be applied to not-normally distributed samples if the sample size is greater than 25 [52, 53]). The Wilks' lambda test is the popular asymptotic method in discriminant analysis to check the equality of means of groups and is the analog of the F-test for multivariate analysis of variance (ANOVA) [54]. Thus, the Wilks' lambda test examines if the means of behavioral metrics are equal across the two driver-groups (the null hypothesis). Unlike to the bootstrap technique, potential driver bias may be more pronounced in this method because the Wilks' lambda test rely on the original sample data collected without examining individual driver effects by pulling and entering their driving behavior activity metrics.

Due to the small sample size of drivers who were involved in crashes as well as the characteristics of parametric and non-parametric methods (the Wilks' lambda test and the bootstrap technique, respectively), this study selects driving behavior activity exposure metrics showing significant differences between the two driver-groups based on results from either the bootstrap technique or the Wilks' Lambda test. This approach reduces the possibility of losing potential measures that can happen when relying on only one of methods. From the test result, any significant values less than the certain level such as 0.05 indicate that the means of driving behavior activity exposure measures estimated from the two driver-groups are statistically different.

Analysis on Mileage Differences

Based on literature reviews [7, 22], this study first investigated travel mileage between the two driver-groups. The difference in total travel mileage between two driver-groups (with and without crash involvements during the 14-months study period) using the bootstrap technique were statistically significant (α = 0.05) (Figure 37), indicating that the total travel mileage measure can be used for identifying drivers who potentially have higher crash involvement rates from the general driver population.

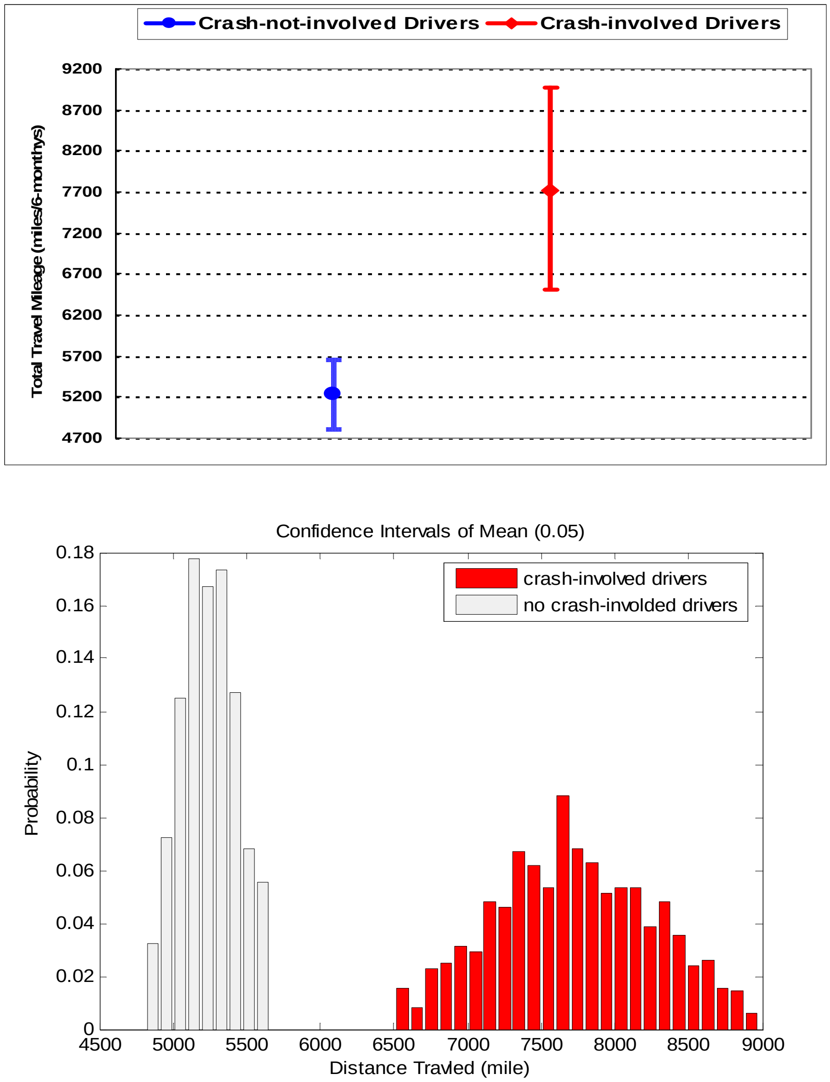


Figure 37: Difference in Total Travel Mileage between the Two Driver-Groups

The average travel mileage of drivers who were involved in crashes was 7,718 miles/six-months and that of drivers who were not involved in crashes was 5,244 miles/six-months, showing a difference of 32 %. This result supports the conventional definition regarding the exposure, where higher exposure on roadways is linked to higher possibility of crash involvements. The Wilks' lambda test also showed that drivers who were involved in crashes had significantly higher travel mileage than drivers who were not involved in crashes (p-value: 0.009).

After assessing a positive relationship between total travel mileage and crash involvement rates, this study examined if the distributions of travel mileage by time-period (time of day) and facility types between the two driver-groups were different to verify any preferred time-periods and facility types of trip-making existed between them who had different crash involvements. For this disaggregated exposure analysis, this study used three roadway types, freeways, arterials, and local roadways, and six different time frames, am peak (6 am ~ 9 am), morning (9 am ~ 12 am), afternoon (12 pm ~ 5 pm), pm peak (5 pm ~ 8 pm), nighttime (8 pm ~ 12 pm)[10], and early morning (12 am ~ 6 am).

After utilizing the Kolmogorov-Smirnov (KS) test, this study found that the distributions of time-specific activities between two driver-groups were not significantly different (p-value: 0.32). Figure 38 shows the distributions of time-specific travel mileages between the two driver-groups who were involved and not involved in crashes. This result indicated that crash involvement may differ by time of day, but crash-involved and non-crash-involved drivers did not travel differently by time of day. Thus, this study suggested that the choices on trip time by drivers between who were involved and not involved in crashes were not significantly different.

[10] Readers are cautioned to keep these time-periods in mind when interpreting the results of the study (i.e., results related to nighttime should not be misconstrued as the total dark/night period).

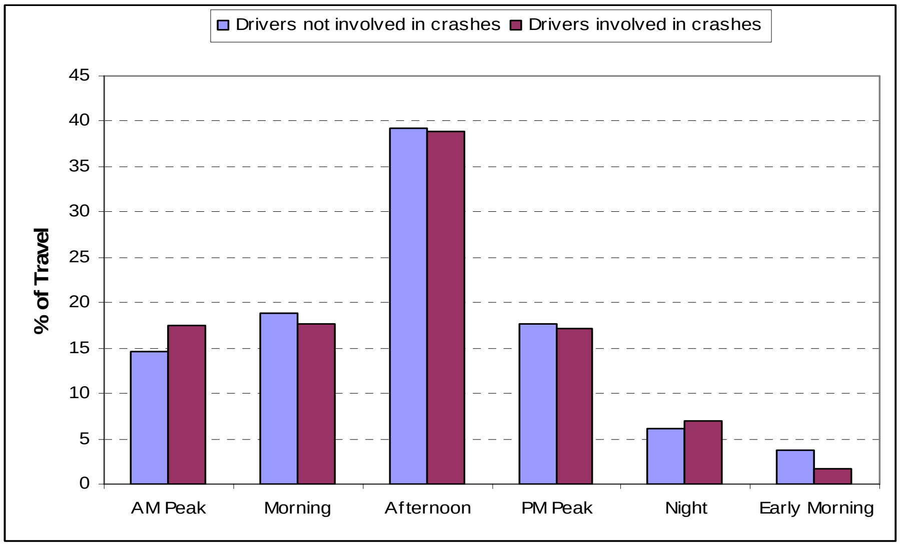


Figure 38: Travel Percentage Regarding Time of Day between Two Driver Groups

Furthermore, this study examined the differences in exposures to facility types to identify if any preferences on roadway types existed between two driver-groups. With the same test method (KS test), this study analyzed the distributions of facility-specific exposures and found that they were significantly different (p-value: 0.03) (Figure 39).

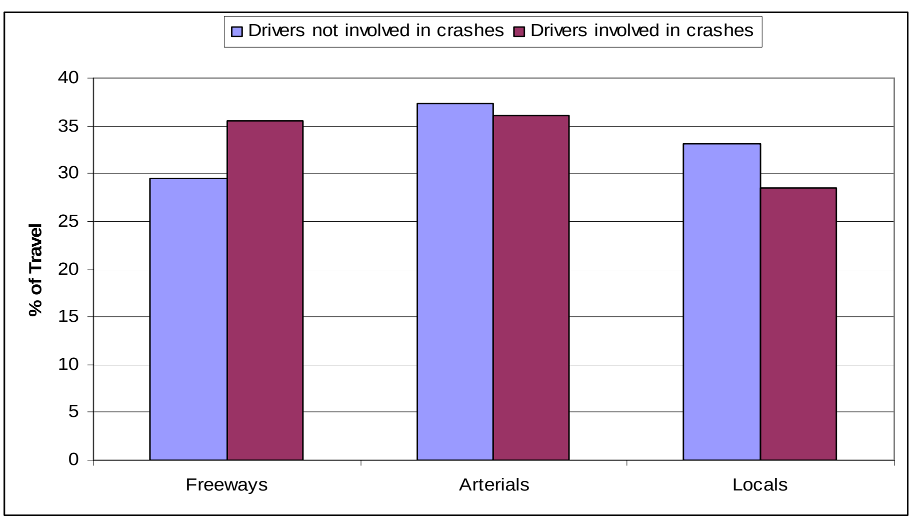


Figure 39: Travel Percentage Regarding Facility Types between Two Driver Groups

Thus, indicating that crash-involved drivers were more likely to undertake freeway mileage than drivers who were not involved in crashes. Further analyses will need to have more information associated with each individual crash event so that crash location (facility type) and time of day can be brought into these types of analyses.

The confidence interval estimates of means using the bootstrap technique showed that the average travel mileage on freeways of drivers who were involved in crashes was significantly different from that of drivers without crash involvements, resulting a difference of 36% (Table 13). The mean of total travel mileage in 13-counties area of crash-involved drivers (5,302 mile per 6 months) was higher (130 %) than that of no-crash-involved drivers (4,085 mile per 6 months), but the difference between the two driver-groups was not significant.

Table 13: Facility and Trip Time Mileage Differences in 13 Counties[11]

Facility Type	Mean of Travel Mileage (miles/6-months)		Mileage Difference (mile)	Difference (%)
	Drivers who *were not involved* in crashes	Drivers who *were involved* in crashes		
Freeways *	**1,204.53**	**1,880.96**	**677.42**	**36**
Arterials	1,525.29	1,914.94	389.65	20
Local Roads	1,355.21	1,509.21	154.01	10

* indicates a significant mean difference ($\alpha = 0.05$).

This study also investigated facility-time-specific average travel mileages. While the distributions of mileages by time of day were not significant by itself, the interaction between facility and time of day did provide significant differences in distributions between the two groups of drivers (p-value: 0.00002, $\alpha = 0.05$) (Figure 40).

[11] Information of facility types is available only inside 13 counties in the GIS database.

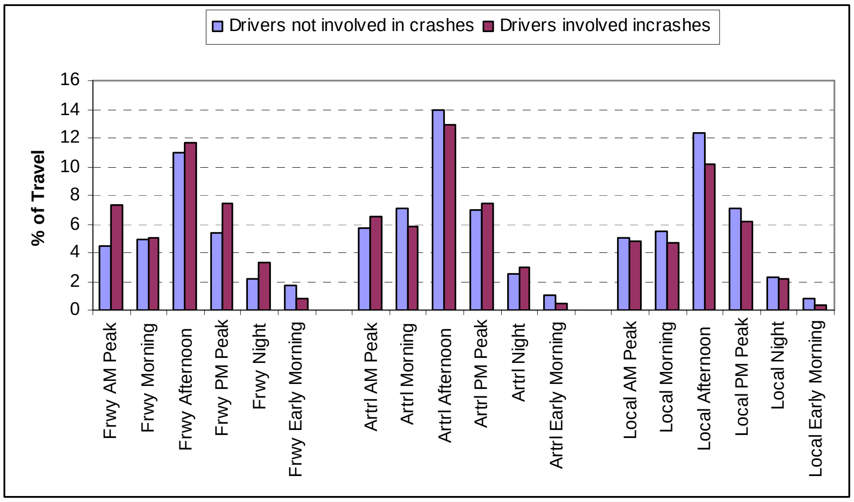


Figure 40: Travel Percentage Regarding Facility and Trip Time

Figure 41[12] shows confidence intervals of average facility-time-specific travel mileage between those who had crashes and did not have crashes during the 14-months period, and Figure 42 illustrates the distributions of the means of travel mileage within the confidence intervals between the two driver-groups using the bootstrap technique.

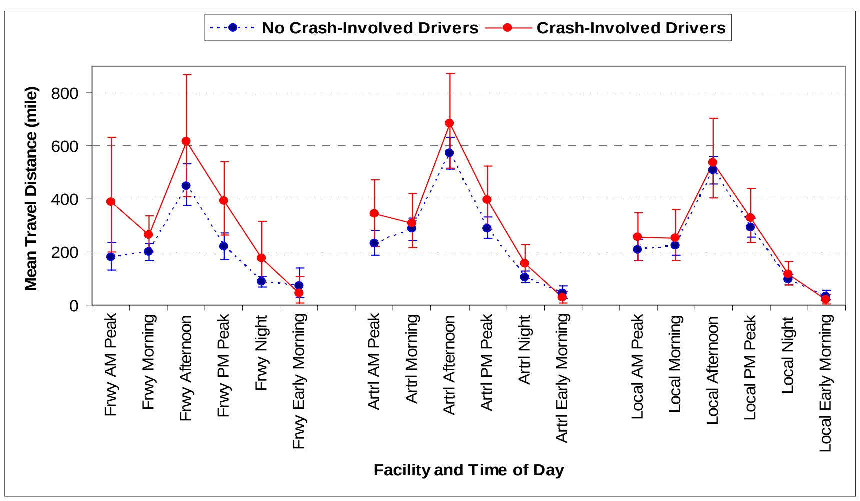
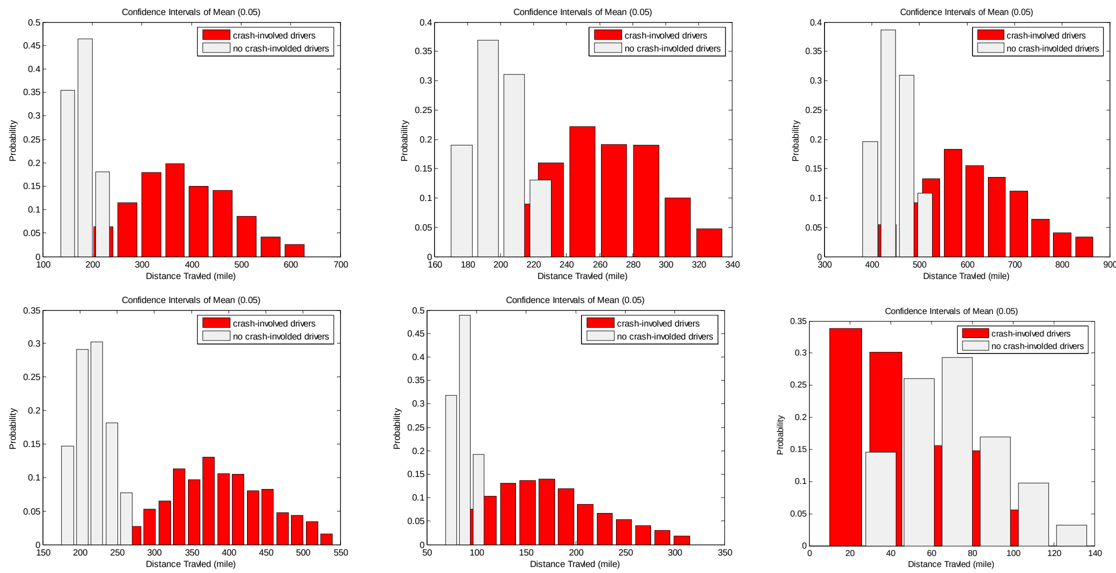
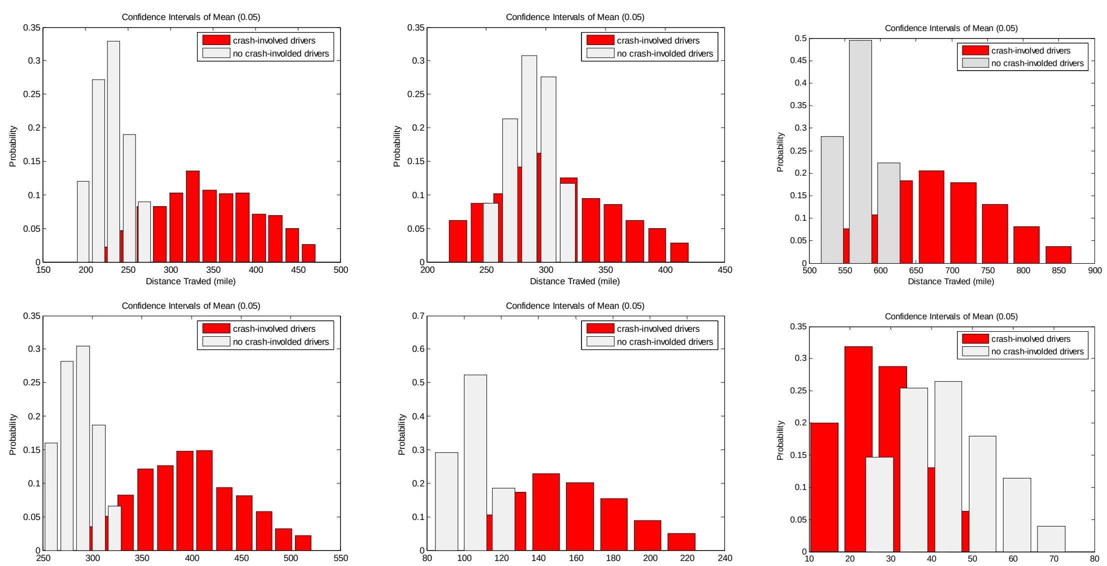


Figure 41: Confidence Intervals of Facility-Time-Specific Mileage Inside 13 Counties based on the Bootstrap Technique

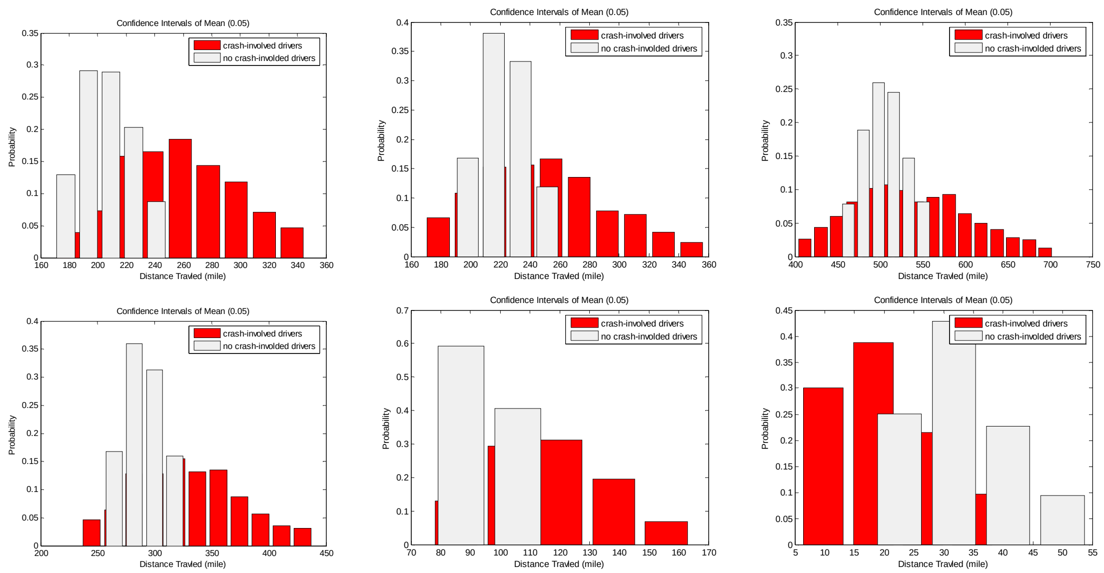
[12] The bootstrap technique did not provide any significant differences in facility-time-specific travel mileage between the two driver-groups.



(A) Mean Distributions within the Confidence Intervals of *Freeway* Mileages based on Time of Day (AM Peak, Morning, Afternoon, PM Peak, Nighttime, and Early Morning)



(B) Mean Distributions within the Confidence Intervals of *Arterial* Mileages based on Time of Day (AM Peak, Morning, Afternoon, PM Peak, Nighttime, and Early Morning)



(C) Mean Distributions within the Confidence Intervals of *Local* Mileages based on Time of Day (AM Peak, Morning, Afternoon, PM Peak, Nighttime, and Early Morning)

Figure 42: Distributions of Means of Average Travel Mileage in 13 Counties Using the Bootstrap Technique

While the bootstrap technique did not provide significant mileage-related exposure metrics with respect to the facility-time-specific travel mileage at the 0.05 significance level (only total travel mileage between the two driver-groups were significantly different), the Wilks' lambda test showed that four facility-time-specific mileage exposures were significantly different between groups (Table 14). This result indicated that drivers who were involved in crashes traveled much more on freeways, especially during peak time periods and nighttime.

Table 14: Tests of Equality of Means of Facility-Time-Specific Travel Mileage based on the Wilks' Lambda Test

Facility	Trip Time	Wilks' Lambda	F	Sig.
Freeway	**AM peak ***	**0.960**	**6.816**	**0.010**
	Morning	0.986	2.427	0.121
	Afternoon	0.984	2.645	0.106
	PM peak *	**0.959**	**6.966**	**0.009**
	Night *	**0.962**	**6.467**	**0.012**
	Early morning	0.999	0.176	0.676
Arterial	AM peak	0.980	3.382	0.068
	Morning	0.999	0.175	0.676
	Afternoon	0.988	2.007	0.158
	PM peak *	**0.974**	**4.361**	**0.038**
	Night	0.983	2.864	0.092
	Early morning	0.998	0.292	0.589
Local Roads	AM peak	0.994	0.945	0.332
	Morning	0.998	0.320	0.572
	Afternoon	0.999	0.137	0.712
	PM peak	0.996	0.672	0.413
	Night	0.995	0.748	0.389
	Early morning	0.997	0.487	0.486

* indicates a significant mean difference ($\alpha = 0.05$)

Table 15 shows the results of the differences in travel mileage based on the pairs of facility and trip time. Based on the Wilks' lambda test, specially, freeways during AM peak provided the largest mileage difference (54%) between crash-involved drivers (388 miles/6-months) and drivers not involved in crashes (180 miles/6-months) and showed the significant difference of the means between the two driver-groups. In addition, travel mileage on freeways during PM peak also showed the significant difference of 44% between the means of the two groups ($\alpha = 0.05$).

Table 15: Facility-Time-Specific Mileage and Differences in 13 Counties

Facility Type	Trip Time	Mean of Travel Mileage (miles/6-months): Drivers who *were not involved* in crashes – Mile	Drivers who *were not involved* in crashes – %	Drivers who *were involved* in crashes – Mile	Drivers who *were involved* in crashes – %	Mileage Difference	% Difference
Freeways	**AM Peak ***	**180.04**	**4.41**	**388.18**	**7.32**	**208.14**	**54**
	Morning	198.66	4.86	264.05	4.98	65.39	25
	Afternoon	448.06	10.97	615.71	11.61	167.65	27
	PM Peak *	**218.29**	**5.34**	**392.26**	**7.39**	**173.97**	**44**
	Night *	**86.51**	**2.12**	**176.72**	**3.33**	**90.21**	**51**
	Early Morning[13]	71.97	1.76	44.04	0.83	-27.94	-63
Arterials	AM Peak	231.46	5.67	345.26	6.51	113.8	33
	Morning	288.02	7.05	308.57	5.82	20.55	7
	Afternoon	570.39	13.97	682.67	12.87	112.28	16
	PM Peak *	**286.5**	**7.02**	**395.94**	**7.46**	**109.44**	**28**
	Night	104.58	2.56	155.49	2.93	50.91	33
	Early Morning	44.34	1.09	27.01	0.51	-17.33	-64
Local Roads	AM Peak	206.73	5.06	256.52	4.84	49.79	19
	Morning	223.21	5.47	250.51	4.72	27.3	11
	Afternoon	506.34	12.40	537.87	10.14	31.53	6
	PM Peak	290.56	7.11	328.55	6.19	37.99	12
	Night	95.04	2.33	116.52	2.20	21.48	18
	Early Morning	33.33	0.82	19.24	0.36	-14.09	-73

* indicates a significant mean difference ($\alpha = 0.05$).

Analysis on Travel Mileage Outside 13-county Area (Outside Regional Mileage)

This study also evaluated travel mileage outside 13 counties with the same statistical procedures, but facility-specific mileages could not be estimated due to the unavailability of roadway characteristics (RC) information in GIS database. Thus, aggregated total outside-regional travel mileage and disaggregated time-specific travel mileage outside 13 counties were estimated.

The total outside-regional travel mileage between crash-involved and no-crash-involved drivers were not significantly different ($\alpha = 0.05$) (Figure 43) while the total

[13] Due to the very low activity during early morning, this study did not examine the difference between the two driver-groups during this period.

travel mileage including inside and outside regions were significantly different. The mean of total travel mileage of drivers who experienced crashes was 2,024 miles/six-months, and that of drivers who were not involved in crashes was 1,047 miles/six-months, resulting a difference of 48 %.

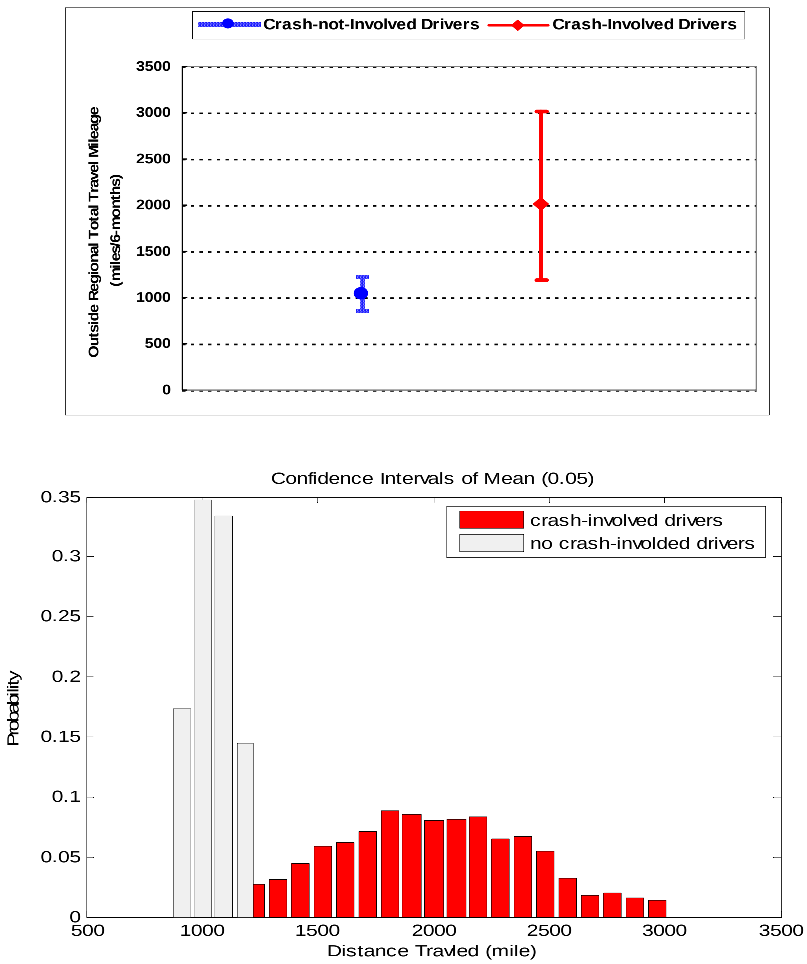


Figure 43: Confidence Intervals and Distributions of the Means of Outside-Regional Travel Mileage Using the Bootstrap Technique

Figure 44 shows the means of trip time-based travel mileage outside 13 counties between the two driver-groups with and without crash involvements. The test result using the bootstrap technique provided average outside-regional travel mileage by time of day between the two driver-groups were not significantly different at the 0.05 significance level.

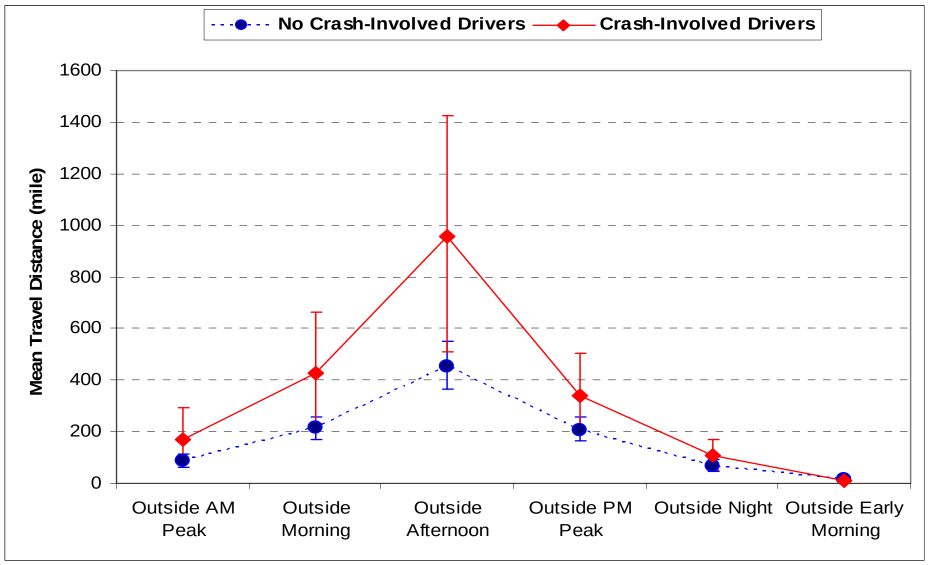


Figure 44: Outside-Regional Average Travel Mileage by Time of Day

However, Table 16 shows the result of the tests of mean equality using the Wilks' lambda test. The result showed that four outside-regional time-specific mileage measures with respect to AM peak, morning, afternoon, and PM peak were significantly different between the two driver-groups.

Table 16: Tests of Equality of Means of Outside-Regional Travel Mileage Based on the Wilks' Lambda Test

Area	Time Period	Wilks' Lambda	F	Sig.
Outside 13 Counties	**Am peak ***	**0.968**	**5.403**	**0.021**
	Morning *	**0.949**	**8.858**	**0.003**
	Afternoon *	**0.932**	**12.045**	**0.001**
	PM peak *	**0.975**	**4.281**	**0.040**
	Night	0.989	1.775	0.185
	Early morning	0.999	0.176	0.676

* indicates a significant mean difference ($\alpha = 0.05$).

Crash-involved drivers traveled 963 miles/six-months during afternoon (12 pm ~ 6 pm) while drivers not involved in crashes traveled only 450 miles/six-months, providing the largest difference of 53 % (Table 17). The AM peak mileages produced the difference of 50 %, and the mileages during morning provided the difference of 49 % between the two driver-groups with the significance (Table 17).

Table 17: Outside-Regional Time-Specific Mileage and Differences

Trip Time	Mean of Outside-Regional Travel Mileage (miles/6-months): Drivers who *were not involved* in crashes — Mile	Drivers who *were not involved* in crashes — %	Drivers who *were involved* in crashes — Mile	Drivers who *were involved* in crashes — %	Mileage Difference	% Difference
AM Peak *	**85**	**8.14**	**171**	**8.43**	**85.26**	**50**
Morning *	**217**	**20.72**	**427**	**21.06**	**209.32**	**49**
Afternoon *	**450**	**42.95**	**963**	**47.58**	**513.10**	**53**
PM Peak *	**210**	**20.02**	**344**	**17.01**	**134.64**	**39**
Night	68	6.45	108	5.32	40.06	37
Early Morning	18	1.73	12	0.60	-5.91	-49

* indicates a significant mean difference ($\alpha = 0.05$).

Linear Discriminant Analysis (LDA) using Travel Mileage Exposure Metrics

After selecting all significant mileage-related exposure metrics from the two driver-groups, this study utilized the technique of linear discriminant analysis (LDA) to classify and predict group memberships (potentially high crash-involvement group or low crash-involvement group) from a given set of independent variables such as GPS-observed driving behavior activity metrics. The general purposes of this technique are 1) to classify two or more groups using a discriminant equation or probability, 2) to examine and select independent variables showing differences among groups, and 3) to predict appropriate group memberships of future observations.

The LDA technique, essentially the inverse process of Multivariate Analysis of Variance (MANOVA), required that independent variables are normally distributed and have equal variance within each group [54]. However, Tabachnick et al. [55] showed that the linear discriminant analysis (LDA) is robust to violations of normality and homogeneity of variances within each group. Lachenbruch [56] also indicated that LDA is relatively robust even when there are modest violations of these assumptions, and Klecka [57] pointed out that dichotomous variables (two categorical dependent variables), which often violate multivariate assumptions, are not likely to affect conclusions based on LDA. However, they claimed that the violation of multi-collinearity between independent variables must be avoided or the analysis will yield unstable results [55].

Table 18 shows the results of correlations among the significant nine mileage-related exposure metrics obtained from either the bootstrap method or the Wilks' lambda test. There is no rule about the relationship threshold that determines what amount of correlations indicates strong correlation between random variables. Cohen et al. [58]

asserted that correlations (in absolute values) should not be determined too strictly due to the arbitrary concept and suggested that correlations between 0.5 and 1, between 0.3 and 0.49, and between 0.1 and 0.29 can be considered large, medium, and small correlations, respectively.

Due to the model assumption of the LDA technique, those correlated variables causing multi-collinearity should be controlled before performing the analysis. This study used 0.6 as a threshold that determined a strong correlation between independent variables.

Table 18: Correlation Analysis of Travel Mileage Exposure Metrics

		Freeways			Arterials	Outside Regional				All
		AM Peak	PM Peak	Night	PM Peak	AM Peak	Morning	Afternoon	PM Peak	Total
Freeways	AM Peak	1.00	-	-	-	-	-	-	-	-
	PM Peak	**0.69**	1.00	-	-	-	-	-	-	-
	Night	0.10	0.44	1.00	-	-	-	-	-	-
Arterials	PM Peak	0.21	0.40	0.12	1.00	-	-	-	-	-
Outside	AM Peak	0.12	0.04	-0.11	0.05	1.00	-	-	-	-
	Morning	-0.07	0.03	-0.07	-0.08	0.54	1.00	-	-	-
	Afternoon	-0.08	0.01	-0.02	-0.12	0.47	**0.90**	1.00	-	-
	PM Peak	-0.01	0.15	0.12	0.09	0.58	**0.70**	**0.80**	1.00	-
All	Total	0.46	0.58	0.34	0.57	0.00	-0.04	-0.06	0.05	1.00

To select better discriminating variables without correlations and also to remain the accuracy of the discrimination process between the two driver-groups, this study examined structure coefficients, also called discriminant loading power, indicating the relative importance of each independent variable in discriminating between the groups [54]. Thus, this study selected variables having higher structure coefficients (loading power) among the correlated variables. Table 19 provided the result of structure coefficients using the nine selected mileage-related exposure metrics[14].

Based on the structure coefficients, this study finally obtained six mileage-related exposure metrics that can be potentially used for classifying drivers into two different crash-groups. This study repeated the linear discriminant analysis (LDA) with six mileage exposure metrics to remove the impact of correlations on the classification result.

Table 19: Structure Coefficients from the Linear Discriminant Analysis Using Travel Mileage Exposures Metrics

Facility	Time of Day	Structure Coefficients	Rank	Variable Selection
Outside	**Afternoon**	**0.547**	**1**	√
Outside	Morning	0.469	2	-
All	**All**	**0.418**	**3**	√
Freeway	**PM Peak**	**0.416**	**4**	√
Freeway	AM Peak	0.411	5	-
Freeway	**Night**	**0.401**	**6**	√
Outside	**AM Peak**	**0.366**	**7**	√
Arterial	**PM Peak**	**0.329**	**8**	√
Outside	PM Peak	0.326	9	-

Finally, Table 20 shows the accuracy of the discriminant analysis using the only selected variables. As a result, 80.9 % of drivers who were not involved in crashes and 57.7 % of drivers who were involved in crashes were correctly classified based on mileage-exposure metrics. Overall performance of the model employing six mileage exposure metrics was 77.2 %. On the other hand, 19.1 % of drivers who were not involved in crashes were classified as a potential crash involvement group based on their

[14] This study used the SPSS (Version 12.0.1) statistics software for the linear discriminant analysis (LDA).

mileage exposures, indicating that their mileage exposures were close to those of drivers who were involved in crashes, yet they had not experienced a crash.

For the 42.3 % of drivers who were involved in crashes, they were theoretically classified into low crash-involvement group based on their observed mileage exposures. This result indicates that the model has a limitation for classifying drivers who have potentially high crash involvement rate based on mileage exposures only and more likely, other interactions between driving behavior activities and drivers (or vehicles) may need to be examined.

Table 20: Classification Results Using Travel Mileage Exposure Metrics

Crash Involvements	Predicted Group Membership	
	Drivers who *were not involved* in crashes	Drivers who *were involved* in crashes
Drivers who *were not involved* in crashes	80.9 %	19.1 %
Drivers who *were involved* in crashes	42.3 %	57.7 %

Summary of the Travel Mileage Exposure

This study evaluated the differences in travel mileage of drivers with and without crash involvements during the 14-months study period based on the GPS-observed activity data in order to verify whether their travel mileage were significantly different in terms of where and when and to what extent. As a result, this study found that travel mileage of drivers who were involved in crashes were higher than those of drivers who were not involved in any crashes. The summary regarding travel mileage exposure is as follows:

1. Travel mileage appears to have a positive relationship with the crash involvement rate. This study showed that the mileages of crash-involved drivers were significantly higher (32 %) than those of drivers who were not involved in crashes. Thus, this result supports the conventional exposure theory, where higher mileage exposure means higher opportunity of being involved in a crash, and suggests that travel mileage exposure metrics can be used as one of potential behavioral crash exposure measures for identifying potentially high crash risk drivers.
2. The distributions of travel mileage by facility types between the two driver-groups were significantly different ($\alpha = 0.05$), indicating that crash-involved drivers were more likely to undertake freeway mileages than drivers who were not involved in crashes (the difference of 32 %). Further analyses will need to have more information associated with each individual crash event so that crash location (facility type) and time of day can be brought into these types of analyses.
3. The distributions of travel activities by trip-time periods were not significantly different between the two driver-groups, indicating that the choices on trip time by drivers between who were involved and not involved in crashes were not different. However, when examining the facility-time-specific travel mileage, drivers involved in crashes tended to travel significantly more on freeways during peak times (possibly congested periods) having many possible conflicts (am peak: 54 % and pm peak: 44 %). Details on crash event locations would be useful in further analyses.

4. Drivers involved in crashes also tended to travel significantly more on freeways during the nighttime, possibly reducing visual abilities and increasing glare sensitivity, resulting in the difference of 51 %.
5. Although total outside-regional travel mileage of the two driver-groups was not significantly different, outside-regional travel mileage by time of day, especially during morning and afternoon showed significant differences (morning: 49 % and afternoon: 53 %).
6. Insurance companies simply using only total mileage estimates for insurance premium structures such as the PAYD insurance program may enhance current insurance classification decision rules with the disaggregated mileage exposure metrics if roadway characteristics are available.

This chapter discussed the relationship between travel mileage exposure and crash involvement and examined the differences in mileage exposures between crash-involved and non-crash-involved drivers at detail. However, using the selected six mileage exposure metrics, this study found that only 57.7 % of crash-involved drivers were correctly classified by the linear discriminant analysis (LDA), indicating that other potential metrics such as travel duration (travel time) need to be investigated. This study examines the relationship between travel duration and crash involvement rate in next chapter.

Chapter Six

POTENTIAL BEHAVIORAL EXPOSURE II: Travel Duration

Analysis on Travel Duration Differences

In addition to the travel mileage exposure, this chapter examined whether travel duration between the two driver-groups were different. As total travel mileage showed the significant difference between the two driver-groups in the previous chapter, the difference of total travel duration (travel time) between crash-involved and crash-not-involved drivers were also significant based on both bootstrap technique (Figure 45) and the Wilks' lambda test (p-value: 0.006, $\alpha = 0.05$). This result indicates that the total travel duration could be used for clustering drivers who potentially have high crash involvement rates from the driver population.

The mean of total travel duration estimated from drivers who were involved in crashes was 921,972 seconds/six-months (256 hours/six-months) and that of drivers who were not involved in crashes was 722,056 seconds/six-months (201 hours/six-months), showing a difference of 21 %. This difference was smaller that that of travel mileage (32 %) in Chapter 5. Similar to the total mileage exposures, this result also supports the general definition of the exposure where higher exposure on roadways means higher risk of crash involvements.

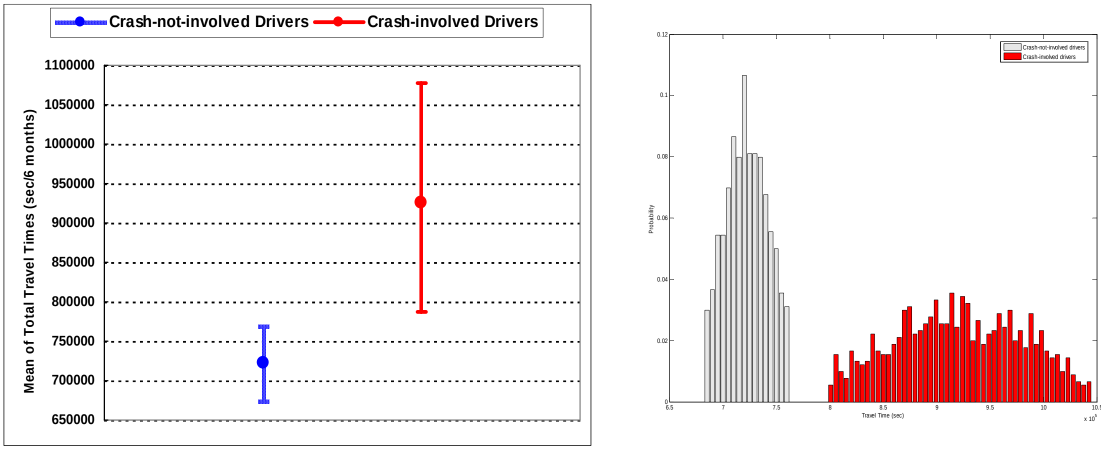


Figure 45: Differences of Means of Total Travel Duration

Similar to the mileage exposure metrics, the bootstrap technique did not provide significant facility-time-specific duration exposure metrics at the 0.05 significance level (Figure 46), but the Wilks' lambda test provided six duration exposure metrics showing differences between the two driver-groups (Table 21).

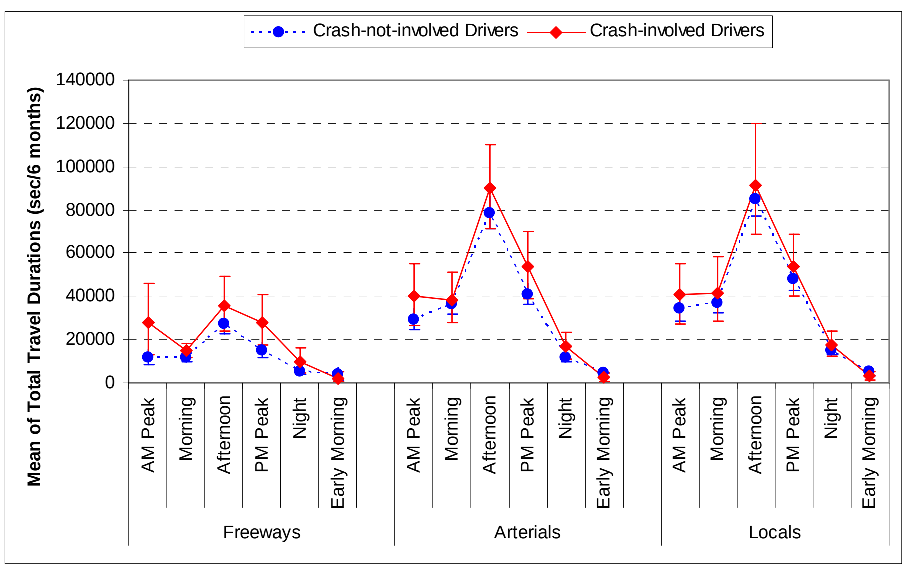


Figure 46: Confidence Intervals of Facility-Time-Specific Travel Duration inside 13 Counties using the Bootstrap Technique

Table 21: Tests of Equality of Means of Facility-Time-Specific Travel Duration Based on the Wilks' Lambda Test

Area	Facility	Wilks' Lambda	F	Sig.
Freeways	**AM peak ***	**0.948**	**8.979**	**0.003**
	Morning	0.980	3.337	0.070
	Afternoon	0.982	2.957	0.087
	PM peak *	**0.948**	**9.085**	**0.003**
	Night *	**0.967**	**5.632**	**0.019**
	Early morning	0.998	0.317	0.574
Arterials	**AM peak ***	**0.969**	**5.276**	**0.023**
	Morning	0.996	0.594	0.442
	Afternoon	0.990	1.592	0.209
	PM peak *	**0.970**	**5.042**	**0.026**
	Night *	**0.971**	**4.961**	**0.027**
	Early morning	0.998	0.390	0.533
Local Roads	AM peak	0.988	2.040	0.155
	Morning	0.978	3.796	0.053
	Afternoon	1.000	0.002	0.961
	PM peak	0.998	0.342	0.560
	Night	1.000	0.004	0.949
	Early morning	0.996	0.689	0.408

* indicates a significant mean difference ($\alpha = 0.05$).

While travel mileage between the two groups during AM peak and nighttime on arterials were not different, travel duration during the same periods on arterials provided significant differences between the two driver-groups, which may imply that crash-involved drivers operated more under congested conditions, but operation conditions such as speed limit and the corresponding speed patterns should be further investigated. Similar to the mileage exposure metrics, duration exposure metrics on local roadways did not provide any significant differences. Travel duration exposure metrics such as peak periods and nighttime freeway travel duration were significantly different between the two driver-groups, but those differences were already showed by the travel mileage, so the differences in those travel duration may be correlated with differences in travel mileage.

Table 22 shows the result of the differences in travel duration based on the pairs of facility and trip time. Travel durations on freeways during AM peak provided the largest difference (58%) between crash-involved drivers (27,415 seconds/six-months or 7.62 hrs/six-months) and drivers who were not involved in crashes (11,644 seconds/six-months or 3.23 hrs/six-months) and showed a significant difference in the means between the two driver-groups[15] based on the Wilks' lambda test (Table 21 and 22). In addition, travel duration on freeways during PM peak showed the difference of 47% between the two groups with statistical significance ($\alpha = 0.05$) (Table 21 and 22).

Table 22: Facility-Time-Specific Travel Duration and Differences in 13 Counties

Facility	Travel Time	Mean of Total Travel Duration (seconds/6-months): Drivers who *were not involved* in crashes		Mean of Total Travel Duration (seconds/6-months): Drivers who *were involved* in crashes		Duration Difference	% Difference
		Second	%	Second	%		
Freeways	**AM Peak ***	**11,644**	**2.33**	**27,415**	**4.52**	**15,771**	**58**
	Morning	11,515	2.31	14,787	2.44	3,272	22
	Afternoon	26,895	5.39	35,890	5.92	8,995	25
	PM Peak *	**14,649**	**2.94**	**27,765**	**4.58**	**13,116**	**47**
	Night *	**4,871**	**0.98**	**9,416**	**1.55**	**4,545**	**48**
	Early Morning	3,689	0.74	2,231	0.37	-1,458	-65
Arterials	**AM Peak ***	**29,631**	**5.94**	**40,333**	**6.65**	**10,702**	**27**
	Morning	36,475	7.31	38,245	6.31	1,774	5
	Afternoon	78,446	15.73	90,282	14.89	11,836	13
	PM Peak *	**40,892**	**8.20**	**53,803**	**8.87**	**12,911**	**24**
	Night *	**11,727**	**2.35**	**16,471**	**2.72**	**4,744**	**29**
	Early Morning	4,108	0.82	2,370	0.39	-1,738	-73
Local Roads	AM Peak	33,973	6.81	41,359	6.82	7,386	18
	Morning	37,277	7.47	41,233	6.80	3,956	10
	Afternoon	85,275	17.09	90,473	14.92	5,197	6
	PM Peak	47,892	9.60	53,576	8.83	5,684	11
	Night	14,843	2.98	17,712	2.92	2,869	16
	Early Morning	5,061	1.01	3,057	0.50	-2,003	-66

* indicates a significant mean difference ($\alpha = 0.05$).

[15] Travel mileages on freeways during AM peak also provided the largest difference (54%) between crash-involved and crash-not-involved drivers with statistically significance between the two driver-groups.

Analysis on Travel Duration outside 13-County Area (Outside Regional Duration)

Figure 47 shows the means of trip time-based travel duration outside 13-counties between the two driver-groups with and without crash involvements using the bootstrap technique.

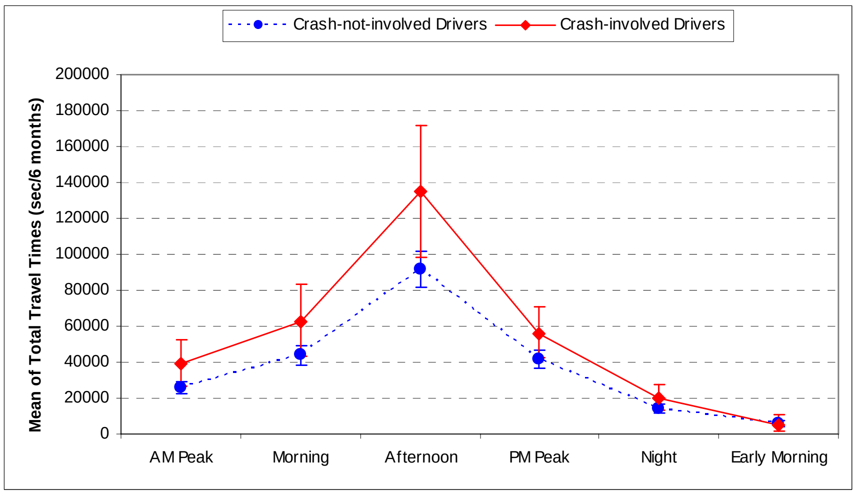


Figure 47: Confidence Intervals of Means of Outside-Regional Travel Duration Using the Bootstrap Technique

Similar to the outside-regional travel mileage exposure metrics in Chapter 5, the bootstrap technique did not provide significant outside-regional duration-related exposure metrics at the 0.05 significance level, but the Wilks' lambda test provided five potential duration exposure metrics. Table 23 shows the result of the tests of mean equality using the Wilks' lambda test, and Table 24 illustrates differences in outside-regional travel duration between the two driver-groups.

Table 23: Tests of Equality of Means of Outside-Regional Travel Duration Based on the Wilks' Lambda Test

Area	Facility	Wilks' Lambda	F	Sig.
Outside 13 Counties	**AM peak ***	**0.928**	**12.835**	**0.000**
	Morning *	**0.916**	**15.192**	**0.000**
	Afternoon *	**0.916**	**15.215**	**0.000**
	PM peak *	**0.954**	**7.971**	**0.005**
	Night *	**0.951**	**8.590**	**0.004**
	Early morning	0.996	0.635	0.427

* indicates a significant mean difference ($\alpha = 0.05$).

As shown in Table 24, the mean of travel duration of drivers who were involved in crashes was 134,520 seconds/six-months (37.37 hrs/six-months) during afternoon (12pm ~ 6pm) while that of drivers who were not involved in crashes was 91,952 seconds/six-months (25.54 hrs/six-months), providing the difference of 32 %.

Table 24: Outside Regional Time-Specific Duration and Differences

Area	Trip Time	Mean of Travel Duration (Secconds/6-months): Drivers who *were not involved* in crashes		Mean of Travel Duration (Secconds/6-months): Drivers who *were involved* in crashes		Duration Difference	% Difference
		Second	%	Second	%		
Outside 13 Counties	**AM Peak ***	**25,755**	**11.53**	**39,117.68**	**12.31**	**13,362.59**	**34**
	Morning *	**43,891**	**19.65**	**62,428.44**	**19.64**	**18,536.52**	**30**
	Afternoon *	**91,952.78**	**41.17**	**134,520.26**	**42.32**	**42,567.47**	**32**
	PM Peak *	**42,064.17**	**18.83**	**56,310.81**	**17.72**	**14,246.64**	**25**
	Night *	**13,898.71**	**6.22**	**20,024.89**	**6.30**	**6,126.18**	**31**
	Early Morning	5,787.45	2.59	5,438.24	1.71	-349.21	-6

* indicates a significant mean difference ($\alpha = 0.05$).

Except nighttime outside-regional duration, all outside-regional duration exposure metrics were already found from the outside-regional mileage metrics, which may indicate that crash-involved drivers traveled on outside-regional roadways having low speed limits during nighttime or traveled more congested roadways. However, due to the unavailability of roadway characteristics information in outside region, this issue cannot be investigated at this time. Further studies regarding this issue need to be performed in future.

The Linear Discriminant Analysis using Travel Duration Exposure

Similar to the analysis of mileage exposure metrics, this study also performed the correlation analysis with 12 pre-selected duration-related exposure metrics before utilizing the linear discriminant analysis (LDA) (Table 25).

Table 25: Correlation Analysis of Travel Duration Exposure Metrics

		Freeway			Arterial			Outside					All
		AM Peak	PM Peak	Night	AM Peak	PM Peak	Night	AM Peak	Morning	Afternoon	PM Peak	Night	Total
Freeways	AM Peak	1.00	-	-	-	-	-	-	-	-	-	-	-
	PM Peak	**0.70**	1.00	-	-	-	-	-	-	-	-	-	-
	Night	0.05	0.40	1.00	-	-	-	-	-	-	-	-	-
Arterials	AM Peak	0.34	0.18	-0.10	1.00	-	-	-	-	-	-	-	-
	PM Peak	0.23	0.39	0.12	0.44	1.00	-	-	-	-	-	-	-
	Night	-0.06	0.12	0.36	0.07	0.49	1.00	-	-	-	-	-	-
Outside 13 Counties	AM Peak	0.19	0.08	-0.08	0.42	0.29	0.12	1.00	-	-	-	-	-
	Morning	-0.13	-0.02	-0.02	-0.13	-0.01	0.11	0.34	1.00	-	-	-	-
	Afternoon	-0.12	0.00	0.07	-0.15	-0.05	0.11	0.21	**0.77**	1.00	-	-	-
	PM Peak	0.06	0.24	0.19	0.10	0.42	0.40	**0.60**	0.44	0.51	1.00	-	-
	Night	0.06	0.23	0.44	-0.03	0.32	**0.67**	0.26	0.36	0.42	**0.66**	1.00	-
All	Total	0.36	0.42	0.17	0.57	0.57	0.27	0.36	0.21	0.25	0.31	0.29	1.00

Using the structure coefficients obtained from the linear discriminant analysis (Table 26), this study tried to select potential travel-duration exposure metrics having higher structure coefficients (loading power) among the correlated variables and finally obtained eight duration-related exposure metrics that can be potentially used for classifying drivers into the two risk groups.

Table 26: Structure Coefficients from the Linear Discriminant Analysis Using Travel Duration Exposure Metrics

Facility	Time of Day	Structure Coefficients	Rank	Variable Selection
Outside	**Afternoon**	**0.550**	**1**	√
Outside	Morning	0.550	2	-
Outside	**AM Peak**	**0.505**	**3**	√
Freeway	**PM Peak**	**0.425**	**4**	√
Freeway	AM Peak	0.423	5	-
Outside	**Night**	**0.414**	**6**	√
Outside	PM Peak	0.398	7	-
All	**All**	**0.392**	**8**	√
Freeway	**Night**	**0.335**	**9**	√
Arterial	**AM Peak**	**0.324**	**10**	√
Arterial	**PM Peak**	**0.317**	**11**	√
Arterial	Night	0.314	12	-

Finally, this study repeated the linear discriminant analysis with the selected eight duration-related exposure metrics (Table 26 and 27). As a result, 76.6 % of drivers who were not involved in crashes and 65.4 %[16] of drivers who were involved in crashes were correctly classified. Overall accuracy of the model using those duration-related exposure metrics was 74.9 % (Overall performance of the model using those mileage-related exposure metrics was 77.2 %). Similar to the analysis on the mileage metrics, 23.4 % of drivers who were not involved in crashes were classified into a group who has a potentially high crash involvement rate based on their travel-duration exposures.

For the 34.6 % of drivers who were involved in crashes, they were theoretically classified into a low crash involvement group based on their travel duration exposures although they actually had crash involvements. Similar to the modeling process using travel mileage exposure metric, this result also indicates that the model has a limitation for classifying drivers who have potentially high crash involvement rate based on

[16] It was 57.7 % when using the travel mileage exposure metrics.

duration exposures only and more likely, other interactions between drivers and vehicles such as speeding and acceleration patterns may be more helpful.

Table 27: Classification Results Using Travel Duration Exposure Metrics

Crash Involvements	Predicted Group Membership	
	Drivers who *were not involved* in crashes	Drivers who *were involved* in crashes
Drivers who *were not involved* in crashes	76.6 %	23.4 %
Drivers who *were involved* in crashes	34.6 %	65.4 %

Summary of the Travel Duration Exposure

This study evaluated differences in travel duration of drivers with and without crash involvements during the 14-months study period to assess whether their travel duration exposures were significantly different with respect to the facility type and time of day. As a result, this study found that travel duration of drivers who were involved in crashes were higher than those of drivers who were not involved in any crashes. The detailed summary of the travel duration exposure is followings:

1. Similar to the travel mileage exposure, the travel duration exposure also appeared to have a positive relationship with the crash involvement rate. Based on total travel duration metrics, crash-involved drivers traveled significantly more hours (21 %) compared to non-crash-involved drivers. This result can support the exposure theory where higher exposure to roadways indicates higher opportunity of being involved in a crash.

2. Travel duration on freeways during peak periods and nighttime and outside-regional travel duration of crash-involved drivers were significantly higher than those of non-crash-involved drivers, but those differences were already showed by the travel mileage. It may indicate that those differences in travel duration might be explained by differences in travel mileage.
3. While travel mileage between the two driver-groups during AM peak and nighttime on arterials were not different, duration measures during the same periods on arterials provided significant differences between the two driver-groups, implying that crash-involved drivers might operate more under congested conditions. However, before that, speed patterns and frequency of turn movement activities on arterials during those periods between the two driver-groups need to be further examined.
4. The overall performances of classification using travel duration metrics was 74.9 % and using travel mileage metric was 77.2 %. However, the performances of classification for crash-involved drivers were 65.4 % and 57.7 % by employing travel duration and mileage metrics, respectively.
5. This study suggests that travel duration exposure metrics can be used as one of behavioral crash exposure measures for identifying drivers who have potentially high crash involvement rates.

This study discussed the relationship between crash involvement and travel mileage and duration exposure metrics and examined the differences in those behavior activities between crash-involved and non-crash-involved drivers at detail. However, these mileage and duration measures cannot represent how drivers interact with other traffic or roadway conditions. To evaluate the relationships between these interactions and crash involvement rate, this study assess speed activities of the two driver-groups in next chapter.

Chapter Seven

POTENTIAL BEHAVIORAL EXPOSURE III: Speed

This study examines the differences of facility-time-specific speed behavior between the two driver-groups since speeding is the most contributing factor of motor vehicle crashes as mentioned in the previous literature review chapter. To evaluate the differences in speed activity patterns between the two driver-groups, this study initially selected eight potential speed-related exposure measures based upon literature reviews [5, 8].

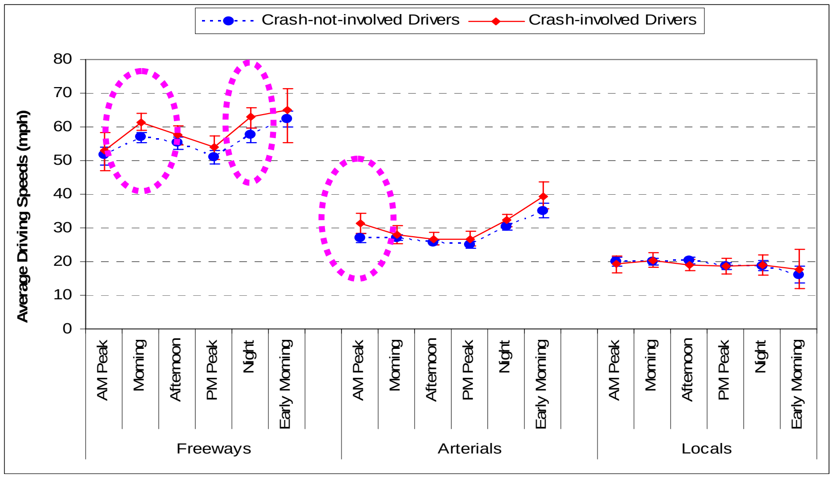
- Mean Speed: The mean speed indicates an average driving speed based on the specific facility and time of day, so this measure includes all speed data.
- Running Speed: The mean of running speed indicates an average speed excluding idling and delay speeds travel at than 5 mph.
- Delta Speed between Driving Speed and Speed Limit: This measure indicates differences between GPS-observed driving speed and the corresponding speed limit.
- Positive Delta Speed between Driving Speed and Speed Limit: This measure includes the only positive delta speeds between driving speeds and the corresponding speeds, which indicates the amount of over speeding behavior
- Frequency of Over-Speed Behavior Activities per Mile: This measure indicates how frequently driving speeds are greater than the posted speed limit.

- Frequency of 10 mph Over-Speed Behavior Activities per Mile: This measure indicates how frequently a driver travels at driving speeds of 10 mph over than the posted speed limit.
- Frequency of 15 mph Over-Speed Behavior Activities per Mile: This measure indicates how frequently a driver travels at driving speeds of 15 mph over than the posted speed limit.
- Frequency of 20 mph Over-Speed Behavior Activities per Mile: This measure indicates how frequently a driver travels at driving speeds of 20 mph over than the posted speed limit.

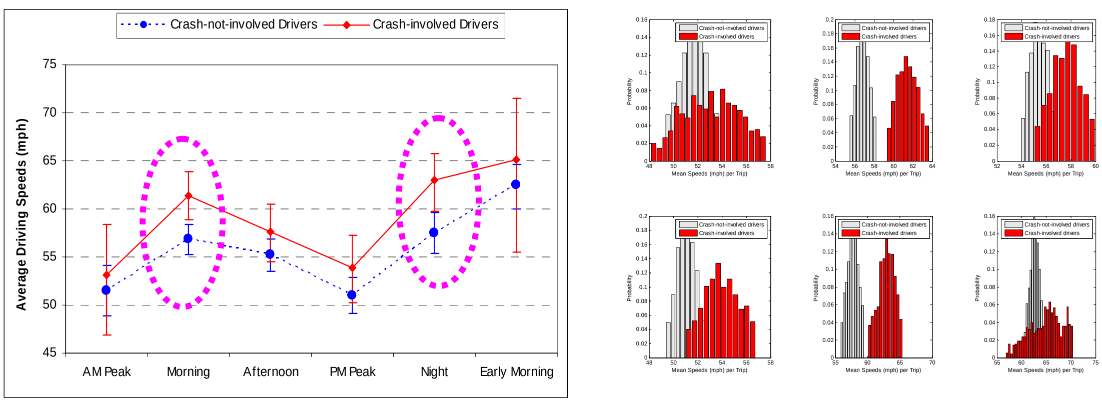
Differences in Mean Speeds

Figure 48 shows the overall means of driving speeds based on facility types and the corresponding time of day between the two driver-groups and indicates that the mean of driving speeds of drivers who were involved in crashes was generally higher than those of drivers who were not involved in crashes.

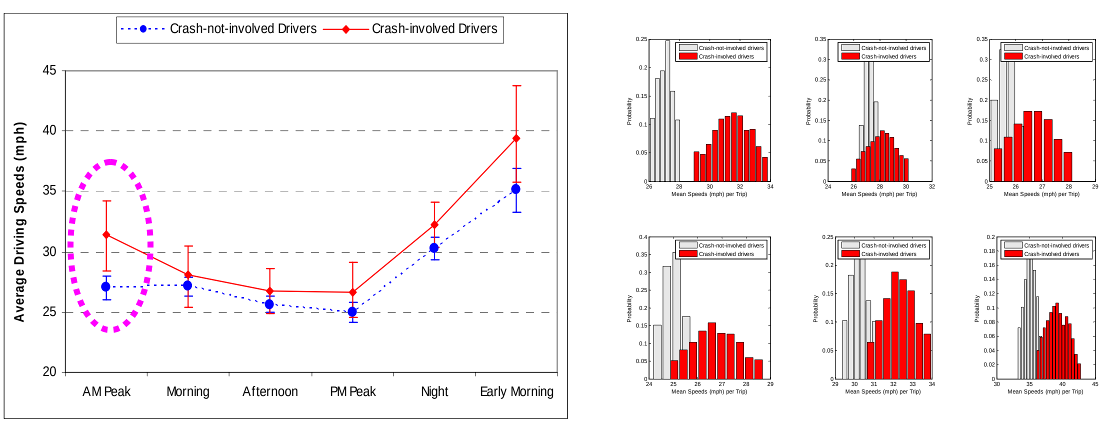
Based on the bootstrap technique and the Wilks' lambda test, this study found three driving speed exposure metrics were significantly different between the two driver-groups (Figure 48 and Table 28).



(A) Means of Average Driving Speeds based on all Facilities between Groups



(B) Means of Average Driving Speeds on Freeways between Groups



(C) Means of Average Driving Speeds on Arterials between Groups

Figure 48: Confidence Intervals and Mean Distributions of Average Driving Speeds Using the Bootstrap Technique

Table 28: Tests of Equality of Means of Average Driving Speeds based on the Wilks' Lambda Test

Facility	Trip Time	Wilks' Lambda	F	Sig.
Freeways	AM peak	0.998	0.198	0.657
	Morning *	**0.969**	**5.155**	**0.025**
	Afternoon	0.992	1.358	0.246
	PM peak	0.990	1.516	0.220
	Night *	**0.968**	**4.471**	**0.036**
	Early morning	0.991	0.590	0.445
Arterials	**AM peak ***	**0.958**	**6.993**	**0.009**
	Morning	0.996	0.685	0.409
	Afternoon	0.994	1.077	0.301
	PM peak	0.989	1.906	0.169
	Night	0.984	2.562	0.112
	Early morning	0.979	2.115	0.149
Local Roads	AM peak	0.999	0.197	0.658
	Morning	1.000	0.058	0.811
	Afternoon	0.994	0.933	0.336
	PM peak	1.000	0.003	0.958
	Night	1.000	0.007	0.933
	Early morning	0.998	0.266	0.607

* indicates a significant mean difference ($\alpha = 0.05$).

As shown in Table 29, crash-involved drivers were more likely to travel at higher speeds on each facility type and time of day (a surrogate for traffic flow and speeds) than drivers who were not involved in crashes. The average driving speed on freeway during nighttime of drivers who were involved in crashes was 63 mph and that of drivers who were not involved in crashes was 58 mph, showing a difference of 5 mph (21 %). Mileage and duration exposures on freeways during morning did not show any differences between the two driver-groups, but average driving speeds showed the significant difference. Drivers who had crash involvements traveled at almost 5 mph higher than drivers who did not have crashes.

In previous chapters, this study discussed that travel duration on arterials during nighttime provided significant differences between the two driver-groups while travel mileage between the two groups were not different. The current chapter shows that average driving speeds on arterials during nighttime between the two driver-groups were not significantly different.

Finally, Similar to mileage and duration exposure measures, average driving speeds on local roadways did not provide any significant differences between the two driver-groups. Other potential exposure measures need to be examined for the local roadways.

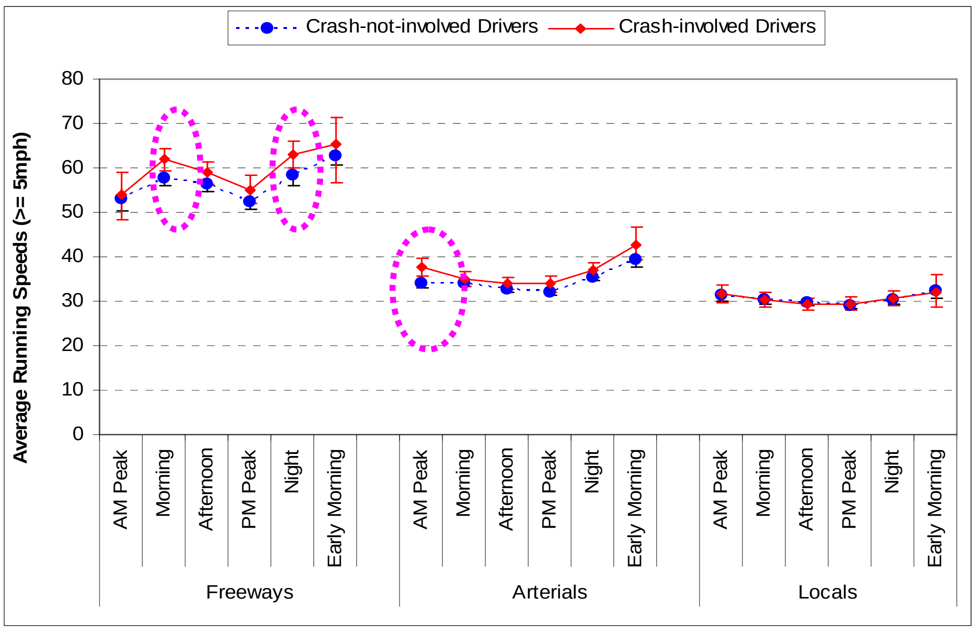
Table 29: Differences and Means of Average Speeds based on Facility and Time

Facility	Trip Time	Mean of Speeds		Speed Difference	% Difference
		Drivers who *were not involved* in Crashes	Drivers who *were involved* in Crashes		
Freeways	AM Peak	51.64	53.08	1.45	3
	Morning*	**56.84**	**61.46**	**4.63**	**8**
	Afternoon	55.33	57.57	2.24	4
	PM Peak	51.01	53.95	2.94	5
	Night*	**57.69**	**62.91**	**5.22**	**8**
	Early Morning	62.55	64.73	2.18	3
Arterials	**AM Peak ***	**27.06**	**31.42**	**4.36**	**14**
	Morning	27.16	28.10	0.94	3
	Afternoon	25.65	26.64	0.99	4
	PM Peak	24.96	26.77	1.81	7
	Night	30.25	32.32	2.07	6
	Early Morning	35.08	39.28	4.21	11
Local Roads	AM Peak	19.94	19.17	-0.78	-4
	Morning	20.16	20.47	0.31	2
	Afternoon	20.28	19.12	-1.16	-6
	PM Peak	18.71	18.61	-0.10	-1
	Night	18.76	18.99	0.22	1
	Early Morning	16.00	17.60	1.60	9

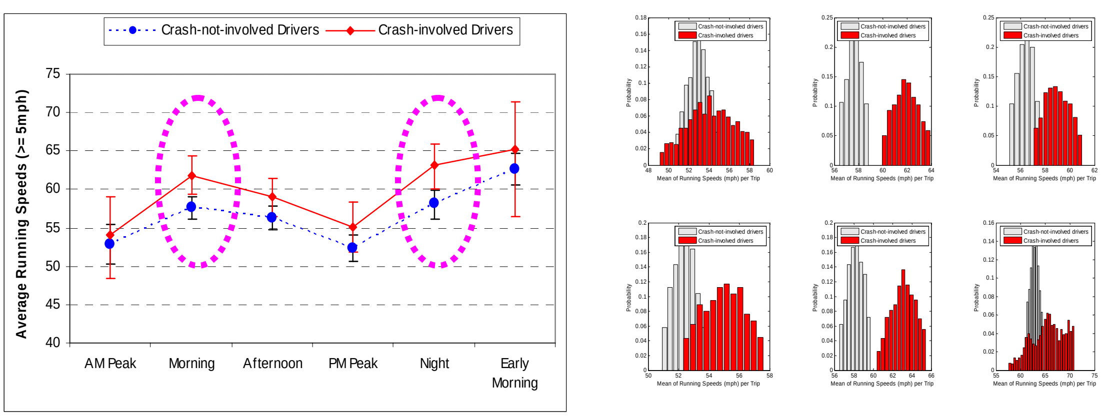
* indicates a significant mean difference ($\alpha = 0.05$).

Differences in Average Running Speeds

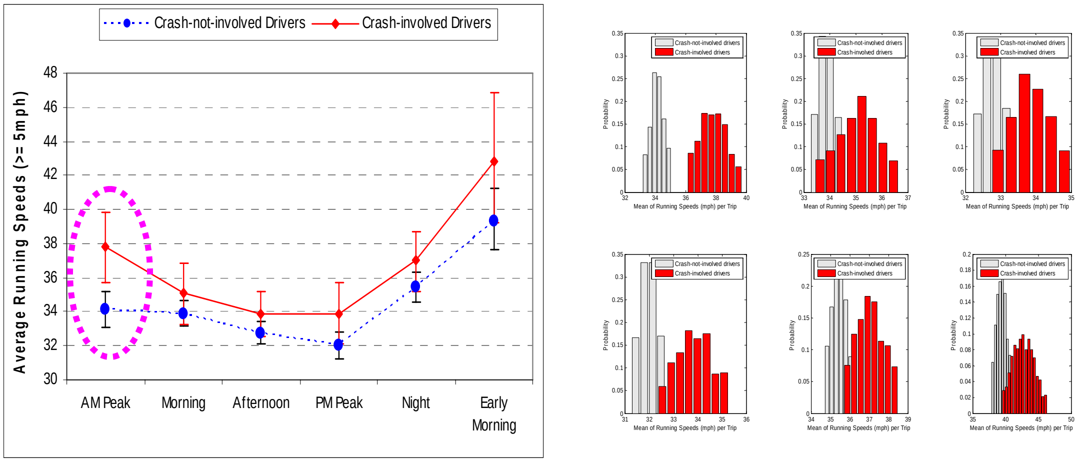
When studying central values of speeding behavior, it is also important to consider the effects of stop delays and slow moving traffic on overall speeding pattern. Thus, average running speeds (excluding all speeds less than 5 mph) (Figure 49) were used to examine the speeding behavior differences between the two driver-groups.



(A) Means of Average Running Speeds based on all Facilities between Groups



(B) Means of Average Running Speeds on Freeways between Groups



(C) Means of Average Running Speeds on Arterials between Groups

Figure 49: Confidence Intervals and Mean Distributions of Average Running Speeds Using the Bootstrap Technique

Similar to the bootstrap analysis, the Wilks' lambda test also showed that three exposure metrics, average running speeds on freeways during morning and nighttime and arterials during AM peak, were significantly different (Table 30).

Table 30: Tests of Equality of Means of Average Running Speeds based on the Wilks' Lambda Test

Facility	Trip time	Wilks' Lambda	F	Sig.
Freeways	AM Peak	0.999	0.104	0.748
	Morning *	**0.970**	**4.903**	**0.028**
	Afternoon	0.988	1.975	0.162
	PM Peak	0.991	1.414	0.236
	Night *	**0.968**	**4.434**	**0.037**
	Early Morning	0.991	0.609	0.438
Arterials	**AM Peak ***	**0.957**	**7.078**	**0.009**
	Morning	0.991	1.551	0.215
	Afternoon	0.989	1.764	0.186
	PM Peak	0.981	3.116	0.079
	Night	0.987	2.046	0.155
	Early Morning	0.979	2.105	0.150
Local Roads	AM Peak	0.999	0.180	0.672
	Morning	1.000	0.021	0.885
	Afternoon	1.000	0.035	0.852
	PM Peak	0.999	0.139	0.709
	Night	0.999	0.100	0.752
	Early Morning	1.000	0.006	0.936

* indicates a significant mean difference ($\alpha = 0.05$).

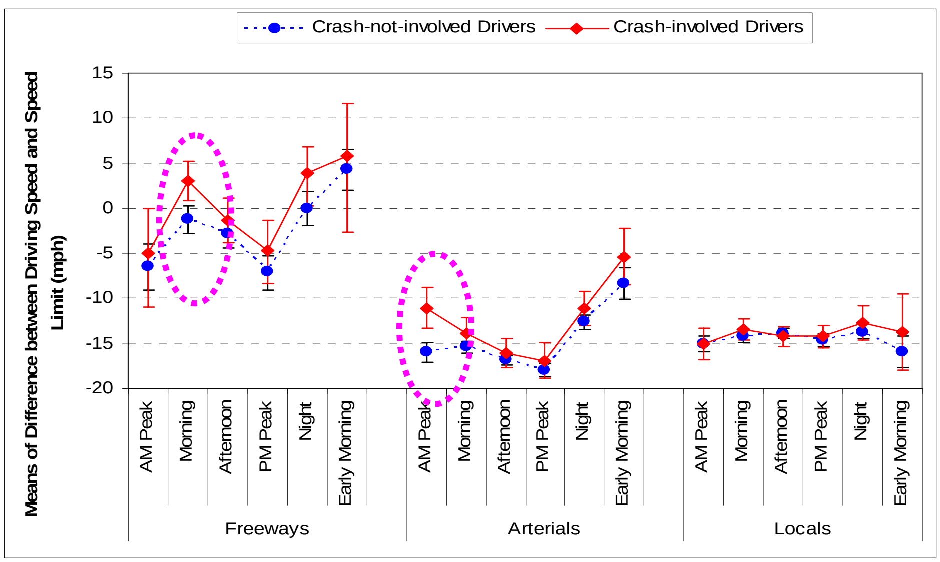
Table 31: Means of Average Running Speeds based on Facility and Time

Facility	Trip Time	Mean of Running Speeds		Speed Difference	% Difference
		Drivers who *were not involved* in Crashes	Drivers who *were involved* in Crashes		
Freeways	AM Peak	52.88	54.07	1.19	2
	Morning *	**57.62**	**61.94**	**4.31**	**7**
	Afternoon	56.36	58.98	2.62	4
	PM Peak	52.35	54.98	2.63	5
	Night *	**58.17**	**63.08**	**4.91**	**8**
	Early Morning	62.73	65.33	2.60	4
Arterials	**AM Peak ***	**34.14**	**37.84**	**3.70**	**10**
	Morning	33.87	35.08	1.21	3
	Afternoon	32.76	33.87	1.11	3
	PM Peak	31.97	33.85	1.88	6
	Night	35.43	37.05	1.62	4
	Early Morning	39.38	42.78	3.40	8
Local Roads	AM Peak	31.14	31.84	0.71	2
	Morning	30.18	30.32	0.14	0
	Afternoon	29.62	29.48	-0.14	0
	PM Peak	29.00	29.42	0.42	1
	Night	30.30	30.71	0.41	1
	Early Morning	32.18	32.08	-0.10	0

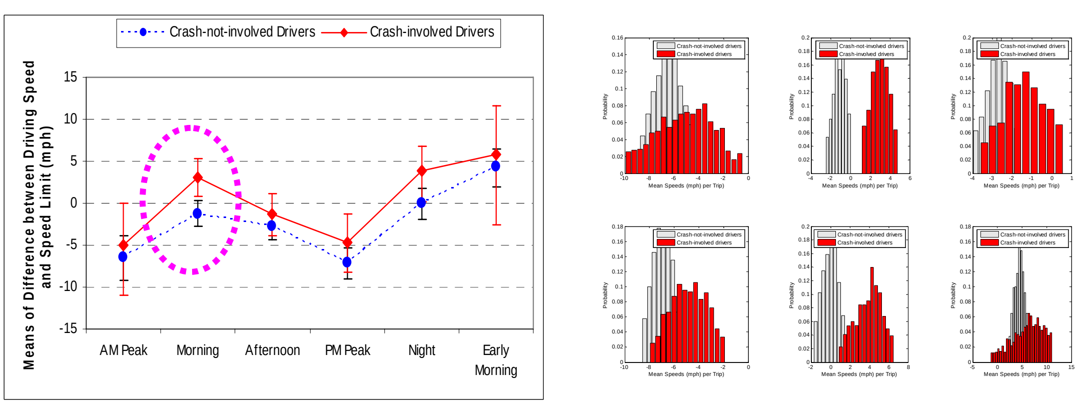
* indicates a significant mean difference ($\alpha = 0.05$).

Differences in Delta Speed between Driving Speed and Speed Limit

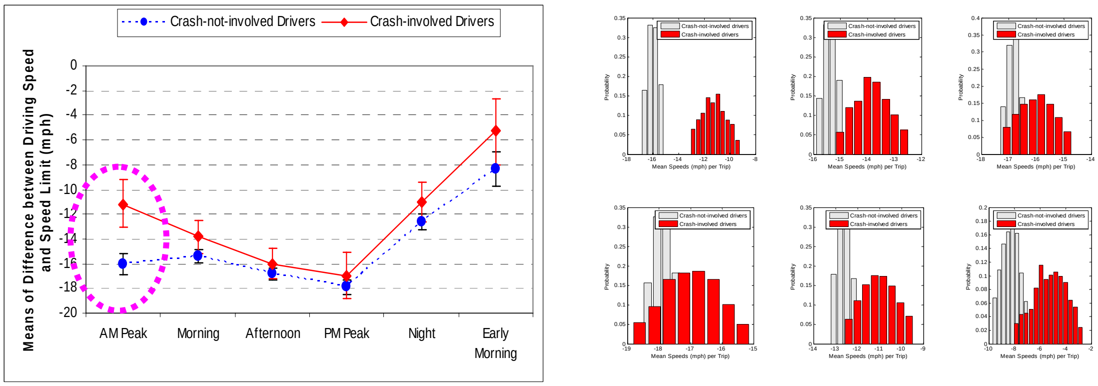
In addition to average driving speeds and running speeds, delta speeds indicating the speed difference from the posted speed limit are important to compare speeding behavior between the two driver-groups. The bootstrap technique (Figure 50) and the Wilks' lambda test (Table 32) showed that delta speeds on freeways during morning and on arterials during AM peak were significantly different between the two driver-groups at the 0.05 significance level.



(A) Average Delta Speeds based on all Facilities between Groups



(B) Average Delta Speeds on Freeways between Groups



(C) Average Delta speeds on Arterials between Groups

Figure 50: Confidence Intervals and Mean Distributions of Delta Speeds Using the Bootstrap Technique

Table 32: Tests of Equality of Means of Delta Speeds based on the Wilks' Lambda Test

Facility	Trip Time	Wilks' Lambda	F	Sig.
Freeway	AM peak	0.999	0.179	0.673
	Morning *	**0.968**	**5.213**	**0.024**
	Afternoon	0.996	0.634	0.427
	PM peak	0.994	0.959	0.329
	Night	0.979	2.841	0.094
	Early morning	0.997	0.194	0.661
Arterial	**AM peak ***	**0.926**	**12.392**	**0.001**
	Morning	0.977	3.860	0.051
	Afternoon	0.991	1.397	0.239
	PM peak	0.993	1.110	0.294
	Night	0.982	2.760	0.099
	Early morning	0.983	1.702	0.195
Local	AM peak	1.000	0.001	0.981
	Morning	0.996	0.687	0.408
	Afternoon	0.999	0.239	0.625
	PM peak	0.999	0.126	0.723
	Night	0.995	0.840	0.361
	Early morning	0.993	0.947	0.332

* indicates a significant mean difference ($\alpha = 0.05$).

Table 33 showed that the mean of delta speeds during morning on freeways of drivers who were involved in crashes was +2.99 mph and that of drivers who were not involved in crashes was -1.15 mph, resulting the difference of 4 mph with the statistical significance. In addition, the mean of delta speeds during AM peak on arterials of drivers who were involved in crashes was -11 mph and that of drivers who were not involved in crashes was -16 mph, resulting the difference of 5 mph with the statistical significance.

Although not showing significant difference between the two driver-groups, the mean of delta speeds on freeways during nighttime of drivers having crash involvements was slightly higher (4 mph) than that of drivers who did not have crash involvements.

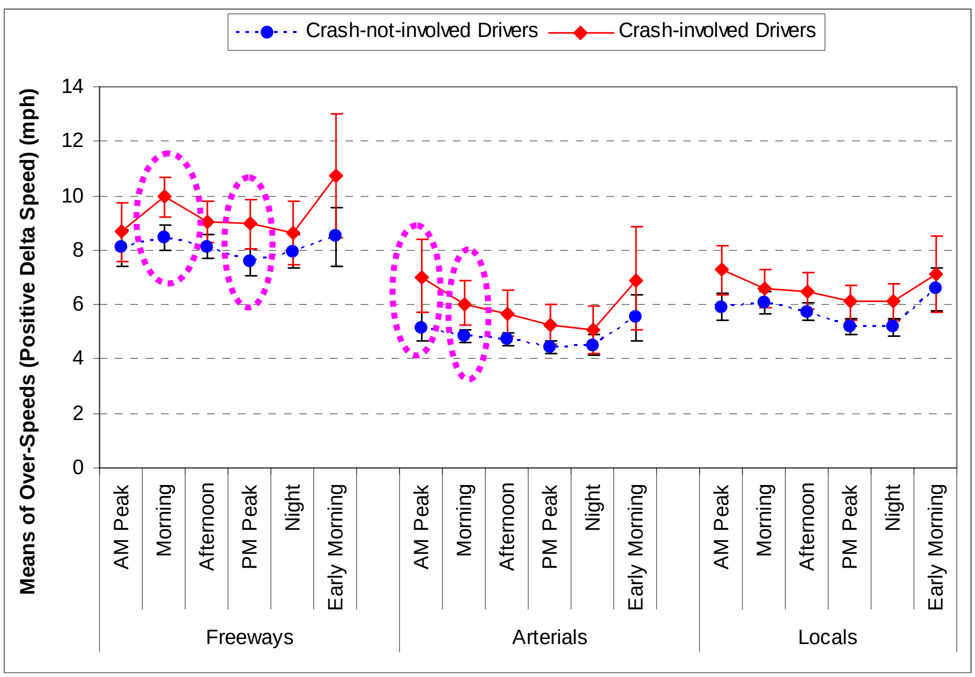
Table 33: Means of Average Delta Speeds based on Facility and Time

Facility	Trip Time	Mean of Delta Speeds		Speed Difference	% Difference
		Drivers who *were not involved* in Crashes	Drivers who *were involved* in Crashes		
Freeways	AM Peak	-6.53	-5.12	1.41	27
	Morning *	**-1.15**	**2.99**	**4.14**	**139**
	Afternoon	-2.75	-1.41	1.35	96
	PM Peak	-7.02	-4.84	2.18	45
	Night	-0.07	3.88	3.96	102
	Early Morning	4.38	5.88	1.50	26
Arterials	**AM Peak ***	**-16.00**	**-11.22**	**4.78**	**43**
	Morning	-15.40	-13.85	1.55	11
	Afternoon	-16.84	-15.99	0.85	5
	PM Peak	-17.90	-16.96	0.94	6
	Night	-12.63	-11.01	1.62	15
	Early Morning	-8.33	-5.25	3.07	59
Local Roads	AM Peak	-15.07	-15.07	-0.01	0
	Morning	-14.20	-13.47	0.73	5
	Afternoon	-13.85	-14.20	-0.35	2
	PM Peak	-14.58	-14.24	0.34	2
	Night	-13.69	-12.78	0.91	7
	Early Morning	-15.92	-13.68	2.24	16

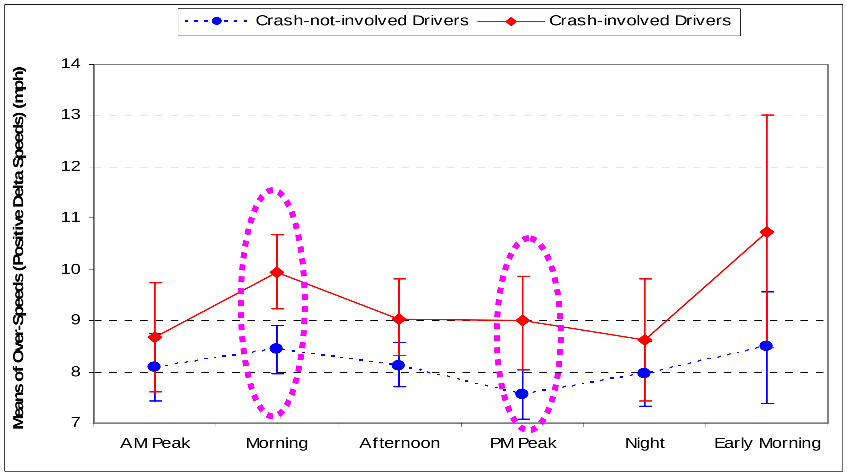
* indicates a significant mean difference ($\alpha = 0.05$).

Differences in Positive Delta Speeds between Driving Speed and Speed Limit

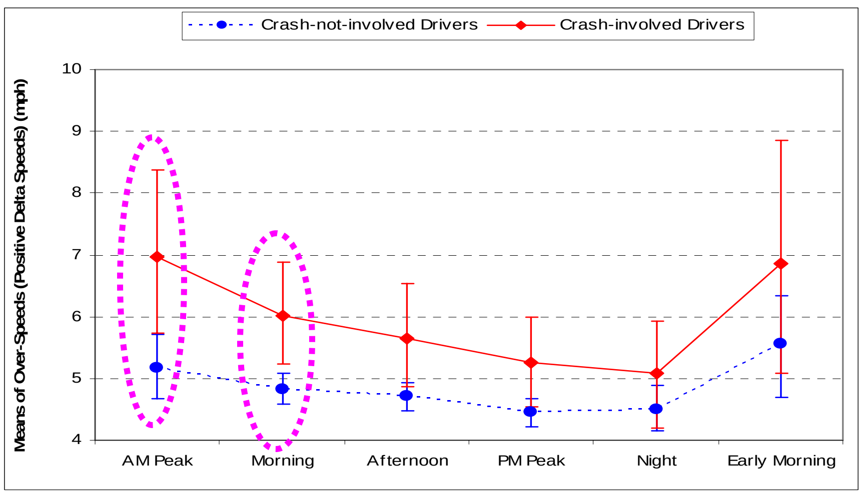
This study assessed positive delta speeds indicating the amount of over-speed between the two driver-groups. Figure 51 shows all positive delta speeds based on facility and time of the two groups. The highest over-speed difference was obtained from freeway travel as expected. This study also found that the amount of over-speed on local roadways was generally higher than that on arterials. This result was caused from the relatively low speed limits on local roadways because the average driving speeds and the average running speeds on local roadways were lower than those on arterials.



(A) Average Positive Delta Speeds based on all Facilities between Groups



(B) Average Positive Delta Speeds on Freeways between Groups



(C) Average Positive Delta Speeds on Arterials between Groups

Figure 51: Confidence Intervals of Average Positive Delta Speeds Using the Bootstrap Technique

While the bootstrap technique provided four significant metrics with respect to the positive delta speeds[17], the Wilks' lambda test showed 10 significant exposure

[17] With the 0.1 significance level, the bootstrap technique provided significant differences in positive delta speeds on local roadways.

metrics (Table 34). This study uses those metric for developing the linear discriminant model.

Table 34: Tests of Equality of Means of Positive Delta Speeds based on the Wilks' Lambda Test

Facility	Trip Time	Wilks' Lambda	F	Sig.
Freeways	AM peak	0.996	0.534	0.466
	Morning *	**0.959**	**6.733**	**0.010**
	Afternoon	0.981	3.103	0.080
	PM peak *	**0.963**	**5.582**	**0.019**
	Night	0.994	0.752	0.387
	Early morning	0.956	2.887	0.094
Arterials	**AM peak ***	**0.958**	**6.906**	**0.009**
	Morning *	**0.935**	**11.494**	**0.001**
	Afternoon *	**0.956**	**7.640**	**0.006**
	PM peak *	**0.964**	**6.101**	**0.015**
	Night	0.989	1.749	0.188
	Early morning	0.985	1.452	0.231
Local Roads	**AM peak ***	**0.970**	**4.885**	**0.029**
	Morning	0.994	1.036	0.310
	Afternoon *	**0.977**	**3.847**	**0.052**
	PM peak *	**0.967**	**5.543**	**0.020**
	Night *	**0.967**	**5.358**	**0.022**
	Early morning	0.996	0.349	0.556

* indicates a significant mean difference ($\alpha = 0.05$).

Table 35 shows the differences in the means of positive delta speeds between the two driver-groups. For the case of speeding behavior on freeways, drivers involved in crashes usually traveled at 10 mph over-speeds than the posted speed limits during morning, and drivers who were not involved in crashes traveled at 8 mph over-speeds. Although their difference was only 2 mph (15 %), this difference was significant, based on both test methods. Over-speeding behavior during PM peak on freeways also showed significant differences at the 0.05 significance level. Specially, this study obtained the differences in over-speed activities on local roadways between the two driver-groups.

Table 35: Means and Differences in Positive Delta Speeds between Groups

Facility	Trip Time	Mean of Positive Delta Speeds		Speed Difference	% Difference
		Drivers who *were not involved* in Crashes	Drivers who *were involved* in Crashes		
Freeways	AM peak	8.09	8.70	0.61	7
	Morning *	**8.44**	**9.93**	**1.49**	**15**
	Afternoon	8.11	9.07	0.96	11
	PM peak *	**7.58**	**8.98**	**1.40**	**16**
	Night	7.95	8.55	0.60	7
	Early morning	8.47	10.80	2.33	22
Arterials	**AM peak ***	**5.16**	**6.95**	**1.79**	**26**
	Morning *	**4.82**	**5.96**	**1.14**	**19**
	Afternoon *	**4.71**	**5.63**	**0.92**	**16**
	PM peak *	**4.45**	**5.28**	**0.83**	**16**
	Night	4.49	5.06	0.57	11
	Early morning	5.52	6.84	1.32	19
Local Roads	**AM peak ***	**5.89**	**7.28**	**1.39**	**19**
	Morning	6.06	6.58	0.52	8
	Afternoon *	**5.75**	**6.50**	**0.75**	**11**
	PM peak *	**5.19**	**6.08**	**0.88**	**15**
	Night *	**5.17**	**6.14**	**0.97**	**16**
	Early morning	6.54	7.14	0.60	8

* indicates a significant mean difference ($\alpha = 0.05$).

Frequency of 10 mph Over-Speed Activities

So far, this study tested differences between the two driver-groups with and without crash involvements based on the "amount" measures. In addition to those "amount" measures, this study examined the differences in driving activity (frequency) of speeding behavior between the two driver-groups. Since the threshold of a legislated violation (ticketable speed limit) in Georgia is equal to the speed 10 mph above than the posted speed limit, this study examined the frequency of driving activities in excess of 10 mph above the posted speed limit [5]. This study used the second-by-second GPS-observed data, so the frequency also expressed for seconds of operation.

Figure 52 showed the average frequency of 10 mph over-speeding activities between the two driver-groups based on facility and travel time. Large numbers of over-speeding activities occurred on freeways than arterials and local roadways while travel mileage and travel duration on arterials were larger than those of freeways. In addition, it was shown that drivers who were involved in crashes were more willing to drive at above than the posted speed limits.

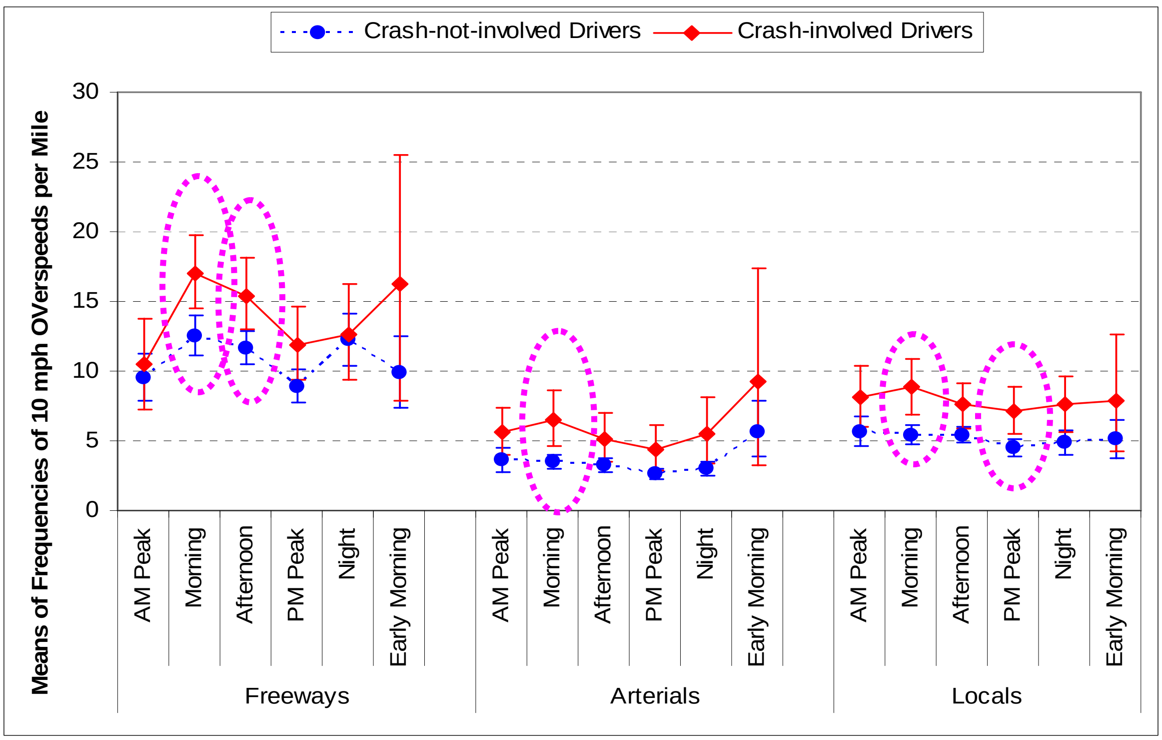


Figure 52: Average Frequencies of 10 mph Over-Speed Activities

While the bootstrap technique provided five exposure metrics, the Wilks' lambda test resulted in 10 significantly different exposure metrics based on 10 mph over-speed activities between the two driver-groups (Table 36).

Table 36: Tests of Equality of Means of Frequencies of 10 mph Over-Speed Activities between the Groups Based on the Wilks' Lambda Test

Facility Type	Trip Time	Wilks' Lambda	F	Sig.
Freeways	AM peak	0.998	0.230	0.632
	Morning *	**0.962**	**6.388**	**0.012**
	Afternoon *	**0.963**	**6.305**	**0.013**
	PM peak	0.980	3.296	0.071
	Night	1.000	0.041	0.840
	Early morning	0.964	3.139	0.080
Arterials	AM peak	0.976	3.920	0.050
	Morning *	**0.901**	**18.072**	**0.000**
	Afternoon *	**0.956**	**7.667**	**0.006**
	PM peak *	**0.948**	**9.018**	**0.003**
	Night *	**0.942**	**9.804**	**0.002**
	Early morning	0.984	1.829	0.179
Local Roads	AM peak	0.980	3.289	0.072
	Morning *	**0.927**	**13.017**	**0.000**
	Afternoon *	**0.959**	**6.991**	**0.009**
	PM peak *	**0.945**	**9.552**	**0.002**
	Night *	**0.965**	**5.952**	**0.016**
	Early morning	0.986	2.007	0.159

* indicates a significant mean difference ($\alpha = 0.05$).

Table 37 shows that drivers with crashes traveled an average of 17 times of over-speeding operation every mile (215 seconds every specific trip regarding facility and time) and driver without crashes traveled an average of 12.5 times of over-speeding operation every mile (134 seconds every trip) during morning on freeways. This result indicates that the over-speeding activity of drivers with crashes was 26 % higher than that of drivers without crash involvements. Over-speeding activities on freeways during afternoon also showed the significant difference (24 %) between the two driver-groups.

For the arterial activities, drivers involved in crashes traveled an average of 7 times of over-speeding operation every mile during morning, and drivers without crashes traveled an average of 4 times of over-speeding operation every mile, the difference of 46 %. For the local roadways, morning and PM peak showed the significant differences in over-speeding activities between the two driver-groups. Drivers involved in crashes traveled an average of 8.9 times of over-speeding operation every mile during morning, and drivers without crashes traveled an average of 5.4 times of over-speeding operation every mile, the difference of 39 %.

Table 37: Frequencies of 10 mph Over-Speeds per Mile based on Facility and Time

Facility	Trip Time	Average of 10 mph Over-speed Activity per Mile		Frequency Difference	% Difference
		Drivers who *were not involved* in Crashes	Drivers who *were involved* in Crashes		
Freeways	AM peak	9.5	10.5	1.0	9
	Morning *	**12.5**	**17.0**	**4.5**	**26**
	Afternoon *	**11.7**	**15.4**	**3.8**	**24**
	PM peak	8.9	11.9	2.9	25
	Night	12.2	12.6	0.4	3
	Early morning	9.8	16.3	6.4	39
Arterials	AM peak	3.6	5.7	2.1	37
	Morning *	**3.5**	**6.5**	**3.0**	**46**
	Afternoon *	**3.3**	**5.1**	**1.9**	**37**
	PM peak *	**2.6**	**4.4**	**1.8**	**40**
	Night *	**3.0**	**5.6**	**2.6**	**47**
	Early morning	5.6	9.3	3.7	40
Local Roads	AM peak	5.7	8.1	2.5	30
	Morning *	**5.4**	**8.9**	**3.4**	**39**
	Afternoon *	**5.4**	**7.6**	**2.2**	**29**
	PM peak *	**4.5**	**7.1**	**2.6**	**37**
	Night *	**4.9**	**7.6**	**2.8**	**36**
	Early morning	5.1	7.9	2.8	36

* indicates a significant mean difference ($\alpha = 0.05$).

Frequency of 15 and 20 mph Over-Speed Activities

In addition to the over-speed activities of 10 mph over than the posted speed limit, this study further investigated the speeding behavior with larger thresholds such as 15 mph and 20 mph. Based on the non-parametric bootstrap technique, speeding activities (15 mph over-speeding) on all facilities during morning were significantly different (Figure 53) and speeding activities (20 mph over-speeding) on only arterials during morning were significantly different between the two driver-groups (Figure 54).

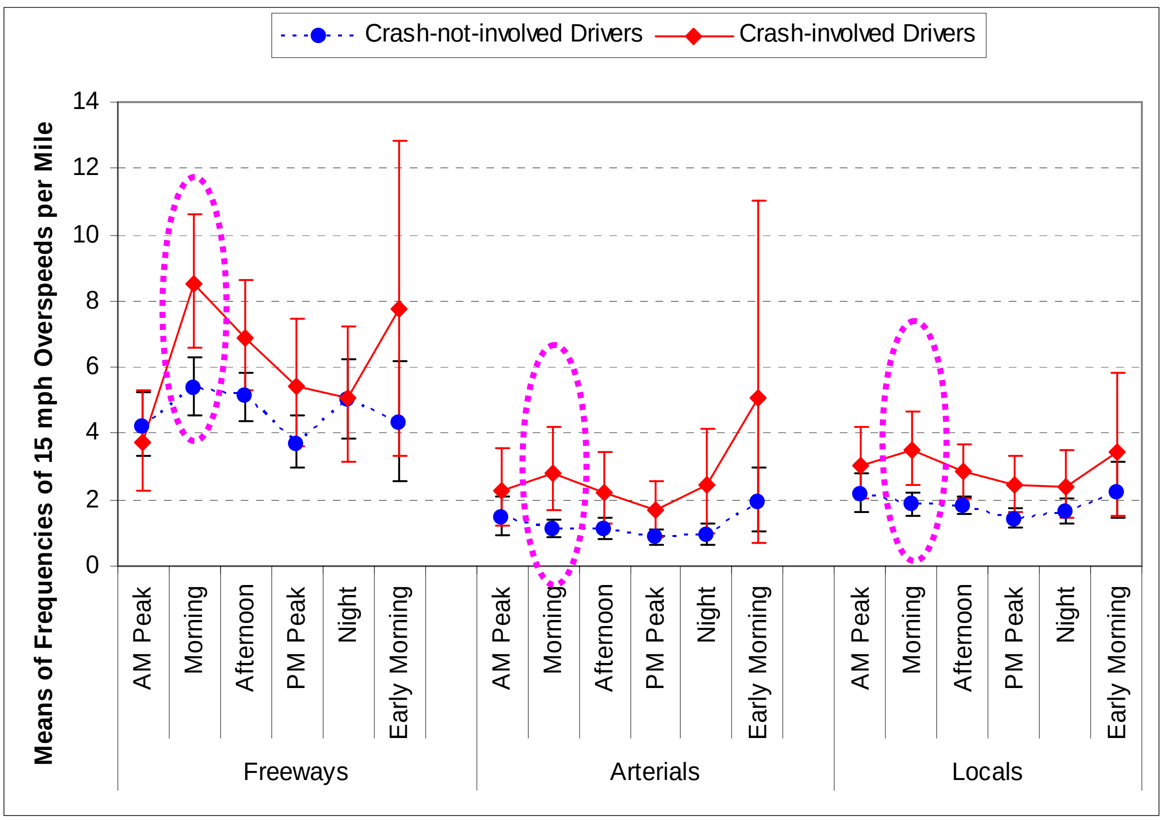


Figure 53: Average Frequencies of 15 mph Over-Speed Activities per Mile

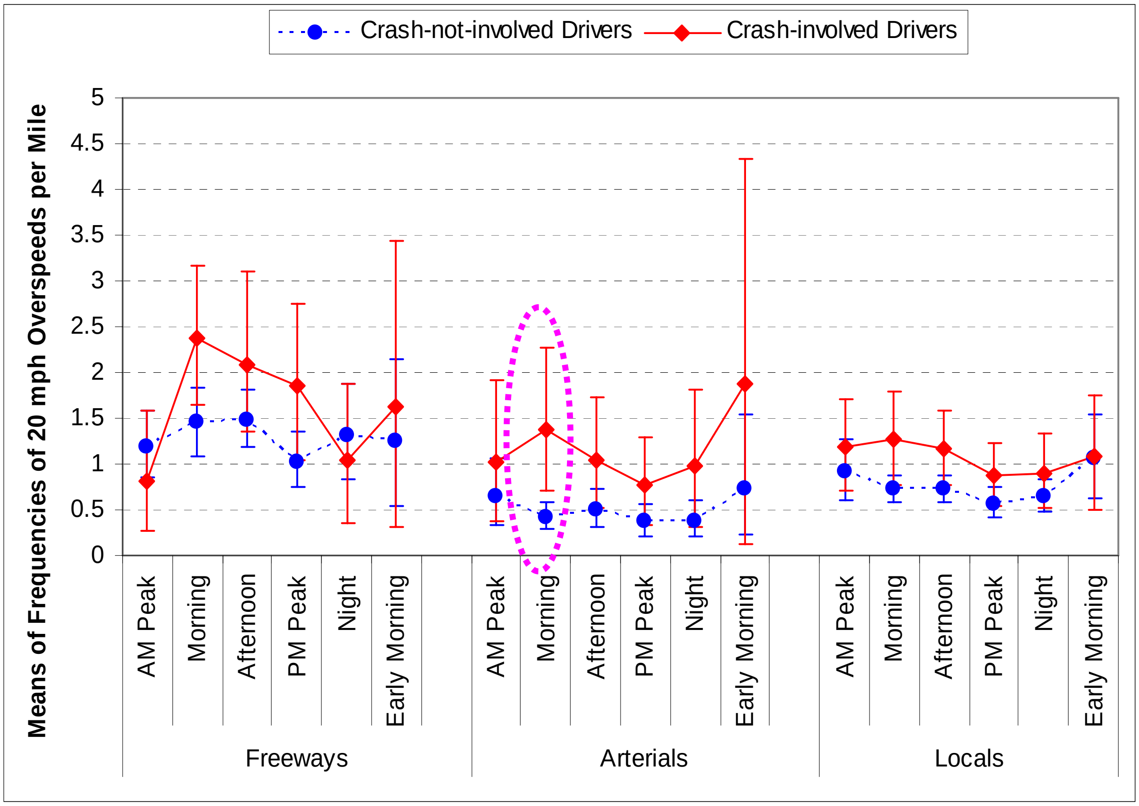


Figure 54: Average Frequencies of 20 mph Over-Speed Activities per Mile

The Walks' lambda test (Table 38) showed numerous significant over-speed activity exposures (15 mph and 20 mph over-speed activities) including those selected from the bootstrap technique.

Table 38: Tests of Equality of Means of Average Frequencies of 15 and 20 mph Over-Speeding per Mile based on the Wilks' Lambda Test

Facility	Time	Means of Frequencies of 15 mph Over-Speeds			Means of Frequencies of 20 mph Over-Speeds		
		Wilks' Lambda	F	Sig.	Wilks' Lambda	F	Sig.
Freeway	AM peak	0.999	0.168	0.682	0.995	0.710	0.401
	Morning	**0.949**	**8.613**	**0.004**	0.978	3.696	0.056
	Afternoon	0.981	3.123	0.079	0.987	2.156	0.144
	PM peak	0.982	2.926	0.089	0.975	4.026	0.050
	Night	1.000	0.003	0.956	0.999	0.168	0.682
	Early morning	0.976	2.012	0.160	0.998	0.150	0.700
Arterial	AM peak	0.992	1.268	0.262	0.996	0.655	0.419
	Morning *	**0.912**	**16.018**	**0.000**	**0.917**	**14.946**	**0.000**
	Afternoon *	**0.965**	**6.028**	**0.015**	0.980	3.390	0.067
	PM peak	**0.961**	**6.677**	**0.011**	0.980	3.440	0.065
	Night *	**0.946**	**9.053**	**0.003**	**0.970**	**4.938**	**0.028**
	Early morning	0.969	3.602	0.060	0.990	1.141	0.288
Local	AM peak	0.992	1.294	0.257	0.997	0.453	0.502
	Morning *	**0.929**	**12.624**	**0.000**	**0.961**	**6.743**	**0.010**
	Afternoon	**0.958**	**7.185**	**0.008**	**0.971**	**4.946**	**0.028**
	PM peak	**0.962**	**6.570**	**0.011**	0.985	2.586	0.110
	Night	0.987	2.154	0.144	0.993	1.171	0.281
	Early morning	0.992	1.241	0.267	1.000	0.000	0.986

* indicates a significant mean difference ($\alpha = 0.05$).

The Linear Discriminant Analysis Using Speed-related Behavior Activity Exposures

Based on the nonparametric bootstrap method and parametric Wilks' lambda test, this study found that 40 speed-related exposure metrics were significantly different between drivers who were involved and were not involved in crashes and suggested that those metrics could be used to verify the drivers who potentially have crash involvement risk. However, for the modeling process, issues on the correlations (threshold: 0.6) should be remedied as this study did for the mileage and duration exposure metrics. Table 39 illustrates the result of the correlation analysis with the 40 speed-related exposure metrics.

Using the structure coefficients in Table 40, this study selected metrics having higher structure coefficients (loading power) among the correlated variables. Based on the structure coefficients, this study finally obtained 11 speed-related metrics that can be potentially used for classifying drivers into two risk groups.

Table 39: Correlation Analysis with Speed-related Exposure Metrics

Measure	Facility/Time	Mean Speed			Average Running Speed			Delta Speed	
		Frwy/Morning	Frwy/Night	Artrl/AM	Frwy/Morning	Frwy/Night	Artrl/AM	Frwy/Morning	Artrl/AM
Mean Speed	Frwy/Morning	1.0	-	-	-	-	-	-	-
Mean Speed	Frwy/Night	0.5	1.0	-	-	-	-	-	-
Mean Speed	Artrl/AM	0.2	0.1	1.0	-	-	-	-	-
Average Running Speed	Frwy/Morning	**1.0**	0.5	0.2	1.0	-	-	-	-
Average Running Speed	Frwy/Night	0.5	**1.0**	0.1	0.5	1.0	-	-	-
Average Running Speed	Artrl/AM	0.2	0.1	**0.9**	0.2	0.1	1.0	-	-
Delta Speed	Frwy/Morning	**0.9**	0.4	0.2	**0.9**	0.5	0.2	1.0	-
Delta Speed	Artrl/AM	0.1	0.1	**0.8**	0.1	0.1	**0.7**	0.1	1.0
Positive Delta Speed	Frwy/Morning	0.4	0.2	-0.1	0.4	0.3	0.0	**0.6**	-0.1
Positive Delta Speed	Frwy/PM	0.4	0.3	0.0	0.4	0.3	0.0	0.4	0.0
Positive Delta Speed	Artrl/AM	0.0	0.0	0.2	0.0	0.0	0.2	0.1	0.3
Positive Delta Speed	Artrl/Morning	0.3	0.2	0.1	0.3	0.2	0.2	0.4	0.2
Positive Delta Speed	Artrl/Afternoon	0.3	0.2	0.2	0.3	0.2	0.3	0.3	0.3
Positive Delta Speed	Artrl/PM	0.3	0.2	0.2	0.3	0.2	0.2	0.4	0.3
Positive Delta Speed	Local/AM	0.1	0.1	0.1	0.2	0.1	0.2	0.2	0.1
Positive Delta Speed	Local/Afternoon	0.4	0.1	0.0	0.4	0.2	0.1	0.4	0.0
Positive Delta Speed	Local/PM	0.2	0.1	0.1	0.3	0.2	0.1	0.3	0.0
Positive Delta Speed	Local/Night	0.2	0.2	0.1	0.3	0.3	0.1	0.3	0.1
Frequency of 10 mph over-speed per mile	Frwy/Morning	0.4	0.2	0.0	0.4	0.2	0.1	0.5	0.0
Frequency of 10 mph over-speed per mile	Frwy/Afternoon	0.3	0.2	0.0	0.3	0.2	0.1	0.4	0.1
Frequency of 10 mph over-speed per mile	Artrl/Morning	0.3	0.2	0.2	0.3	0.2	0.2	0.4	0.3
Frequency of 10 mph over-speed per mile	Artrl/Afternoon	0.3	0.2	0.3	0.3	0.2	0.3	0.3	0.4
Frequency of 10 mph over-speed per mile	Artrl/PM	0.3	0.2	0.2	0.3	0.2	0.3	0.3	0.3
Frequency of 10 mph over-speed per mile	Artrl/Night	0.2	0.1	0.1	0.2	0.1	0.1	0.2	0.2
Frequency of 10 mph over-speed per mile	Local/Morning	0.3	0.1	0.1	0.3	0.2	0.2	0.3	0.1
Frequency of 10 mph over-speed per mile	Local/Afternoon	0.3	0.1	0.2	0.3	0.2	0.3	0.3	0.1
Frequency of 10 mph over-speed per mile	Local/PM	0.2	0.1	0.2	0.3	0.2	0.2	0.2	0.1
Frequency of 10 mph over-speed per mile	Local/Night	0.3	0.3	0.2	0.3	0.3	0.1	0.3	0.1
Frequency of 15 mph over-speed per mile	Frwy/Morning	0.4	0.2	0.0	0.4	0.3	0.1	0.5	0.0
Frequency of 15 mph over-speed per mile	Artrl/Morning	0.2	0.2	0.3	0.2	0.2	0.3	0.3	0.4
Frequency of 15 mph over-speed per mile	Artrl/Afternoon	0.2	0.1	0.3	0.2	0.1	0.3	0.2	0.4
Frequency of 15 mph over-speed per mile	Artrl/PM	0.2	0.1	0.3	0.2	0.1	0.3	0.2	0.4
Frequency of 15 mph over-speed per mile	Artrl/Night	0.1	0.1	0.1	0.1	0.1	0.0	0.1	0.1
Frequency of 15 mph over-speed per mile	Local/Morning	0.3	0.1	0.1	0.3	0.1	0.1	0.3	0.0
Frequency of 15 mph over-speed per mile	Local/Afternoon	0.3	0.1	0.1	0.3	0.1	0.2	0.3	0.1
Frequency of 15 mph over-speed per mile	Local/PM	0.2	0.1	0.1	0.2	0.1	0.1	0.2	0.1
Frequency of 20 mph over-speed per mile	Artrl/Morning	0.1	0.1	0.3	0.1	0.1	0.2	0.2	0.4
Frequency of 20 mph over-speed per mile	Artrl/Night	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2
Frequency of 20 mph over-speed per mile	Local/Morning	0.2	0.1	0.0	0.2	0.1	0.0	0.3	0.0
Frequency of 20 mph over-speed per mile	Local/Afternoon	0.2	0.1	0.0	0.2	0.1	0.0	0.3	0.0

Cont'd

Measure	Facility/Time	Positive Delta Speed									
		Frwy/Morning	Frwy/PM	Artrl/AM	Artrl/Morning	Artrl/Afternoon	Artrl/PM	Local/AM	Local/Afternoon	Local/PM	Local/Night
Positive Delta Speed	Frwy/Morning	1.0	-	-	-	-	-	-	-	-	-
	Frwy/PM	**0.6**	1.0	-	-	-	-	-	-	-	-
	Artrl/AM	0.1	0.1	1.0	-	-	-	-	-	-	-
	Artrl/Morning	0.4	0.3	0.5	1.0	-	-	-	-	-	-
	Artrl/Afternoon	0.4	0.3	0.5	0.8	1.0	-	-	-	-	-
	Artrl/PM	0.3	0.3	**0.6**	**0.8**	**0.8**	1.0	-	-	-	-
	Local/AM	0.4	0.3	0.4	0.4	0.4	0.4	1.0	-	-	-
	Local/Afternoon	0.5	0.4	0.2	0.5	0.4	0.4	**0.6**	1.0	-	-
	Local/PM	0.4	0.4	0.2	0.5	0.4	0.5	**0.6**	**0.8**	1.0	-
	Local/Night	0.4	0.3	0.2	0.4	0.4	0.4	0.5	**0.7**	**0.7**	1.0
Frequency of 10 mph over-speed per mile	Frwy/Morning	**0.7**	0.5	0.1	0.4	0.4	0.4	0.3	0.4	0.4	0.3
	Frwy/Afternoon	**0.6**	0.5	0.1	0.4	0.4	0.4	0.3	0.3	0.3	0.2
	Artrl/Morning	0.4	0.3	0.4	**0.8**	**0.7**	**0.7**	0.3	0.4	0.4	0.4
	Artrl/Afternoon	0.3	0.3	0.5	**0.7**	**0.8**	**0.7**	0.3	0.3	0.3	0.3
	Artrl/PM	0.3	0.3	0.6	**0.7**	**0.8**	**0.9**	0.3	0.3	0.4	0.3
	Artrl/Night	0.2	0.2	0.5	**0.6**	**0.7**	**0.6**	0.3	0.3	0.3	0.4
	Local/Morning	0.5	0.4	0.1	0.5	0.4	0.5	0.4	**0.7**	**0.7**	0.5
	Local/Afternoon	0.4	0.4	0.2	0.4	0.4	0.5	0.5	**0.8**	**0.8**	0.5
	Local/PM	0.4	0.4	0.3	0.5	0.4	0.5	0.5	**0.7**	**0.9**	**0.7**
	Local/Night	0.4	0.3	0.1	0.4	0.3	0.4	0.4	**0.6**	**0.7**	**0.8**
Frequency of 15 mph over-speed per mile	Frwy/Morning	**0.8**	**0.6**	0.1	0.4	0.4	0.4	0.3	0.5	0.4	0.4
	Artrl/Morning	0.3	0.2	0.5	**0.8**	**0.7**	**0.7**	0.3	0.3	0.3	0.3
	Artrl/Afternoon	0.1	0.1	**0.6**	**0.6**	**0.8**	**0.7**	0.2	0.2	0.2	0.2
	Artrl/PM	0.1	0.1	0.5	**0.6**	**0.7**	**0.8**	0.2	0.2	0.2	0.2
	Artrl/Night	0.1	0.1	0.5	0.5	**0.6**	0.5	0.2	0.1	0.2	0.2
	Local/Morning	0.5	0.4	0.2	0.5	0.4	0.4	0.5	**0.7**	**0.7**	0.5
	Local/Afternoon	0.5	0.5	0.2	0.4	0.4	0.4	**0.6**	**0.8**	**0.8**	0.5
	Local/PM	0.4	0.4	0.2	0.4	0.4	0.5	**0.6**	**0.7**	**0.8**	**0.6**
Frequency of 20 mph over-speed per mile	Artrl/Morning	0.1	0.1	0.5	**0.7**	**0.7**	**0.6**	0.2	0.1	0.2	0.2
	Artrl/Night	0.1	0.1	0.5	0.4	**0.6**	0.5	0.2	0.1	0.2	0.2
	Local/Morning	0.5	0.3	0.0	0.3	0.3	0.3	0.4	**0.6**	0.5	0.4
	Local/Afternoon	0.5	0.5	0.1	0.3	0.3	0.3	**0.6**	**0.7**	**0.6**	0.5

Cont'd

Measure	Facility/Time	Frequency of 10 mph over-speed per mile								
		Frwy/Morning	Frwy/Afternoon	Artrl/Morning	Artrl/Afternoon	Artrl/PM	Artrl/Night	Local/Morning	Local/Afternoon	Local/PM
Frequency of 10 mph over-speed per mile	Frwy/Morning	1.0	-	-	-	-	-	-	-	-
	Frwy/Afternoon	**0.6**	1.0	-	-	-	-	-	-	-
	Artrl/Morning	0.4	0.4	1.0	-	-	-	-	-	-
	Artrl/Afternoon	0.4	0.4	**0.8**	1.0	-	-	-	-	-
	Artrl/PM	0.4	0.4	**0.8**	**0.9**	1.0	-	-	-	-
	Artrl/Night	0.3	0.3	**0.7**	**0.7**	**0.7**	1.0	-	-	-
	Local/Morning	0.5	0.3	0.5	0.4	0.4	0.3	1.0	-	-
	Local/Afternoon	0.4	0.3	0.4	0.4	0.4	0.3	**0.8**	1.0	-
	Local/PM	0.4	0.3	0.4	0.4	0.4	0.3	**0.7**	**0.8**	1.0
	Local/Night	0.3	0.2	0.4	0.3	0.3	0.3	**0.6**	**0.7**	**0.8**
Frequency of 15 mph over-speed per mile	Frwy/Morning	**0.9**	**0.6**	0.4	0.3	0.4	0.3	0.5	0.4	0.4
	Artrl/Morning	0.3	0.3	**0.9**	**0.8**	**0.8**	**0.6**	0.4	0.3	0.3
	Artrl/Afternoon	0.2	0.3	**0.6**	**0.9**	**0.8**	**0.7**	0.3	0.2	0.2
	Artrl/PM	0.2	0.2	**0.6**	**0.8**	**0.9**	**0.6**	0.3	0.3	0.3
	Artrl/Night	0.1	0.2	0.5	**0.6**	**0.6**	**0.9**	0.1	0.1	0.2
	Local/Morning	0.5	0.4	0.5	0.4	0.4	0.3	**0.9**	**0.7**	**0.7**
	Local/Afternoon	0.4	0.4	0.4	0.4	0.4	0.3	**0.8**	**0.9**	**0.8**
	Local/PM	0.4	0.4	0.4	0.4	0.4	0.3	**0.6**	**0.7**	**0.9**
Frequency of 20 mph over-speed per mile	Artrl/Morning	0.2	0.2	**0.8**	**0.7**	**0.7**	0.5	0.3	0.2	0.2
	Artrl/Night	0.2	0.2	0.4	0.5	0.5	**0.7**	0.1	0.0	0.1
	Local/Morning	0.5	0.4	0.4	0.2	0.2	0.1	**0.7**	0.5	0.5
	Local/Afternoon	0.4	0.5	0.3	0.3	0.3	0.2	**0.6**	**0.6**	**0.6**

Cont'd

Measure	Facility/Time	Frequency of 15 mph over-speed per mile							
		Frwy/Morning	Artrl/Morning	Artrl/Afternoon	Artrl/PM	Artrl/Night	Local/Morning	Local/Afternoon	Local/PM
Frequency of 15 mph over-speed per mile	Frwy/Morning	1.0	-	-	-	-	-	-	-
	Artrl/Morning	0.3	1.0	-	-	-	-	-	-
	Artrl/Afternoon	0.2	0.8	1.0	-	-	-	-	-
	Artrl/PM	0.2	0.8	0.9	1.0	-	-	-	-
	Artrl/Night	0.1	0.6	0.6	0.5	1.0	-	-	-
	Local/Morning	0.5	0.4	0.2	0.2	0.1	1.0	-	-
	Local/Afternoon	0.5	0.3	0.2	0.2	0.1	0.8	1.0	-
	Local/PM	0.5	0.3	0.2	0.3	0.2	0.7	0.8	1.0
Frequency of 20 mph over-speed per mile	Artrl/Morning	0.1	0.9	0.8	0.8	0.5	0.3	0.2	0.2
	Artrl/Night	0.1	0.5	0.6	0.5	0.9	0.1	0.1	0.2
	Local/Morning	0.5	0.3	0.1	0.1	0.0	0.9	0.7	0.5
	Local/Afternoon	0.4	0.2	0.2	0.1	0.1	0.7	0.9	0.7

Cont'd

Measure	Facility/Time	Frequency of 20 mph over-speed per mile			
		Artrl/Morning	Artrl/Night	Local/Morning	Local/Afternoon
Frequency of 20 mph over-speed per mile	Artrl/Morning	1.0	-	-	-
	Artrl/Night	0.5	1.0	-	-
	Local/Morning	0.2	0.0	1.0	-
	Local/Afternoon	0.1	0.1	0.7	1.0

Table 40: Structure Coefficients from the Linear Discriminant Analysis Using Speed-related Activity Exposure Metrics

Metrics	Facility/Time	Structure Coefficients	Rank	Variable Selection
Frequency of 15 mph over-speed per mile	**Arterial/Morning**	**0.451**	**1**	**√**
Frequency of 10 mph over-speed per mile	Arterial/Morning	0.439	2	-
Frequency of 20 mph over-speed per mile	Arterial/Morning	0.423	3	-
Frequency of 10 mph over-speed per mile	Arterial/Night	0.414	4	-
Delta Speed	**Arterial/AM Peak**	**0.375**	**5**	**√**
Frequency of 15 mph over-speed per mile	Arterial/Night	0.374	6	-
Positive Delta Speed	Arterial/Morning	0.348	7	-
Frequency of 10 mph over-speed per mile	Arterial/PM Peak	0.318	8	-
Frequency of 10 mph over-speed per mile	**Freeway/Morning**	**0.309**	**9**	**√**
Frequency of 10 mph over-speed per mile	Arterial/Afternoon	0.308	10	-
Frequency of 15 mph over-speed per mile	Arterial/Afternoon	0.303	11	-
Frequency of 10 mph over-speed per mile	**Local/Morning**	**0.298**	**12**	**√**
Positive Delta Speed	Arterial/PM Peak	0.294	13	-
Frequency of 15 mph over-speed per mile	Local/Morning	0.285	14	-
Positive Delta Speed	Arterial/Afternoon	0.263	15	-
Frequency of 20 mph over-speed per mile	Local/Morning	0.261	16	-
Positive Delta Speed	**Arterial/AM Peak**	**0.261**	**17**	**√**
Positive Delta Speed	**Freeway/PM Peak**	**0.257**	**18**	**√**
Frequency of 20 mph over-speed per mile	**Arterial/Night**	**0.251**	**19**	**√**

Cont'd

Frequency of 15 mph over-speed per mile	Arterial/PM Peak	0.250	20	-
Mean Speed	**Freeway/Night**	**0.247**	**21**	√
Frequency of 10 mph over-speed per mile	Local/PM Peak	0.246	22	-
Frequency of 15 mph over-speed per mile	Freeway/Morning	0.246	23	-
Average Running Speed	Freeway/Night	0.245	24	-
Positive Delta Speed	Freeway/Morning	0.242	25	-
Average Running Speed	Arterial/AM Peak	0.240	26	-
Mean Speed	Arterial/AM Peak	0.238	27	-
Frequency of 15 mph over-speed per mile	Local/PM Peak	0.206	28	-
Positive Delta Speed	Local/PM Peak	0.185	29	-
Mean Speed	**Freeway/Morning**	**0.173**	**30**	√
Delta Speed	Freeway/Morning	0.169	31	-
Average Running Speed	Freeway/Morning	0.164	32	-
Frequency of 20 mph over-speed per mile	Local/Afternoon	0.157	33	-
Frequency of 15 mph over-speed per mile	Local/Afternoon	0.148	34	-
Frequency of 10 mph over-speed per mile	Local/Afternoon	0.148	35	-
Positive Delta Speed	**Local/AM Peak**	**0.136**	**36**	√
Frequency of 10 mph over-speed per mile	Local/Night	0.133	37	-
Positive Delta Speed	**Local/Night**	**0.127**	**38**	√
Frequency of 10 mph over-speed per mile	Freeway/Afternoon	0.124	39	-
Positive Delta Speed	Local/Afternoon	0.049	40	-

Table 41 shows the accuracy of the discriminant analysis using the selected 11 potential speed-related crash exposure metrics. As a result, 71.8 % of drivers who were not involved in crashes and 63.6 % of drivers who were involved in crashes were correctly classified. Overall performance of the model using only speed-related exposure metrics was 70.4 %. Similar to the previous analyses in this study, 28.2 % of drivers who were not involved in crashes were classified as a potential crash involvement group based on their speed-related exposures. For the 36.4 % of drivers who were involved in crashes, they were theoretically classified into the low crash-involvement group based on their observed speed exposures. This result indicates that the model has a limitation for classifying drivers who have potentially high crash involvement rate based on speed exposures only or more likely, interactions between other behavior activities such as acceleration patterns need to be investigated.

Table 41: Classification Results Using Speed-related Exposure Metrics

Crash Involvements	Predicted Group Membership	
	Drivers who *were not involved* in crashes	Drivers who *were involved* in crashes
Drivers who *were not involved* in crashes	71.8 %	28.2 %
Drivers who *were involved* in crashes	36.4 %	63.6 %

Summary of the Speed-related Exposure Metrics

This study evaluated differences in speed-related exposure metrics of drivers with and without crash involvements during the 14-months period to verify whether their speed behaviors were significantly different based on facility and time of day. As a result, this study found that drivers who were involved in crashes tended to travel at higher speeds every day than drivers who were not involved in crashes. The findings are summarized as follows:

1. This study indicates that drivers who had experienced crashes appear to be more willing to travel at high speed than drivers who had not crash involvements based on both “amount” and “frequency” measures.
2. This study also indicates that speed has a positive relationship with crash involvement rate, so this result supports the conventional theory, “higher speed is equaled to higher opportunity of crash involvements”.
3. Speeding activities of drivers with crash involvements were larger at most times than drivers who without crash involvements. Morning and afternoon provided the largest differences in speeding behaviors between the two driver-groups.
4. This study suggests that speed-related exposure metrics can be also used as one of potential behavioral crash risk measures for identifying potentially high crash risk drivers.
5. Safety engineers and policy makers need to continue to aim anti-speeding campaigns to drivers, and these behavioral metrics for individual drivers can be incorporated into education campaigns and driver evaluation or monitoring programs.
6. Insurance companies may enhance current insurance classification decision rules with adding the more detailed speed exposure metrics discussed in this study.

Classification with Travel Mileage, Duration, and Speed-related Exposure Metrics

Through the previous and current chapters, this study evaluated individual relationships between crash involvement rate and potential behavioral exposure measures such as travel mileage, travel duration, and speed patterns and provided the prediction powers of each driving behavior activity exposure metric for classifying drivers into the two different crash involvement groups using the linear discriminant analysis (LDA).

However, it may not clearly predict potential crash involvements based on individual behavioral exposure metric only. For example, when using the travel mileage exposure only, some of drivers who involved in crashes may not be correctly classified if they traveled less mileage. However, if they have high over-speed activities, they can be correctly classified based on speed-related behavior exposure metrics. Thus, this study evaluates the differences of prediction performance by combining those behavioral exposure metrics[18].

This study first evaluated the performance of classifications using selected travel mileage and travel duration exposure metrics. Based on the structure coefficients and correlation analysis, this study selected eight behavior activity exposure metrics (Table 42) and performed the linear discriminant analysis. The final classification result is shown in Table 43.

---

[18] A detailed modeling process with larger independent variables is performed in Chapter 11.

Table 42: Structure Coefficient Analysis Using Travel Mileage and Duration Exposure Metrics

Metrics	Facility/Time	Structure Coefficients	Rank	Variable Selection
Travel Mileage	**Outside/Afternoon**	**0.580**	**1**	√
Travel Duration	**Total**	**0.510**	**2**	√
Travel Duration	**Outside/Afternoon**	0.490	3	-
Travel Duration	**Freeway/PM Peak**	**0.449**	**4**	√
Travel Distance	Total	0.443	5	-
Travel Distance	Freeway/PM Peak	0.441	6	-
Travel Distance	**Freeway/Night**	**0.425**	**7**	√
Travel Duration	**Outside/AM Peak**	**0.416**	**8**	√
Travel Duration	Freeway/Night	0.413	9	-
Travel Distance	Outside/AM Peak	0.388	10	-
Travel Distance	**Arterial/PM Peak**	**0.349**	**11**	√
Travel Duration	Arterial/PM Peak	0.323	12	-
Travel Duration	**Outside/Night**	**0.309**	**13**	√
Travel Duration	**Arterial/AM Peak**	**0.265**	**14**	√

As a result, 75.2 % of drivers who were not involved in crashes and 61.5 % of drivers who were involved in crashes were correctly classified (Table 43). Overall performance of the model using mileage and duration exposure metrics was 73.1 %. The result shows that the performance for classifying crash-involved drivers using mileage and duration metrics is higher than using the mileage exposure only (57.7 %) but lower than using the travel duration exposure only (65.4 %) or speed exposure only (63.6 %).

Table 43: Classification Results Using Mileage and Duration Exposure Metrics

Crash Involvements	Predicted Group Membership	
	Drivers who *were not involved* in crashes	Drivers who *were involved* in crashes
Drivers who *were not involved* in crashes	75.2 %	24.8 %
Drivers who *were involved* in crashes	38.5 %	61.5 %

This study also evaluated the performance of classifications using selected travel mileage, travel duration, and speed exposure metrics. Based on the structure coefficients and correlation analysis, this study selected 19 driving behavior activity exposure metrics (Table 44) and performed the linear discriminant analysis. The final classification result is shown in Table 45.

Table 44: Structure Coefficient Analysis Using Travel Mileage, Duration, and Speed Exposure Metrics

Metrics	Facility/Time	Structure Coefficients	Rank	Variable Selection
Travel Duration	Total	0.502	1	√
Frequency of 15 mph over-speed per mile	Arterial/Morning	0.473	2	√
Travel Mileage	Outside/Afternoon	0.415	3	√
Travel Duration	Outside/AM Peak	0.398	4	√
Delta Speed	Arterial/AM Peak	0.388	5	√
Frequency of 10 mph over-speed per mile	Freeway/Morning	0.329	6	√
Frequency of 10 mph over-speed per mile	Local/Morning	0.315	7	√
Travel Duration	Freeway/PM Peak	0.300	8	√
Travel Mileage	Arterial/PM Peak	0.297	9	√
Positive Delta Speed	Arterial/AM Peak	0.274	10	√
Positive Delta Speed	Freeway/PM Peak	0.273	11	√
Travel Duration	Arterial/AM Peak	0.269	12	√
Mean Speed	Freeway/Night	0.266	13	√
Frequency of 20 mph over-speed per mile	Arterial/Night	0.264	14	√
Travel Duration	Outside/Night	0.214	15	√
Mean Speed	Freeway/Morning	0.193	16	√
Positive Delta Speed	Local/AM Peak	0.150	17	√
Travel Duration	Freeway/Night	0.142	18	√
Positive Delta Speed	Local/Night	0.139	19	√

81.4 % of drivers who were not involved in crashes and 69.6 % of drivers who were involved in crashes were correctly classified (Table 45). Overall performance of the model using mileage and duration exposure metrics was 79.4 %. The result shows that the performance for classifying crash-involved drivers using mileage, duration, and speed exposure metrics is higher than using the mileage exposure only (57.7 %), using the travel duration exposure only (65.4 %), or speed exposure only (63.6 %) (Table 46). This result indicates that interactions exist between driving behavior activity exposure metrics and that complementary approach using various behavioral exposures need to be applied to classify potential high crash-involvement drivers.

Table 45: Classification Results Using Mileage, Duration, and Speed Exposure Metrics

Crash Involvements	Predicted Group Membership	
	Drivers who *were not involved* in crashes	Drivers who *were involved* in crashes
Drivers who *were not involved* in crashes	81.4 %	18.6 %
Drivers who *were involved* in crashes	30.4 %	69.6 %

Table 46: Differences in Performance of Classification Using Mileage, Duration, and Speed Exposure Metrics

Behavioral Exposure Measure	Overall Performance (%)	Performance of Predictions on Drivers who *were involved* in crashes (%)
Travel Mileage	77.2	57.7
Travel Duration	68.7	65.4
Speed	70.4	63.6
Mileage and Duration	73.1	61.5
Mileage, Duration, and Speed	**79.4**	**69.6**

Chapter Eight

POTENTIAL BEHAVIORAL EXPOSURE IV: Acceleration

In addition to the travel mileage, duration, and speed-related potential exposure measures in the previous chapters, this study examines differences in acceleration behaviors between the two driver-groups who were involved and were not involved in crashes. To estimate acceleration profiles from the GPS-observed second-by-second speeds, this study used the central difference method [18]. To evaluate potential driving behavior activity exposures using acceleration patterns in detail, this study selects 11 acceleration-related metrics as follows:

- Mean of Accelerations
- Mean of Absolute Acceleration Values
- Acceleration Noise
- Frequency of Hard Accelerations per Mile (>= 4 mph/s, 6 mph/s, 8 mph/s, and 10 mph/s)
- Frequency of Hard Decelerations per Mile (>= 4 mph/s, 6 mph/s, 8 mph/s, and 10 mph/s)

Differences in the Mean of Accelerations, Mean of Absolute Accelerations, and Acceleration Noise

Figure 55 and 56 shows the means of average accelerations based on facility types and the corresponding time of day between the two driver-groups and indicates that they did not have any significantly different acceleration patterns using the bootstrap method.

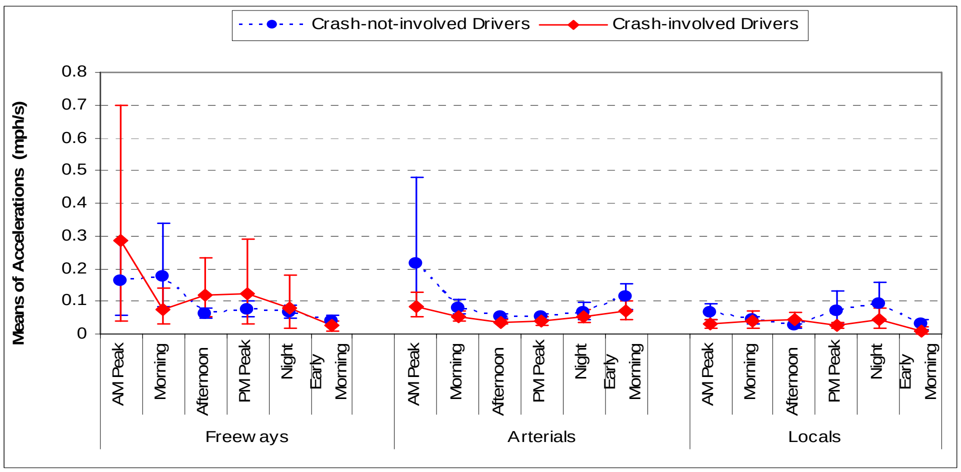


Figure 55: Confidence Intervals of the Means of Accelerations Using the Bootstrap Technique

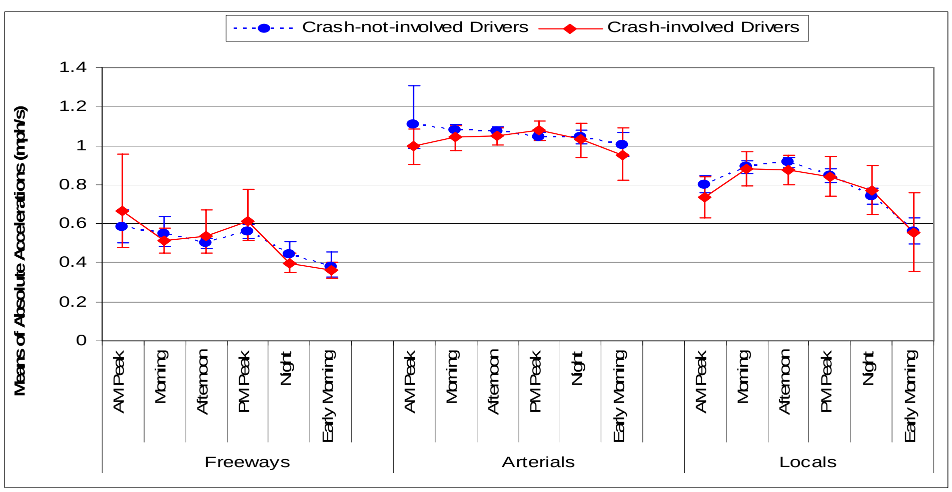


Figure 56: Confidence Intervals of the Means of Absolute Accelerations Using the Bootstrap Technique

Unlike to the bootstrap technique, the Wilks' lambda test showed that the average accelerations on freeways during afternoon were significantly different at the 0.05 significance level (Table 47).

Table 47: Tests of Equality of Means and Means of Absolute Accelerations between Two Groups Based on the Wilks' Lambda Test

		Means of Accelerations			Means of Absolute Accelerations		
Facility Type	Trip Time	Wilks' Lambda	F	Sig.	Wilks' Lambda	F	Sig.
Freeways	am peak	1.000	0.031	0.861	1.000	0.030	0.864
	morning	0.997	0.174	0.678	1.000	0.016	0.900
	Afternoon *	**0.837**	**12.307**	**0.001 ***	0.944	3.717	0.058
	pm peak	1.000	0.000	0.986	1.000	0.016	0.899
	night	1.000	0.031	0.860	0.957	2.841	0.097
	early morning	0.999	0.033	0.856	1.000	0.022	0.882
Arterials	am peak	0.997	0.195	0.660	0.998	0.099	0.754
	morning	0.980	1.277	0.263	0.971	1.857	0.178
	afternoon	0.989	0.710	0.403	0.991	0.565	0.455
	pm peak	1.000	0.002	0.966	1.000	0.001	0.970
	night	0.995	0.306	0.582	1.000	0.023	0.880
	early morning	0.977	1.490	0.227	0.990	0.640	0.427
Local Roads	am peak	0.997	0.215	0.645	1.000	0.003	0.960
	morning	0.996	0.271	0.604	0.991	0.561	0.457
	afternoon	0.987	0.828	0.366	1.000	0.001	0.973
	pm peak	0.998	0.113	0.737	1.000	0.002	0.968
	night	1.000	0.005	0.944	1.000	0.022	0.882
	early morning	0.994	0.382	0.539	0.991	0.589	0.446

* indicates a significant mean difference ($\alpha = 0.05$).

This study also compared differences between the two groups through standard deviation of accelerations, which sometimes referred to acceleration noise (Equation 23).

$$\sigma^2 = \frac{1}{T}\sum_{1}^{T}(a(t) - \bar{a})^2 \quad (23)$$

where σ is acceleration noise, T is a total time, $a(t)$ is acceleration of a vehicle at time t, and $\overline{a}$ is the average of accelerations during a trip.

However, this study did not find any significant differences in acceleration noise between the two driver-groups using both the bootstrap technique (Figure 57) and the Wilks' lambda test (Table 48).

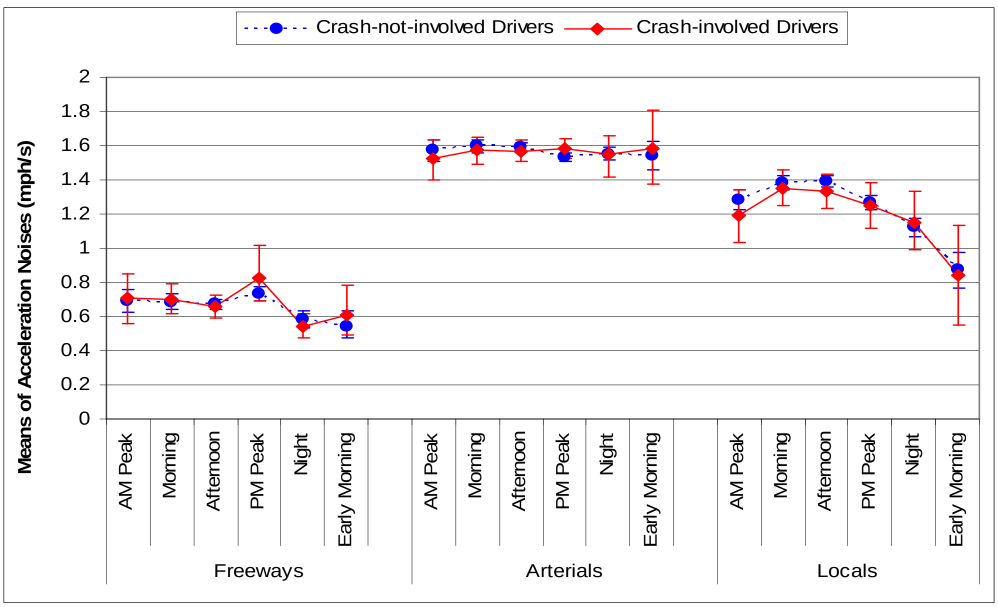


Figure 57: Confidence Intervals of the Means of Acceleration Noises Using the Bootstrap Technique

Table 48: Tests of Equality of Acceleration Noises between Two Groups Based on the Wilks' Lambda Test

Facility Type	Trip Time	Wilks' Lambda	F	Sig.
Freeways	am peak	0.997	0.204	0.653
	morning	1.000	0.020	0.889
	afternoon	0.999	0.036	0.850
	pm peak	0.998	0.156	0.695
	night	0.980	1.286	0.261
	early morning	0.993	0.435	0.512
Arterials	am peak	1.000	0.025	0.874
	morning	0.978	1.415	0.239
	afternoon	0.994	0.383	0.538
	pm peak	1.000	0.001	0.973
	night	1.000	0.028	0.868
	early morning	0.995	0.339	0.563
Local Roads	am peak	0.998	0.129	0.720
	morning	0.995	0.325	0.571
	afternoon	0.999	0.048	0.828
	pm peak	1.000	0.016	0.898
	night	0.997	0.206	0.651
	early morning	0.995	0.293	0.590

* indicates a significant mean difference ($\alpha = 0.05$).

Differences in the Frequencies of Hard Acceleration Activities

This study separately analyzed acceleration values such as accelerations and decelerations and examined the frequency of hard acceleration and deceleration activities similar to the over-speed activities in the previous chapter. This study tested several thresholds for defining the hard acceleration/deceleration activities such as $\pm$ 4 mph/s, $\pm$ 6mph/s, and $\pm$ 8 mph/s. Figure 58 shows the result based on the bootstrap technique, and Table 49 shows the result of the Wilks' lambda test. From both test methods, this study found that the average frequencies of hard acceleration events (4 mph/s) between two driver groups (with and without crash involvements) were significantly different in activities on local roadways during morning (Table 49).

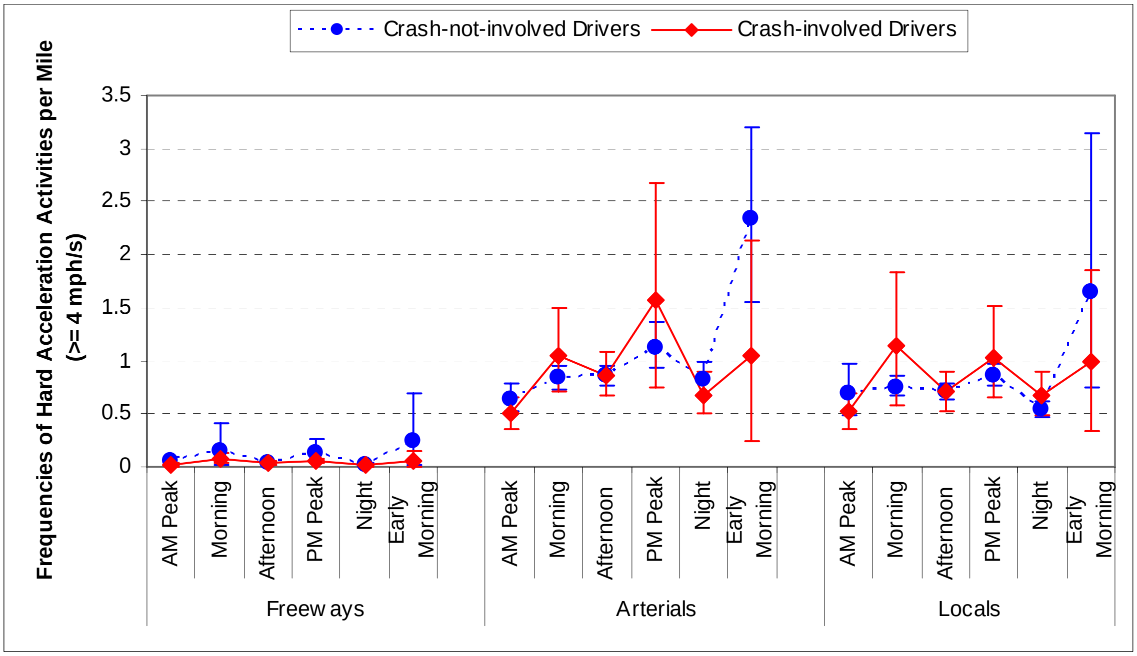


Figure 58: Confidence Intervals of the Means of Frequencies of Hard Acceleration Activities (>= 4 mph/s) per Mile Using the Bootstrap Technique

Table 49: Tests of Equality of Average of Frequency of Hard Acceleration Activities (4mph/s) per Mile between Two Groups Using the Wilks' Lambda Test

Facility Type	Trip Time	Wilks' Lambda	F	Sig.
Freeways	AM peak	0.998	0.332	0.565
	Morning	0.999	0.086	0.770
	Afternoon	0.998	0.298	0.586
	PM peak	0.998	0.263	0.609
	Night	0.998	0.265	0.608
	Early morning	0.998	0.168	0.683
Arterials	AM peak	0.995	0.777	0.379
	Morning	0.989	1.865	0.174
	Afternoon	1.000	0.013	0.909
	PM peak	0.988	1.968	0.163
	Night	0.997	0.504	0.479
	Early morning	0.988	1.356	0.247
Local Roads	AM peak	0.998	0.340	0.561
	Morning *	**0.975**	**4.274**	**0.040**
	Afternoon	1.000	0.005	0.941
	PM peak	0.994	0.987	0.322
	Night	0.988	2.019	0.157
	Early morning	1.000	0.010	0.919

* indicates a significant mean difference ($\alpha = 0.05$).

As shown in Table 50, drivers who were involved in crashes produced hard acceleration such as greater than 4 mph/s on average 1.14 times every mile on local roadways during morning, but drivers who were not involved in crashes provided only 0.76 times per mile. The difference in hard acceleration activities between the two driver-groups was 34 %. For arterials during PM peak, drivers who were involved in crashes produced hard acceleration on average 1.6 times every mile, and drivers who were not involved in crashes provided 1.1 times per mile, but those were not significantly different.

Table 50: Differences and Means of Frequencies of Hard Accelerations (>= 4 mph/s)

Facility Type	Trip Time	Frequency of Hard Accelerations per Mile		Frequency Difference	% Difference
		Drivers who *were not involved* in Crashes	Drivers who *were involved* in Crashes		
Freeways	AM Peak	0.048	0.021	-0.028	-131
	Morning	0.153	0.073	-0.081	-110
	Afternoon	0.033	0.039	0.006	16
	PM Peak	0.127	0.052	-0.075	-145
	Night	0.024	0.014	-0.009	-66
	Early Morning	0.246	0.064	-0.182	-287
Arterials	AM Peak	0.642	0.502	-0.140	-28
	Morning	0.839	1.050	0.211	20
	Afternoon	0.858	0.870	0.012	1
	PM Peak	1.127	1.566	0.439	28
	Night	0.817	0.682	-0.135	-20
	Early Morning	2.332	1.056	-1.275	-121
Local Roads	AM Peak	0.687	0.517	-0.170	-33
	Morning *	**0.758**	**1.141**	**0.383**	**34**
	Afternoon	0.718	0.712	-0.006	-1
	PM Peak	0.867	1.027	0.160	16
	Night	0.539	0.682	0.143	21
	Early Morning	1.645	0.984	-0.661	-67

* indicates a significant mean difference ($\alpha = 0.05$).

When using the different threshold for the hard accelerations such as 6 mph/s, both test methods did not find any differences between the two driver-groups. Figure 59 and Table 51 illustrates the results of the bootstrap technique and the Wilks' lambda test, respectively. Table 52 shows the differences in hard acceleration using 6 mph/s threshold between the two driver-groups.

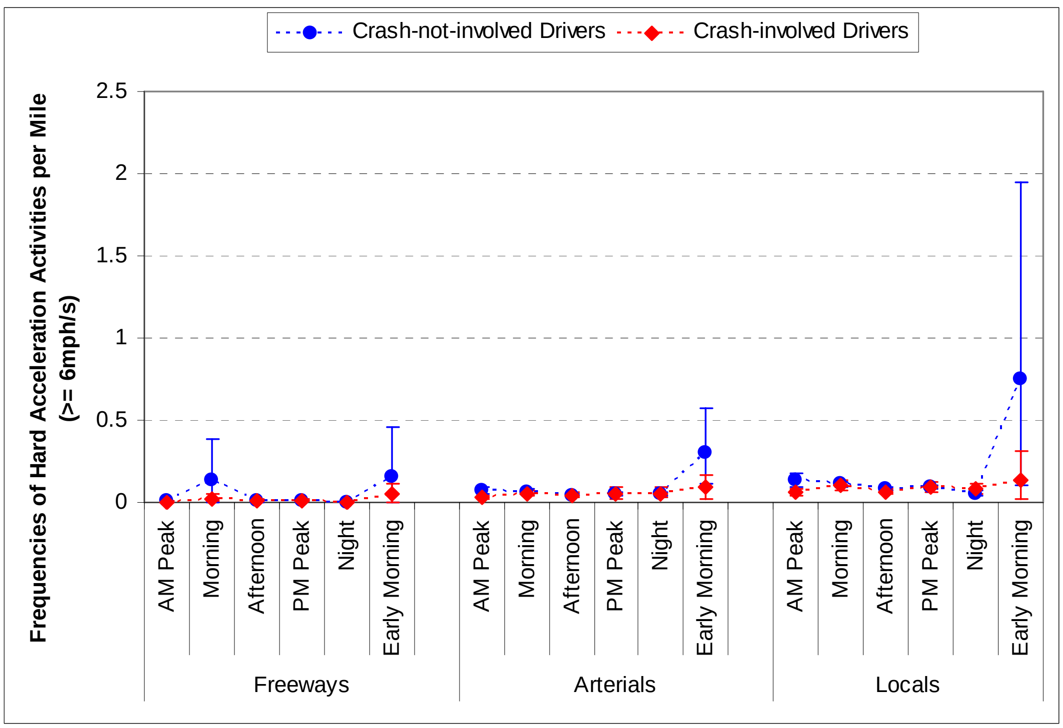


Figure 59: Confidence Intervals of the Means of Frequencies of Hard Acceleration Activities (>= 6 mph/s) Using the Bootstrap Technique

Table 51: Tests of Equality of Average of Frequency of Hard Acceleration Activities (6mph/s) per Mile between Two Groups Using the Wilks' Lambda Test

Facility Type	Trip Time	Wilks' Lambda	F	Sig.
Freeways	AM Peak	1.000	0.041	0.840
	Morning	0.999	0.140	0.709
	Afternoon	0.992	1.343	0.248
	PM Peak	1.000	0.001	0.970
	Night	0.999	0.144	0.705
	Early Morning	0.999	0.098	0.755
Arterials	AM Peak	0.986	2.255	0.135
	Morning	0.998	0.317	0.574
	Afternoon	1.000	0.016	0.900
	PM Peak	1.000	0.005	0.946
	Night	1.000	0.004	0.951
	Early Morning	0.996	0.467	0.496
Local Roads	AM Peak	0.988	1.950	0.164
	Morning	0.998	0.254	0.615
	Afternoon	0.989	1.899	0.170
	PM Peak	1.000	0.003	0.953
	Night	0.987	2.179	0.142
	Early Morning	0.999	0.197	0.658

* indicates a significant mean difference (α = 0.05).

Table 52: Differences and Means of Frequencies of Hard Acceleration Activities (>= 6 mph/s) per Mile

Facility Type	Trip Time	Frequency of Hard Accelerations per Mile		Frequency Difference	% Difference
		Drivers who *were not involved* in Crashes	Drivers who *were involved* in Crashes		
Freeways	AM Peak	0.0062	0.0048	-0.0014	-30
	Morning	0.1363	0.0236	-0.1127	-477
	Afternoon	0.0074	0.0114	0.0040	35
	PM Peak	0.0137	0.0141	0.0004	3
	Night	0.0050	0.0036	-0.0014	-40
	Early Morning	0.1590	0.0497	-0.1093	-220
Arterials	AM Peak	0.0737	0.0293	-0.0443	-151
	Morning	0.0626	0.0528	-0.0098	-19
	Afternoon	0.0460	0.0447	-0.0013	-3
	PM Peak	0.0548	0.0538	-0.0010	-2
	Night	0.0545	0.0557	0.0012	2
	Early Morning	0.3008	0.0907	-0.2101	-232
Local Roads	AM Peak	0.1316	0.0598	-0.0718	-120
	Morning	0.1111	0.1017	-0.0094	-9
	Afternoon	0.0810	0.0649	-0.0160	-25
	PM Peak	0.0909	0.0891	-0.0018	-2
	Night	0.0561	0.0783	0.0223	28
	Early Morning	0.7463	0.1366	-0.6097	-446

* indicates a significant mean difference (α = 0.05).

Similarly, when using the different threshold for the hard accelerations such as 8 mph/s, both test methods did not find any differences between the two driver-groups. Figure 60 and Table 53 illustrates the results of the bootstrap technique and the Wilks' lambda test, respectively. Table 54 shows the differences in hard acceleration using 8 mph/s threshold between the two driver-groups.

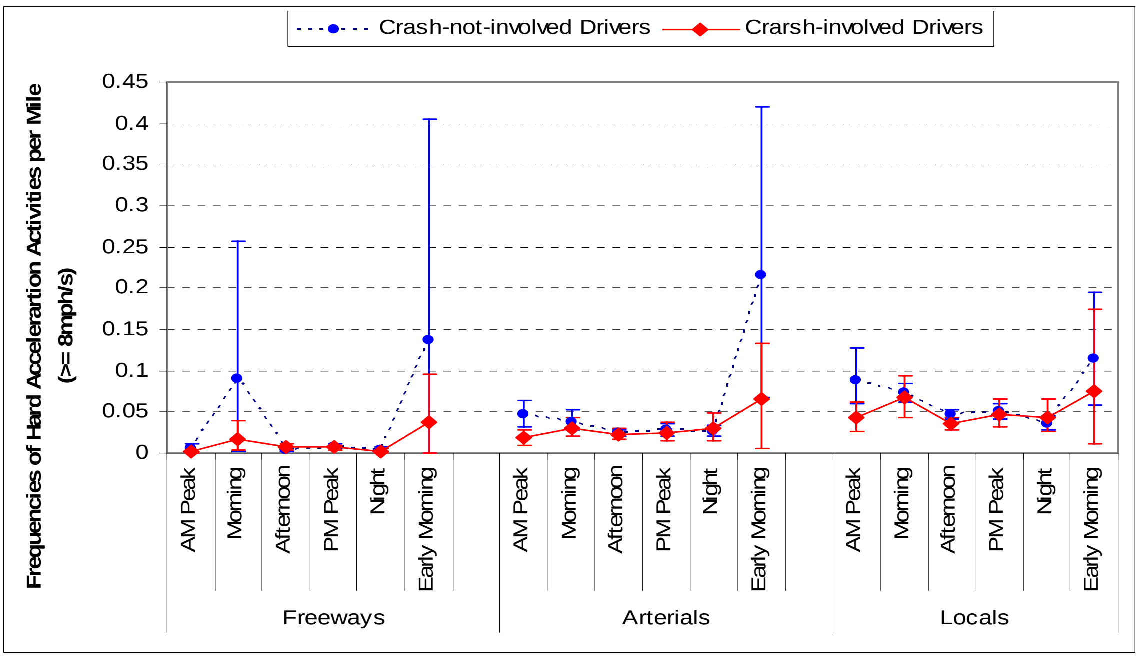


Figure 60: Confidence Intervals of the Means of Frequencies of Hard Acceleration Activities (>= 8 mph/s) per Mile Using the Bootstrap Technique

Table 53: Tests of Equality of Average of Frequency of Hard Acceleration Activities (8mph/s) per Mile between Two Groups Using the Wilks' Lambda Test

Facility Type	Trip Time	Wilks' Lambda	F	Sig.
Freeways	am peak	0.998	0.283	0.595
	Morning	0.999	0.133	0.716
	afternoon	0.981	3.193	0.076
	pm peak	1.000	0.037	0.847
	night	0.999	0.170	0.681
	early morning	0.999	0.107	0.744
Arterials	am peak	0.986	2.221	0.138
	morning	0.998	0.249	0.618
	afternoon	1.000	0.061	0.806
	pm peak	0.999	0.087	0.768
	night	0.999	0.186	0.667
	early morning	0.997	0.385	0.536
Local Roads	am peak	0.993	1.155	0.284
	Morning	0.999	0.159	0.691
	afternoon	0.982	2.999	0.085
	pm peak	0.999	0.114	0.736
	night	0.997	0.469	0.495
	early morning	0.998	0.227	0.634

* indicates a significant mean difference ($\alpha = 0.05$).

Table 54: Differences and Means of Frequencies of Hard Acceleration Activities (>= 8 mph/s) per Mile

Facility Type	Trip Time	Frequency of Hard Accelerations per Mile		Frequency Difference	% Difference
		Drivers who *were not involved* in Crashes	Drivers who *were involved* in Crashes		
Freeways	AM Peak	0.0047	0.0010	-0.0037	-363
	Morning	0.0892	0.0169	-0.0723	-428
	Afternoon	0.0038	0.0069	0.0031	45
	PM Peak	0.0077	0.0069	-0.0007	-10
	Night	0.0040	0.0025	-0.0014	-56
	Early Morning	0.1361	0.0369	-0.0992	-269
Arterials	AM Peak	0.0464	0.0184	-0.0280	-152
	Morning	0.0375	0.0306	-0.0069	-23
	Afternoon	0.0245	0.0234	-0.0010	-4
	PM Peak	0.0275	0.0250	-0.0025	-10
	Night	0.0271	0.0307	0.0036	12
	Early Morning	0.2161	0.0662	-0.1499	-226
Local Roads	AM Peak	0.0878	0.0425	-0.0452	-106
	Morning	0.0727	0.0668	-0.0059	-9
	Afternoon	0.0471	0.0355	-0.0116	-33
	PM Peak	0.0503	0.0465	-0.0038	-8
	Night	0.0355	0.0431	0.0076	18
	Early Morning	0.1146	0.0749	-0.0397	-53

* indicates a significant mean difference ($\alpha = 0.05$).

When using the different threshold for the hard accelerations such as 10 mph/s, while the bootstrap technique did not find any significantly different metrics between the two driver-groups (Figure 61), the Wilks' lambda test showed that hard acceleration activities larger than 10 mph/s of drivers who were involved in crashes were significantly difference from drivers who were not involved in crashes at the 0.05 significance level (Table 55), resulting the difference of 0.003 times per mile (50 %) (Table 56).

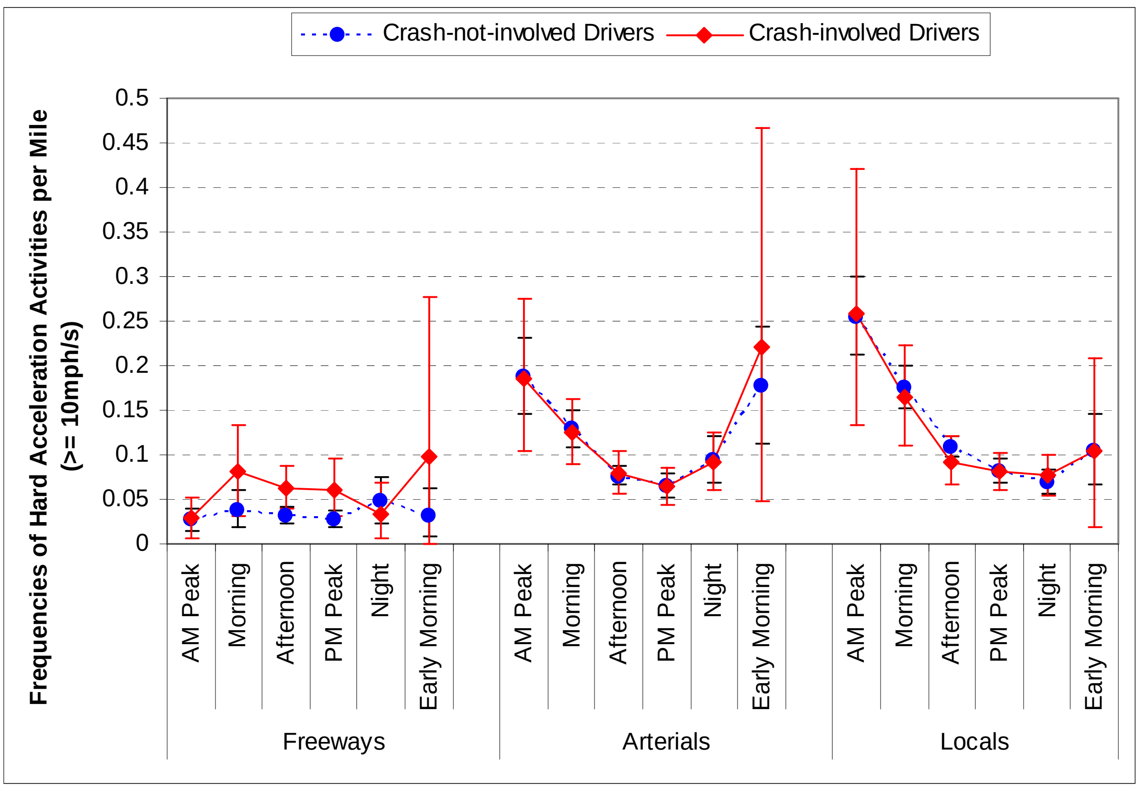


Figure 61: Confidence Intervals of the Means of Frequencies of Hard Acceleration Activities (>= 10 mph/s) per Mile Using the Bootstrap Technique

Table 55: Tests of Equality of Average of Frequency of Hard Acceleration Activities (>= 10 mph/s) per Mile between Two Groups Using the Wilks' Lambda Test

Facility Type	Trip Time	Wilks' Lambda	F	Sig.
Freeways	AM peak	0.998	0.272	0.602
	Morning	0.999	0.106	0.746
	Afternoon *	**0.973**	**4.577**	**0.034**
	PM peak	1.000	0.009	0.926
	Night	1.000	0.056	0.814
	Early morning	0.999	0.118	0.732
Arterials	AM peak	0.988	1.961	0.163
	Morning	1.000	0.051	0.821
	Afternoon	1.000	0.003	0.955
	PM peak	1.000	0.000	0.988
	Night	1.000	0.021	0.884
	Early morning	0.997	0.316	0.575
Local Roads	AM peak	0.993	1.120	0.292
	Morning	1.000	0.020	0.889
	Afternoon	0.977	3.841	0.052
	PM peak	1.000	0.000	0.990
	Night	0.996	0.687	0.408
	Early morning	1.000	0.016	0.901

* indicates a significant mean difference ($\alpha = 0.05$).

Table 56: Differences and Means of Frequencies of Hard Acceleration Activities (>= 8 mph/s) per Mile

Facility Type	Trip Time	Frequency of Hard Accelerations per Mile		Frequency Difference	% Difference
		Drivers who *were not involved* in Crashes	Drivers who *were involved* in Crashes		
Freeways	AM Peak	0.0261	0.0281	0.0020	7
	Morning	0.0383	0.0808	0.0425	53
	Afternoon *	**0.0311**	**0.0621**	**0.0309**	**50**
	PM Peak	0.0271	0.0611	0.0340	56
	Night	0.0475	0.0342	-0.0133	-39
	Early Morning	0.0320	0.0985	0.0665	67
Arterials	AM Peak	0.1881	0.1848	-0.0033	-2
	Morning	0.1288	0.1256	-0.0033	-3
	Afternoon	0.0760	0.0786	0.0026	3
	PM Peak	0.0652	0.0636	-0.0016	-3
	Night	0.0944	0.0923	-0.0021	-2
	Early Morning	0.1778	0.2200	0.0421	19
Local Roads	AM Peak	0.2547	0.2577	0.0029	1
	Morning	0.1756	0.1643	-0.0112	-7
	Afternoon	0.1088	0.0923	-0.0165	-18
	PM Peak	0.0814	0.0811	-0.0003	0
	Night	0.0689	0.0771	0.0082	11
	Early Morning	0.1037	0.1038	0.0001	0

* indicates a significant mean difference ($\alpha = 0.05$).

Differences in the Frequencies of Hard Deceleration Activities

This study evaluates hard deceleration activities between the two driver-groups. Although the bootstrap technique (Figure 62) did not provide any significant differences in hard deceleration events (4 mph/s), the Wilks' lambda test provided three potential exposure metrics, freeways during morning, arterials during morning, and local roadways during nighttime (Table 57). As shown in Table 58, drivers who were involved in crashes produced hard deceleration (greater than 4 mph/s) activities on average 0.23 times every mile on freeways during morning, but drivers who were not involved in crashes provided only 0.07 times per mile. The difference in hard acceleration activities between the two driver-groups was 68 %.

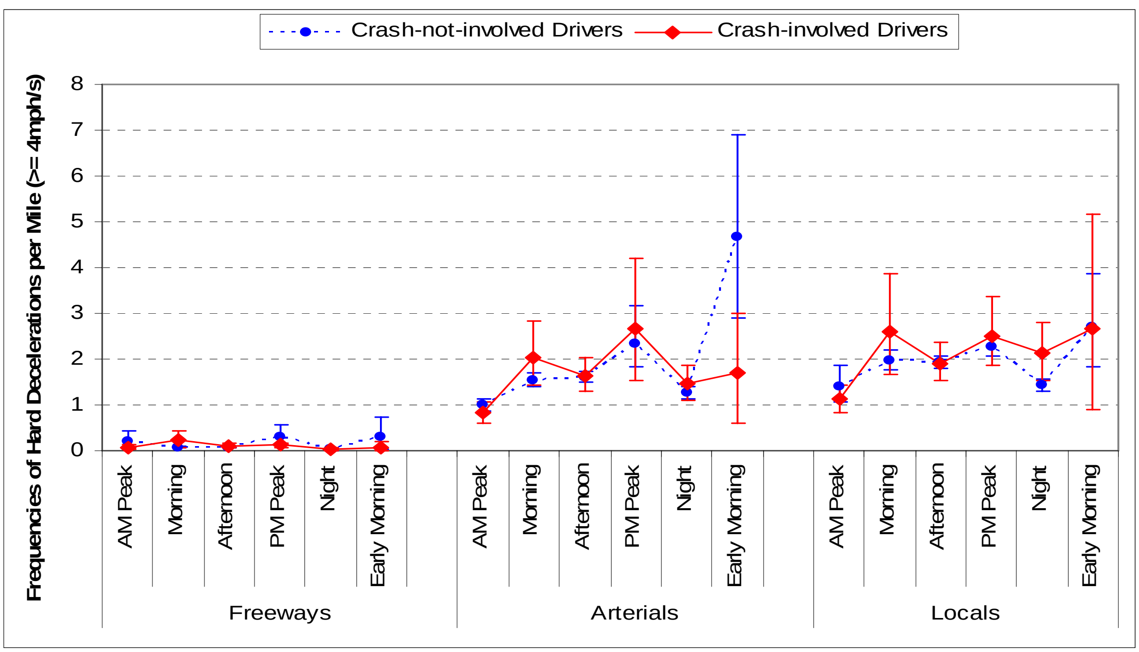


Figure 62: Confidence Intervals of the Means of Frequencies of Hard Deceleration Activities (>= 4 mph/s) per Mile Using the Bootstrap Technique

Table 57: Tests of Equality of Average of Frequency of Hard Deceleration Activities (>= 4 mph/s) per Mile between Two Groups Using the Wilks' Lambda Test

Facility Type	Trip Time	Decelerations larger than 4 mph/s		
		Wilks' Lambda	F	Sig.
Freeways	AM peak	0.998	0.330	0.567
	Morning *	**0.943**	**9.711**	**0.002**
	Afternoon	0.980	3.278	0.072
	PM peak	0.998	0.361	0.549
	Night	1.000	0.004	0.950
	Early morning	0.997	0.289	0.592
Arterials	AM peak	0.994	0.954	0.330
	Morning *	**0.971**	**4.871**	**0.029**
	Afternoon	1.000	0.015	0.903
	PM peak	0.999	0.154	0.695
	Night	0.992	1.337	0.249
	Early morning	0.987	1.447	0.231
Local Roads	AM peak	0.998	0.262	0.610
	Morning	0.982	2.981	0.086
	Afternoon	1.000	0.011	0.916
	PM peak	0.997	0.473	0.493
	Night *	**0.941**	**10.306**	**0.002**
	Early morning	0.998	0.305	0.581

* indicates a significant mean difference ($\alpha = 0.05$).

Table 58: Differences and Means of Frequencies of Hard Deceleration Activities (>= 4 mph/s) per Mile

Facility Type	Trip Time	Frequency of Hard Decelerations per Mile		Frequency Difference	% Difference
		Drivers who *were not involved* in Crashes	Drivers who *were involved* in Crashes		
Freeways	AM Peak	0.198	0.074	-0.124	-169
	Morning *	**0.074**	**0.229**	**0.154**	**68**
	Afternoon	0.067	0.102	0.035	34
	PM Peak	0.296	0.126	-0.170	-134
	Night	0.028	0.029	0.001	3
	Early Morning	0.304	0.066	-0.238	-363
Arterials	AM Peak	1.005	0.824	-0.182	-22
	Morning *	**1.536**	**2.042**	**0.507**	**25**
	Afternoon	1.611	1.636	0.025	2
	PM Peak	2.319	2.659	0.339	13
	Night	1.269	1.465	0.196	13
	Early Morning	4.654	1.692	-2.962	-175
Local Roads	AM Peak	1.389	1.138	-0.251	-22
	Morning	1.978	2.602	0.625	24
	Afternoon	1.927	1.897	-0.030	-2
	PM Peak	2.281	2.491	0.210	8
	Night *	**1.448**	**2.137**	**0.689**	**32**
	Early Morning	2.709	2.652	-0.057	-2

* indicates a significant mean difference ($\alpha = 0.05$).

When using the different threshold for the hard deceleration such as 6 mph/s, while the bootstrap technique still did not find any significantly different metrics between the two driver-groups (Figure 63), the Wilks' lambda test showed five significant hard deceleration activities metrics, freeways during morning and afternoon, arterials during morning, and local roadways during morning and nighttime (Table 59). As shown in Table 60, drivers who were involved in crashes produced hard deceleration (greater than 6 mph/s) activities on average 0.029 times every mile on freeways during morning, but drivers who were not involved in crashes provided only 0.009 times per mile. The difference in hard acceleration activities between the two driver-groups was 71 %.

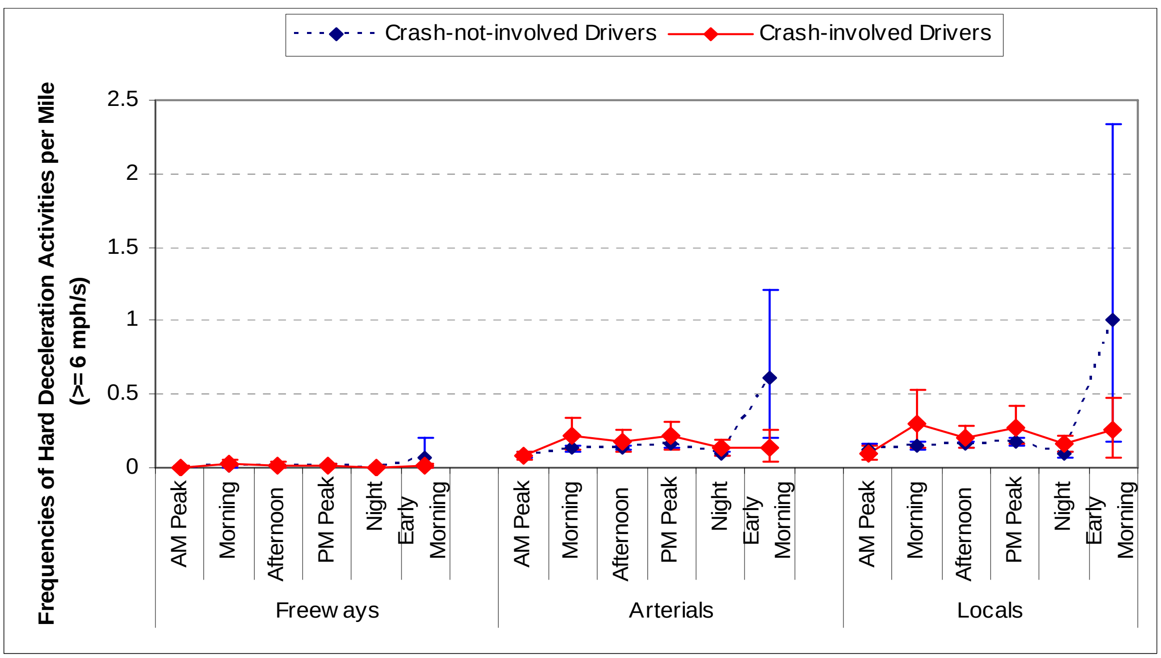


Figure 63: Confidence Intervals of the Means of Frequencies of Hard Deceleration Activities (>= 6 mph/s) per Mile Using the Bootstrap Technique

Table 59: Tests of Equality of Average of Frequency of Hard Deceleration Activities (>= 6 mph/s) per Mile between Two Groups Using the Wilks' Lambda Test

Facility Type	Trip Time	Decelerations larger than 6 mph/s		
		Wilks' Lambda	F	Sig.
Freeways	AM peak	0.998	0.335	0.564
	Morning *	**0.952**	**8.134**	**0.005**
	Afternoon *	**0.966**	**5.676**	**0.018**
	PM peak	1.000	0.011	0.915
	Night	0.999	0.092	0.763
	Early morning	0.998	0.165	0.686
Arterials	AM peak	0.999	0.227	0.634
	Morning *	**0.956**	**7.649**	**0.006**
	Afternoon	0.981	3.155	0.078
	PM peak	0.996	0.675	0.413
	Night	0.982	2.863	0.093
	Early morning	0.995	0.531	0.468
Local Roads	AM peak	0.999	0.242	0.623
	Morning *	**0.947**	**9.214**	**0.003**
	Afternoon	0.988	2.031	0.156
	PM peak	**0.976**	**4.114**	**0.044**
	Night *	**0.954**	**7.923**	**0.005**
	Early morning	0.998	0.283	0.596

* indicates a significant mean difference ($\alpha = 0.05$).

Table 60: Differences and Means of Frequencies of Hard Deceleration Activities (>= 6 mph/s) per Mile

Facility Type	Trip Time	Frequency of Hard Decelerations per Mile		Frequency Difference	% Difference
		Drivers who *were not involved* in Crashes	Drivers who *were involved* in Crashes		
Freeways	AM Peak	0.005	0.006	0.001	20
	Morning *	**0.009**	**0.029**	**0.021**	**71**
	Afternoon	0.008	0.018	0.010	55
	PM Peak	0.017	0.017	-0.001	-4
	Night	0.004	0.003	-0.001	-36
	Early Morning	0.070	0.010	-0.060	-618
Arterials	AM Peak	0.087	0.076	-0.011	-14
	Morning *	**0.131**	**0.221**	**0.090**	**41**
	Afternoon	0.132	0.172	0.040	23
	PM Peak	0.166	0.214	0.048	23
	Night	0.096	0.136	0.040	29
	Early Morning	0.611	0.138	-0.473	-341
Local Roads	AM Peak	0.122	0.100	-0.022	-22
	Morning *	**0.152**	**0.303**	**0.151**	**50**
	Afternoon	0.162	0.203	0.042	21
	PM Peak	0.179	0.272	0.092	34
	Night *	**0.091**	**0.161**	**0.070**	**44**
	Early Morning	1.006	0.255	-0.752	-295

* indicates a significant mean difference ($\alpha = 0.05$).

When using the different threshold for the hard deceleration such as 8 mph/s, while the bootstrap technique still did not find any significantly different metrics between the two driver-groups (Figure 64), the Wilks' lambda test showed only one significant hard deceleration activities metrics, freeways during afternoon (Table 61). As shown in Table 62, drivers who were involved in crashes produced hard deceleration (greater than 6 mph/s) activities on average 0.0041 times every mile on freeways during afternoon, but drivers who were not involved in crashes provided only 0.0018 times per mile. The difference in hard acceleration activities between the two driver-groups was 56 %.

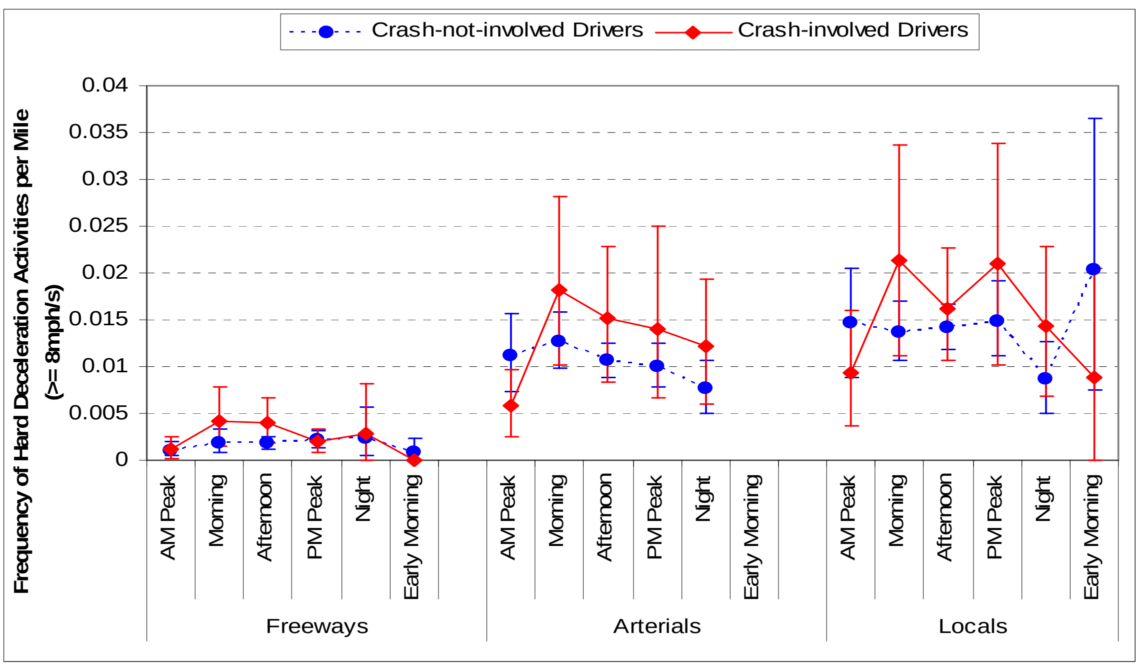


Figure 64: Confidence Intervals of the Means of Frequencies of Hard Deceleration Activities (>= 6 mph/s) Using the Bootstrap Technique

Table 61: Tests of Equality of Average of Frequency of Hard Deceleration Activities (>= 8 mph/s) per Mile between Two Groups Using the Wilks' Lambda Test

Facility Type	Trip Time	Decelerations larger than 8 mph/s		
		Wilks' Lambda	F	Sig.
Freeways	AM Peak	1.000	0.000	0.990
	Morning	0.989	1.854	0.175
	Afternoon *	**0.969**	**5.250**	**0.023**
	PM Peak	1.000	0.020	0.888
	Night	1.000	0.035	0.852
	Early Morning	0.995	0.416	0.521
Arterials	AM Peak	0.993	1.062	0.304
	Morning	0.989	1.870	0.173
	Afternoon	0.984	2.598	0.109
	PM Peak	0.992	1.278	0.260
	Night	0.991	1.396	0.239
	Early Morning	0.998	0.255	0.615
Local Roads	AM Peak	0.997	0.549	0.460
	Morning	0.984	2.752	0.099
	Afternoon	0.997	0.425	0.515
	PM Peak	0.991	1.472	0.227
	Night	0.991	1.415	0.236
	Early Morning	0.998	0.340	0.561

* indicates a significant mean difference ($\alpha = 0.05$).

Table 62: Differences and Means of Frequencies of Hard Deceleration Activities (>= 8 mph/s)

Facility Type	Trip Time	Frequency of Hard Decelerations per Mile		Frequency Difference	% Difference
		Drivers who *were not involved* in Crashes	Drivers who *were involved* in Crashes		
Freeways	AM Peak	0.0011	0.0011	0.0000	2
	Morning	0.0019	0.0042	0.0024	56
	Afternoon *	**0.0018**	**0.0041**	**0.0023**	**56**
	PM Peak	0.0021	0.0020	-0.0002	-9
	Night	0.0023	0.0029	0.0006	20
	Early Morning	0.0009	0.0000	-0.0009	-
Arterials	AM Peak	0.0111	0.0058	-0.0053	-91
	Morning	0.0126	0.0181	0.0055	30
	Afternoon	0.0106	0.0151	0.0045	30
	PM Peak	0.0100	0.0139	0.0039	28
	Night	0.0076	0.0122	0.0045	37
	Early Morning	0.1651	0.0083	-0.1569	-1900
Local Roads	AM Peak	0.0146	0.0093	-0.0053	-58
	Morning	0.0136	0.0214	0.0078	36
	Afternoon	0.0141	0.0162	0.0021	13
	PM Peak	0.0148	0.0210	0.0062	29
	Night	0.0086	0.0143	0.0057	40
	Early Morning	0.0203	0.0089	-0.0114	-129

* indicates a significant mean difference ($\alpha = 0.05$).

Similar to the hard acceleration patterns using the 10 mph/s threshold, the bootstrap technique did not find any significantly different metrics from the average frequencies of hard decelerations using the same threshold. The Wilks' lambda test (Table 63) also did not provide any significant metrics regarding hard deceleration activities with this threshold (10 mph/s).

Table 63: Tests of Equality of Frequency of Hard Deceleration Activities (10 mph/s) per Mile between Two Groups Using the Wilks' Lambda Test

Facility Type	Trip Time	Decelerations larger than 10 mph/s		
		Wilks' Lambda	F	Sig.
Freeways	AM Peak	1.000	0.004	0.952
	Morning	0.986	2.214	0.139
	Afternoon	1.000	0.006	0.937
	PM Peak	0.999	0.142	0.707
	Night	0.984	2.343	0.128
	Early Morning	0.995	0.385	0.537
Arterials	AM Peak	0.991	1.398	0.239
	Morning	1.000	0.014	0.905
	Afternoon	0.996	0.630	0.429
	PM Peak	1.000	0.063	0.803
	Night	0.999	0.139	0.709
	Early Morning	0.997	0.350	0.555
Local Roads	AM Peak	0.997	0.482	0.488
	Morning	1.000	0.013	0.909
	Afternoon	0.999	0.137	0.711
	PM Peak	1.000	0.002	0.965
	Night	0.992	1.325	0.251
	Early Morning	0.997	0.499	0.481

* indicates a significant mean difference ($\alpha = 0.05$).

Finally, this study found that most significant deceleration activities were occurred during morning and afternoon, which may imply that drivers with crash involvements may be undertaking tailgating behavior or other factors such as cellular phone use since the periods of morning and afternoon did not generally indicate the congested traffic condition. Brookhuis et al. [59] investigated the relationship between cellular phone use and driver performance of 12 drivers in an instrumented passenger car on the road measured every work day for 3 weeks and found that statistically significant increase in brake reaction time to adapt to a slowing lead vehicle. They also found that drivers who were using a cellular phone while driving did not decrease their driving speeds.

The Linear Discriminant Analysis using Acceleration-related Behavior Exposures

Based on the bootstrap technique and the Wilks' lambda test, this study selected 11 potential acceleration-related exposure metrics showing significant differences between the two driver-groups. While those 11 exposure metrics can be individually used to verify the potential crash risk drivers, for the modeling process, the issues on the correlations should be treated. Table 64 illustrates the result of the correlation analysis using significantly different 11 acceleration-related exposure metrics.

Using the structure coefficients in Table 65, this study selected metrics having higher structure coefficients (loading power) among the correlated variables. Based on the structure coefficients, this study finally obtained six acceleration-related metrics that can be potentially used for classifying drivers into two crash involvement groups. Those six selected metrics are not correlated each other and show significantly differences between the two driver-groups. Those metrics are used for developing the final discriminant model in Chapter 11.

Table 64: Correlation Analysis with the 11 Acceleration-related Exposure Metrics

	Frwy afternoon mean	Local morning hard acceleration (4mp/s)	Frwy morning hard acceleration (10mph/s)	Frwy morning hard deceleration (4mph/s)	Artrl morning hard deceleration (4mph/s)	Local night deceleration (4mph/s)
Frwy afternoon mean	1.0	-	-	-	-	-
Local morning hard acceleration (4mp/s)	0.4	1.0	-	-	-	-
Frwy morning hard acceleration (10mph/s)	0.0	0.1	1.0	-	-	-
Frwy morning hard deceleration (4mph/s)	0.3	0.5	0.0	1.0	-	-
Artrl morning hard deceleration (4mph/s)	0.3	0.7	0.0	0.4	1.0	-
Local night deceleration (4mph/s)	0.0	0.5	0.1	0.1	0.5	1.0
frwy_morning_dec6	0.3	0.3	0.0	0.9	0.3	0.1
artrl_morning_dec6	0.2	0.7	0.0	0.1	0.8	0.6
local_morning_dec6	0.3	0.7	0.1	0.3	0.7	0.6
local_night_dec6	0.1	0.3	0.0	0.0	0.3	0.6
frwy_afternoon_dec8	0.0	0.3	0.2	0.3	0.1	0.2

	Frwy morning deceleration (6mph/s)	Artrl morning deceleration (6mph/s)	Local morning deceleration (6mph/s)	Local night deceleration (6mph/s)	Frwy afternoon deceleration (8mph/s)
Frwy afternoon mean	-	-	-	-	-
Local morning hard acceleration (4mp/s)	-	-	-	-	-
Frwy morning hard acceleration (10mph/s)	-	-	-	-	-
Frwy morning hard deceleration (4mph/s)	-	-	-	-	-
Artrl morning hard deceleration (4mph/s)	-	-	-	-	-
Local night deceleration (4mph/s)	-	-	-	-	-
Frwy morning deceleration (6mph/s)	1.0	-	-	-	-
Artrl morning deceleration (6mph/s)	0.2	1.0	-	-	-
Local morning deceleration (6mph/s)	0.3	0.8	1.0	-	-
Local night deceleration (6mph/s)	-0.1	0.4	0.4	1.0	-
Frwy afternoon deceleration (8mph/s)	0.2	0.1	0.1	0.1	1.0

Table 65: Structure Coefficients from the Linear Discriminant Analysis (Acceleration Exposures)

Metric	Measure	Threshold	Facility Type	Time	Structure Coefficients	Rank	Variable Selection
Frequency	**Deceleration**	**4 mph/s**	**Local**	**Night**	**0.617**	**1**	√
Frequency	**Deceleration**	**4 mph/s**	**Freeway**	**Morning**	**0.592**	**2**	√
Frequency	Deceleration	6 mph/s	Local	Morning	0.563	3	-
Frequency	Deceleration	6 mph/s	Freeway	Morning	0.542	4	-
Frequency	Deceleration	6 mph/s	Local	Night	0.535	5	-
Frequency	Deceleration	6 mph/s	Arterial	Morning	0.511	6	-
Frequency	**Deceleration**	**8 mph/s**	**Freeway**	**Afternoon**	**0.424**	**7**	√
Frequency	**Deceleration**	**4 mph/s**	**Arterial**	**Morning**	**0.414**	**8**	√
Mean	**All**	**-**	**Freeway**	**Afternoon**	**0.412**	**9**	√
Frequency	Accelerations	4 mph/s	Local	Morning	0.379	10	-
Frequency	**Accelerations**	**10 mph/s**	**Freeway**	**Morning**	**0.064**	**11**	√

Finally, Table 66 shows the performance of the linear discriminant analysis using the selected six potential acceleration-related exposure metrics. As a result, 80.6 % of drivers who were not involved in crashes and 50.0 % of drivers who were involved in crashes were correctly classified. Overall performance of the model using acceleration-related metrics was 75.6 %. Similar to the previous analyses in this study, it can be explained that 19.4 % of drivers who were not involved in crashes might be potentially crash risk drivers based on their acceleration exposures.

Table 66: Classification Results Using Acceleration-related Exposures

Crash Involvements	Predicted Group Membership	
	Drivers who _were not involved_ in crashes	Drivers who _were involved_ in crashes
Drivers who _were not involved_ in crashes	80.6 %	19.4 %
Drivers who _were involved_ in crashes	50.0 %	50.0 %

Summary of the Acceleration-related Exposure Metrics

This study evaluated differences in acceleration-related exposure metrics of drivers with and without crash involvements during the 14-months period to verify whether their acceleration behaviors were significantly different with respect to facility types and time of day. As a result, this study found that drivers who were involved in crashes were more likely to produce large acceleration levels and frequent hard acceleration and deceleration events than drivers who were not involved in crashes. The findings on acceleration-related exposure metrics are summarized as follows:

1. Similar to the speed patterns, this study found that drivers who had experienced crashes more frequently produced large acceleration and deceleration activities.
2. This study verified that accelerations have a positive relationship with the crash involvement rate, so this result also supports the conventional theory, "higher acceleration rates and more frequent hard acceleration events (larger speed changes) means higher opportunity of being involved in a crash".
3. Although the amount of accelerations and the frequency of hard acceleration and deceleration events on arterials or local roadways were larger than those on freeways, large behavioral differences between the two driver-groups were found from the activities on freeways instead of on arterials and local roadways.
4. Significant differences especially during morning and afternoon periods between the two driver-groups imply that drivers who were involved in

crashes may be more likely to conduct tailgating behavior or use cellular phones. Further study will be needed to confirm the impacts of tailgating behavior and cellular phone use on hard acceleration and deceleration activities using the more sophisticated instruments.

5. This study also suggests that acceleration exposure metrics can be used as one of potential behavioral crash risk measures for identifying potentially high crash risk drivers.
6. Safety engineers and policy makers need to educate individual drivers through education campaigns and driver evaluation or monitoring programs to avoid tailgating behaviors and to travel with appropriate safety distance.
7. Insurance companies may enhance current insurance classification decision rules with adding the more detailed acceleration exposure metrics defined by this study.
8. Acceleration-related behavioral exposure metrics may be also employed as the safety surrogate measures to select hazardous roadway segments or intersections where hard decelerations frequently occur (This application is discussed in Chapter 11 in detail).

Chapter Nine

POTENTIAL BEHAVIORAL EXPOSURE V: Speed Stability

This study evaluated differences in travel mileage, travel duration, speed behavior, and acceleration patterns in previous chapters. In addition to those potential driving behavior activity exposure measures, this study also tried to evaluate speed stability patterns of drivers who were involved and not involved in crashes. From the literature reviews, it was found that numerous studies claimed that the speed variation is the more important factor than speed itself since large speed variations can cause more frequent traffic conflicts. A driver who usually drives at high speeds such as 15 mph greater than the posted speed limit (this driver is usually regarded as a high speeder indicating high potential crash risk) may be considered as a normal speed driver if the speeds of the surrounding traffic are also 15 mph greater than the speed limits because the speed difference (or variation) between the surrounding traffic and this driver is very low, indicating possibly lower conflict rates.

While the speed variation can be estimated from all speed data from vehicles traveled on the certain type of roadways, in fact, it is hard to measure the speed variation from individual drivers using the GPS data. Thus, this study proposes the surrogate measure of speed variations using the GPS instrumented vehicles [60]. The speed stability pattern implies how long a driver can travel without changing speeds, given the acceleration cut-point, this study considers the speed difference less than ± 1 mph (and also ± 2 mph) between sequential speeds as a speed stability, also called a cruise duration.

However, a variety of other cut-points could be used to differentiate between mild, moderate, and hard acceleration/deceleration activity such as emissions research.

The speed stability pattern also indicates how frequently drivers change their driving speeds. To better describe individual driver behavior and distinguish between non-crash and crash groups, researchers must estimate not only the amount of acceleration or deceleration describing hard acceleration behaviors but also the frequency pattern of speed changes that may indicate abnormal driving habits [60].

Another potential benefit of using the speed stability measure to describe driver behavior is that such metrics may be easier to communicate to individual drivers in the effort to modify driver behavior through the driver training. That is, it may be easier to train a driver to maintain smooth vehicle operation, than it is to train a driver to reduce their average acceleration rate by 0.5 mph/second. The frequency pattern of speed changes may help drivers evaluate and change their own behaviors when participating in value pricing insurance incentive programs.

Hence, this study proposed new time-based speed stability patterns indicating how frequently drivers change their driving speeds and examined the differences in speed stability patterns.

Differences in the Speed Stability Patterns (Cruise Mode Patterns)

Table 67 shows the means of average speed stability patterns based on facility and time of day between the two driver-groups and indicates that all values of drivers who were involved in crashes were lower than those of drivers who were not involved in crashes (Threshold: ± 1 mph). As mentioned earlier, the speed stability is the duration of time without speed changes (cruise speed durations). Conversely, it can be interpreted how frequently drivers change their driving speeds. For example, average cruise duration on freeways during AM peak of drivers who were not involved in crashes was 76 seconds and that of drivers who were involved in crashes was 58 seconds. It indicated that drivers without crash involvements changed their speeds every 76 seconds and drivers with crash involvements changed their speeds every 58 seconds during AM peak on freeways. Thus, the results in Table 67 shows that drivers who were involved in crashes tended to much more frequently change their driving speeds all the time and may undertake tailgating.

Table 67: Differences and Means of Average Speed Stability based on Facility and Time (Threshold: ± 1 mph)

Facility Type	Trip Time	Mean of Average Speed Stability (seconds)		Stability Difference	% Difference
		Drivers who *were not involved* in crashes	Drivers who *were involved* in crashes		
Freeways	AM Peak	75.68	57.67	-18.01	-31
	Morning	73.54	64.69	-8.85	-14
	Afternoon	60.51	55.45	-5.06	-9
	PM Peak	54.26	46.88	-7.39	-16
	Night	91.16	78.27	-12.89	-16
	Early Morning	111.98	95.44	-16.55	-17
Arterials	AM Peak	16.48	16.98	0.50	3
	Morning	14.82	14.87	0.05	0
	Afternoon	13.25	13.17	-0.08	-1
	PM Peak	13.62	13.60	-0.02	0
	Night	18.77	18.12	-0.66	-4
	Early Morning	25.79	28.81	3.03	11
Local Roads	AM Peak	11.01	10.93	-0.07	-1
	Morning	9.86	9.87	0.01	0
	Afternoon	9.31	9.51	0.20	2
	PM Peak	9.75	9.67	-0.09	-1
	Night	10.71	10.43	-0.28	-3
	Early Morning	12.60	11.73	-0.87	-7

* indicates a significant mean difference ($\alpha = 0.05$).

Table 68 also shows the means of average speed stability patterns based on facility and time between the two driver-groups and indicates that all values of drivers who were involved in crashes were lower than those of drivers who were not involved in crashes (Threshold: ± 2 mph).

Table 68: Differences and Means of Average Speed Stability based on Facility and Time (Threshold: ± 2 mph)

Facility Type	Trip Time	Mean of Average Speed Stability (seconds)		Stability Difference	% Difference
		Drivers who *were not involved* in crashes	Drivers who *were involved* in crashes		
Freeways	AM Peak	283	223	-60	-27
	Morning	277	258	-20	-8
	Afternoon	213	224	10	5
	PM Peak	200	196	-4	-2
	Night	323	344	21	6
	Early Morning	391	279	-113	-40
Arterials	AM Peak	30	31	1	2
	Morning	28	26	-1	-5
	Afternoon	25	25	0	2
	PM Peak	26	25	0	-2
	Night	34	34	0	0
	Early Morning	43	49	5	11
Local Roads	AM Peak	22	22	0	2
	Morning	20	20	1	4
	Afternoon	19	19	0	1
	PM Peak	20	20	0	0
	Night	22	20	-1	-7
	Early Morning	25	23	-1	-6

* indicates a significant mean difference ($\alpha = 0.05$).

However, Figure 65 and 66 depicts that all speed stability metrics between the two driver-groups were not significantly different based on the result of the bootstrap technique. Although any significant differences in speed stability metrics were not be obtained, this study found that the speed changes heavily occurred on local roadways and arterials due to lots of traffic conflicts and intersection impacts.

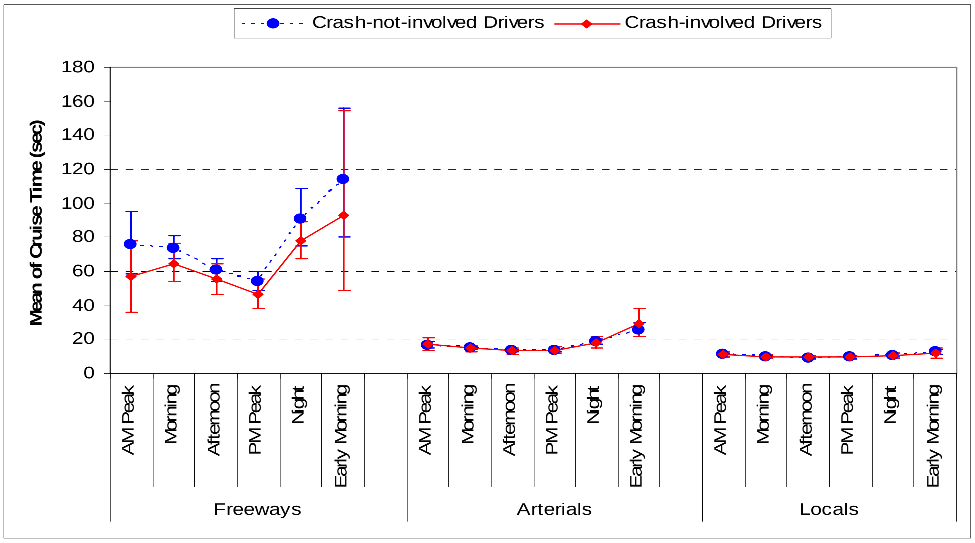


Figure 65: Confidence Intervals of the Means of Average Speed Stability Durations Using the Bootstrap Technique (Threshold: ± 1 mph)

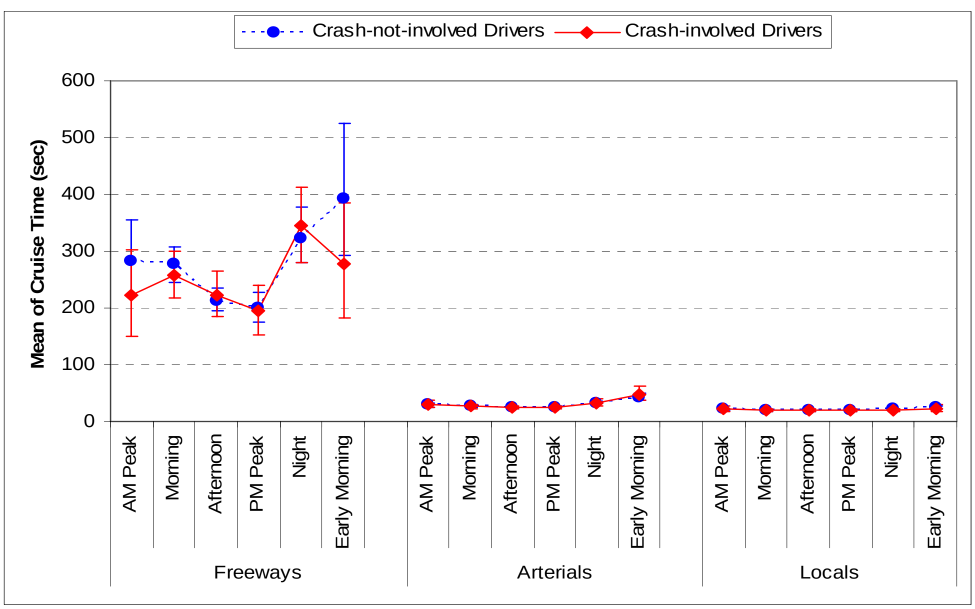


Figure 66: Confidence Intervals of the Means of Average Speed Stability Durations Using the Bootstrap Technique (Threshold: ± 2 mph)

In addition, the Wilks' lambda test did not provide any statistically different speed stability metrics, either (Table 69). Since speed stability exposure metrics did not provide any statistical differences between the two driver-groups, this study did not perform the linear discriminant analysis.

Table 69: Tests of Equality of Speed Stability Metrics Using the Wilks' Lambda Test

Facility Type	Trip Time	Threshold: ± 1 mph			Threshold: ± 2 mph		
		Wilks' Lambda	F	Sig.	Wilks' Lambda	F	Sig.
Freeways	AM Peak	0.999	0.081	0.777	0.998	0.133	0.716
	Morning	0.993	0.387	0.536	0.997	0.185	0.668
	Afternoon	0.995	0.281	0.598	0.997	0.159	0.692
	PM Peak	0.995	0.263	0.61	0.997	0.165	0.686
	Night	1	0.001	0.972	0.988	0.656	0.422
	Early Morning	1	0.005	0.944	0.990	0.560	0.457
Arterials	AM Peak	0.995	0.295	0.589	0.994	0.341	0.562
	Morning	0.993	0.394	0.533	0.999	0.079	0.779
	Afternoon	0.982	1.021	0.317	0.969	1.744	0.192
	PM Peak	0.994	0.304	0.583	0.996	0.206	0.652
	Night	0.98	1.137	0.291	0.991	0.515	0.476
	Early Morning	0.997	0.174	0.678	0.996	0.227	0.636
Local Roads	AM Peak	0.999	0.046	0.832	1.000	0.000	0.993
	Morning	0.998	0.098	0.756	0.992	0.459	0.501
	Afternoon	0.987	0.718	0.4	0.990	0.561	0.457
	PM Peak	1	0.003	0.955	1.000	0.004	0.949
	Night	0.996	0.217	0.643	0.988	0.668	0.417
	Early Morning	0.993	0.363	0.549	0.991	0.493	0.486

* indicates a significant mean difference ($\alpha = 0.05$).

Summary of the Speed Stability Exposure Metrics

To understand better driving behavior activities associated with vehicle speed, acceleration, and deceleration, this study examined a new metric indicating speed stability patterns with the longitudinally observed speed trajectories of individual drivers, which could not be investigated in previous studies due to the difficulty of data collection.

Although this study did not find any statistical differences in speed stability metrics between drivers who were involved and were not involved in crashes, this study found that drivers with crash involvements tended to much more frequently change their driving speeds than drivers without crash involvements.

Thus, this study suggests that researchers who have larger sample data in terms of the number of drivers and the period of data collection in future need to reexamine the speed stability patterns between crash-involved and crash-not-involved drivers. Although this study evaluated disaggregated behavioral exposures based on time of day and facility type, further investigations regarding exposures to roadways having different geometric designs (grade and curvature) and operational designs (speed limit and traffic volume) need to be performed. In addition, speed difference between individual driving speed and surrounding traffic speed may be one of potential behavioral crash-related exposure measures. The linear discriminant analysis was not performed since those exposure metrics were not significantly different between the two driver-groups.

Chapter Ten

POTENTIAL BEHAVIRAL EXPOSURE VI: Unfamiliar Roadway, Left/Right Turns, and Previous Crash Location Exposure

In the previous chapters, this study examined travel mileage, duration, speed, acceleration, and speed stability patterns as potential behavioral exposure measures. In addition to those measures, other potential driving behavior activity exposure measures include unfamiliar roadway exposure, left/right turn movement exposure, and previous crash location exposure. This chapter examines if those measures can show differences between drivers who experienced crashes and who did not experienced crashes during the 14-months study period.

Unfamiliar Roadway Exposure

Drivers may need to be more cautions when traveling unfamiliar roadways since drivers traveling unfamiliar roadways may have difficulty predicting oncoming roadway designs, such as number of lanes and curvature. Poor driver expectancy on those conditions may cause motor vehicle crashes. However, the unfamiliar roadway is not well defined in previous research. In addition, it is difficult to measure the exposure to unfamiliar roadways. Thus, it is challenge to define the unfamiliar exposure roadway and to evaluate differences in exposures to unfamiliar roadways between the two driver-groups who had and had not crash involvements in the study period.

Definition and Estimation of Unfamiliar Roadway Exposure

Among possible definitions of unfamiliar roadway, this study defined roadways where driver traveled only one time (and less than or equal to two times) during the six-months study period as unfamiliar roadways. Using this definition, this study estimated unfamiliar roadway exposures between the two driver-groups using the GPS-observed data and roadway characteristics (RC) information in GIS network database. Figure 67 and Table 70 illustrate how roadways traveled by each driver during the study period can be defined as familiar roadways, unfamiliar roadways, and unused roadways. As shown Figure 67 and Table 70, RC 1, RC 3, RC 4, and RC 6 were classified into familiar roadways, RC 2 was an unused roadway, and RC 5 was decided as an unfamiliar roadway. Each traverse across a roadway constitutes a roadway exposure.

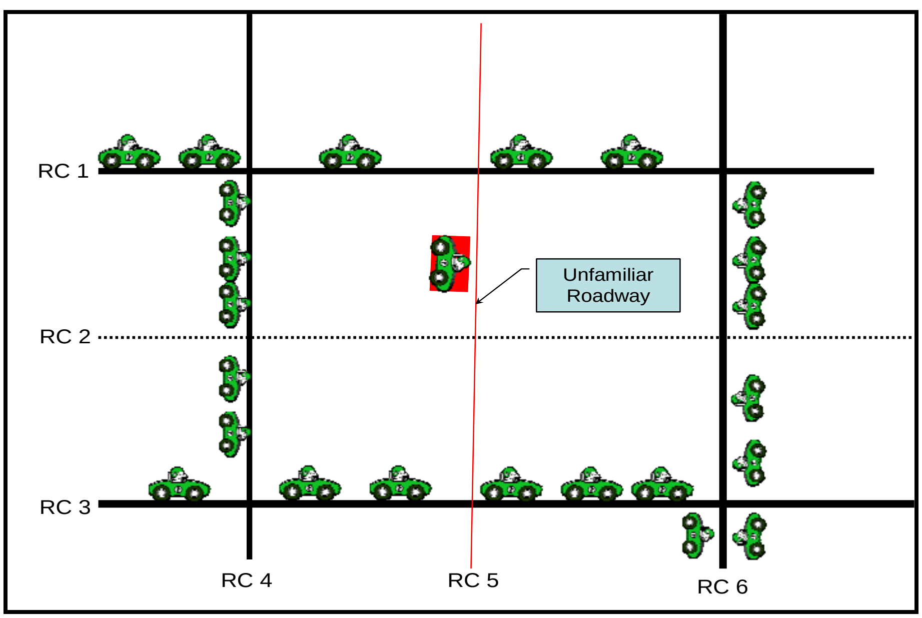


Figure 67: Example of Definition of Unfamiliar Roadways

Table 70: Example of Definitions of Unfamiliar Roadways

(A) Threshold: 1 (One time)

	# of Exposures	Threshold for unfamiliar roadway	Definition
RC 1	5	IF # of Exposures ≥ 2	Familiar roadway
RC 2	0	IF # of Exposures $= 0$	Unused roadway
RC 3	6	IF # of Exposures ≥ 2	Familiar roadway
RC 4	5	IF # of Exposures ≥ 2	Familiar roadway
RC 5	**1**	**IF # of Exposures $= 1$**	**Unfamiliar Roadway**
RC 6	7	IF # of Exposures ≥ 2	Familiar roadway

(B) Threshold: 1 (Two Times)

	# of Exposures	Threshold for unfamiliar roadway	Definition
RC 1	5	IF # of Exposures ≥ 3	Familiar roadway
RC 2	0	IF # of Exposures $= 0$	Unused roadway
RC 3	6	IF # of Exposures ≥ 3	Familiar roadway
RC 4	5	IF # of Exposures ≥ 3	Familiar roadway
RC 5	**1**	**IF # of Exposures ≤ 2**	**Unfamiliar Roadway**
RC 6	7	IF # of Exposures ≥ 3	Familiar roadway

Histogram and Kernel Density Estimation

To determine the shape of unknown distribution F of the frequency of unfamiliar roadways between the two driver-groups, a common histogram method was used with a specified bin width in this study, and a kernel density estimation method was used. The histogram method has potential limitation in that the distribution (or shape) of the histogram strongly depends on data origin, the starting point of the first bin interval, and bin width [17, 45, 46]. Another limitation of the histogram is that some of bin interval may have zero data points, which causes a non-continuity in density distribution if the bin width or the sample size is small. The kernel density estimation is obtained by

$$\hat{f}_k(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{24}$$

where h is a parameter which controls the width of kernel, and $K(t)$ is a kernel function that satisfies

$$\int K(t)dt = 1, \qquad (25)$$

Among various methods that help select a correct width (*h*), this study utilized a normal reference rule due to its simplicity [45, 46]. The normal reference rule for kernel density estimation is given by

$$h_k^* = (\frac{4}{3})^{1/5} \sigma n^{-1/5} \approx 1.06 \sigma n^{-1/5}, \qquad (26)$$

where σ is a standard deviation of sample data.

After running the kernel function through all data points with the same weight $\frac{1}{h}$, each data point x_i has the base density distribution centered at x_i. Then n-density distribution curves are summarized and averaged with the concentration weight [45, 46].

Based on the definition of the unfamiliar roadway, this study estimated numbers of unfamiliar roadway exposures of individual drivers. Figure 68 shows histograms and kernel density distributions of unfamiliar roadway exposures. Based on threshold of one time, the mean of unfamiliar roadway exposures during the 6-months period generated by drivers who were involved in crashes was 106 roadways/six-months and that of drivers who were not involved in crashes was 98 roadways/six-months, a difference of 8 %. In addition, the mode (the highest probability) of unfamiliar roadway exposure of drivers who were involved in crashes was higher than that of drivers who were not involved in crashes.

In addition, Figure 69 and Table 71 show average unfamiliar roadway exposures based on different units (mile and trip).

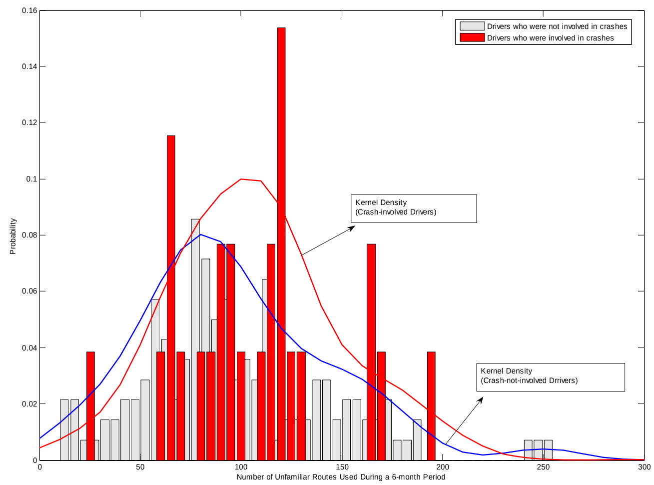
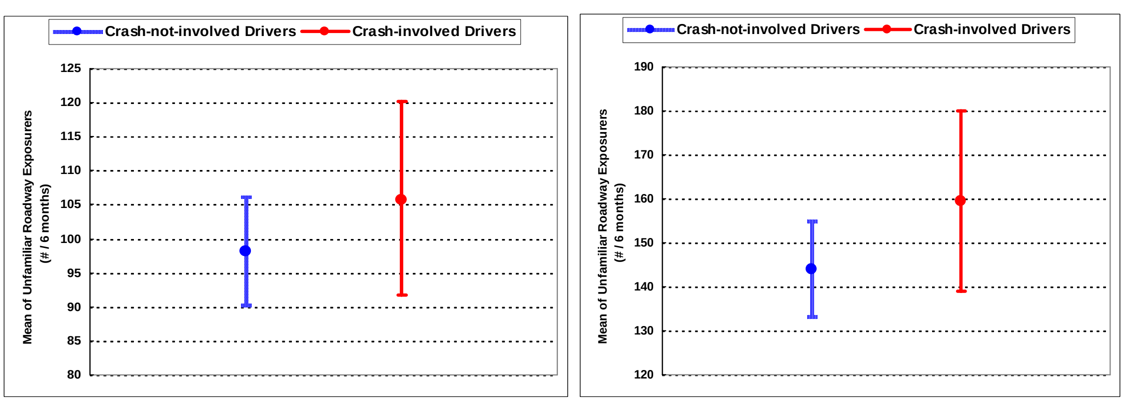
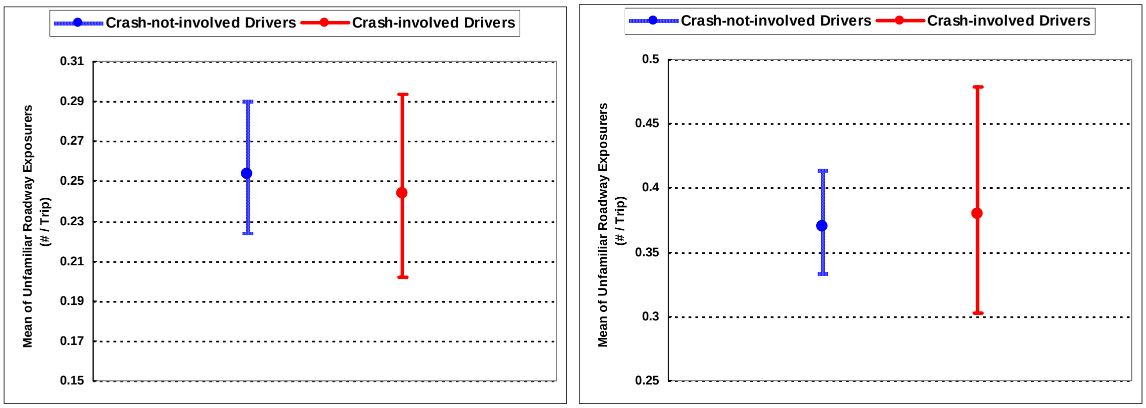


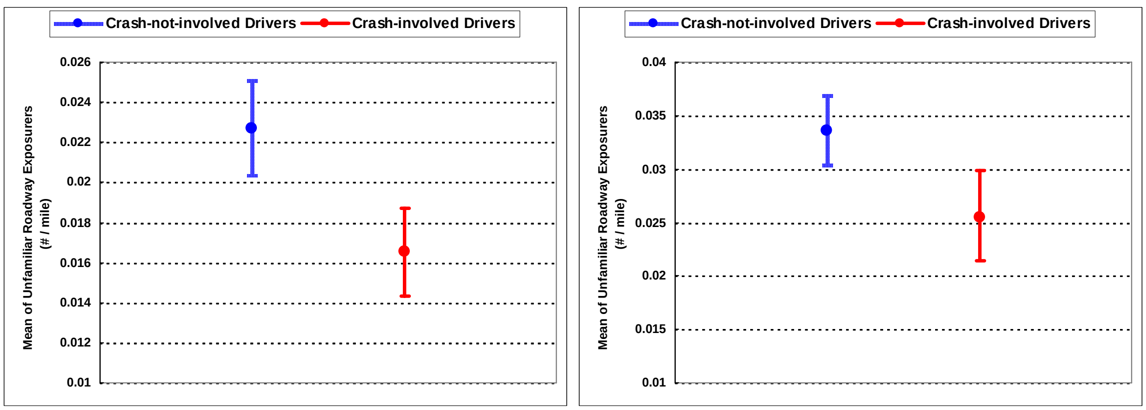
Figure 68: Distributions of Unfamiliar Roadway Exposures of Two Driver Groups using Threshold 1



(A) Total Unfamiliar Exposures (6-months) (Left: Threshold 1 and right: Threshold 2)



(B) Average Unfamiliar Exposures per Trip (Left: Threshold 1 and right: Threshold 2)



(C) Average Unfamiliar Exposures per Mile (Left: Threshold 1 and right: Threshold 2)

Figure 69: Confidence Intervals of Unfamiliar Roadway Exposures Using the Bootstrap Technique

Based on the bootstrap technique and the Wilks' lambda test (Figure 69 and Table 71), unfamiliar roadway exposure by mile provided significant difference between the two driver-groups. However, due to the relatively large difference in travel mileage and small difference in unfamiliar roadway exposures (7% and 10% with respect to thresholds) between the two driver-groups, average unfamiliar roadway exposure of drivers who were not involved in crashes was larger than that of crash-involved drivers. Thus, this study did not consider this metric as a potential behavior exposure measure and suggests that researchers who have larger sample data in terms of the number of drivers and the period of data collection in future need to reexamine this issue. Other metrics such as total unfamiliar exposure and average unfamiliar exposure per trip were not significantly different ($\alpha = 0.05$).

Table 71: Tests of Equality of Unfamiliar Roadway Exposures Using the Wilks' Lambda Test

(A) Threshold: 1 (one time)

	Unfamiliar Roadways		
Normalization	Wilks' Lambda	F	Sig.
by 6-months	0.997	0.576	0.449
by trip	1.000	0.060	0.807
by mile *	**0.973**	**4.538**	**0.035**

* indicates a significant mean difference ($\alpha = 0.05$).

(B) Threshold: 2 (two times)

	Unfamiliar Roadways		
Normalization	Wilks' Lambda	F	Sig.
by 6-months	0.995	0.876	0.351
by trip	1.000	0.057	0.812
by mile *	**0.977**	**3.896**	**0.050**

* indicates a significant mean difference ($\alpha = 0.05$).

Table 72: Differences in Unfamiliar Roadways between the Two Groups

(A) Threshold: 1 (one time)

Normalization	Mean of unfamiliar roadway exposures		Exposure Difference	% Difference
	Drivers who *were not involved* in crashes	Drivers who *were involved* in crashes		
by 6-months	98	106	8	7
by trip	0.25	0.24	-0.01	-4
by mile *	**0.023**	**0.017**	**-0.006**	**-37**

* indicates a significant mean difference ($\alpha = 0.05$).

(B) Threshold: 2 (two times)

Normalization	Mean of unfamiliar roadway exposures		Exposure Difference	% Difference
	Drivers who *were not involved* in crashes	Drivers who *were involved* in crashes		
by 6-months	144	159	16	10
by trip	0.37	0.38	0.01	3
by mile *	**0.034**	**0.025**	**-0.008**	**-32**

In addition to unfamiliar roadway exposures, this study evaluated how many different roadways (like footprint) drivers traveled during the 6-months period. Figure 70 shows histograms and kernel density distributions of number of roadways that the two driver-groups used. The mean of total number of roadways traveled by drivers who were involved in crashes was 288 roadways/six-months and that of drivers who were not involved in crashes was 260 roadways/six-months, a difference of 10% (Table 74). Similar to the unfamiliar roadway exposures, the mode (the highest probability) of number of roadways of drivers who were involved in crashes was higher than that of drivers who were not involved in crashes.

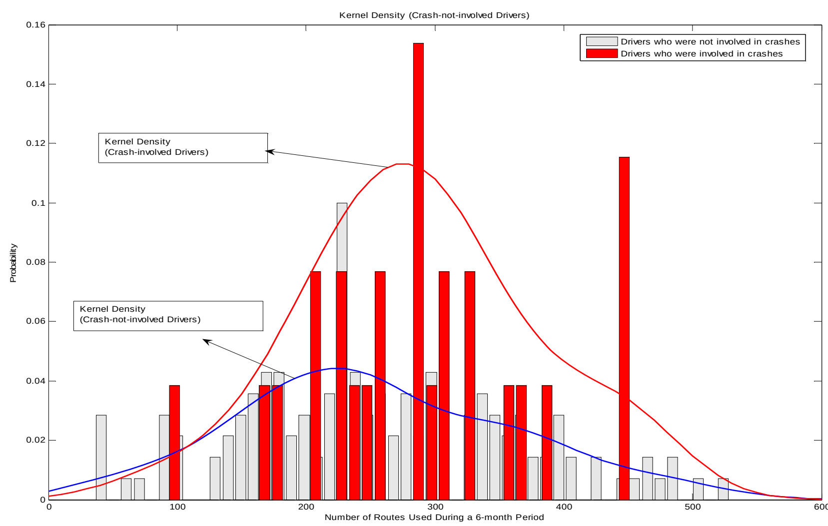
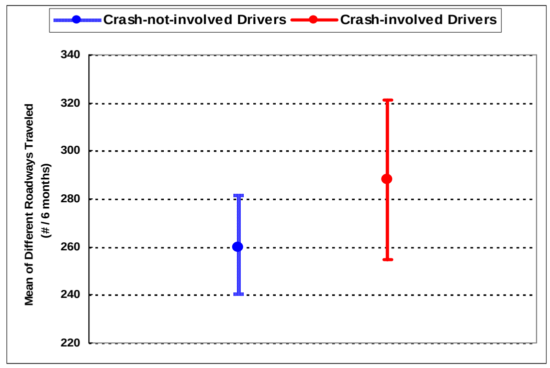
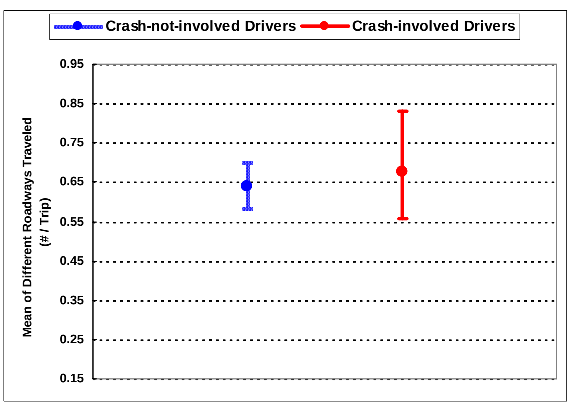


Figure 70: Distributions of Number of Different Roadways Traveled

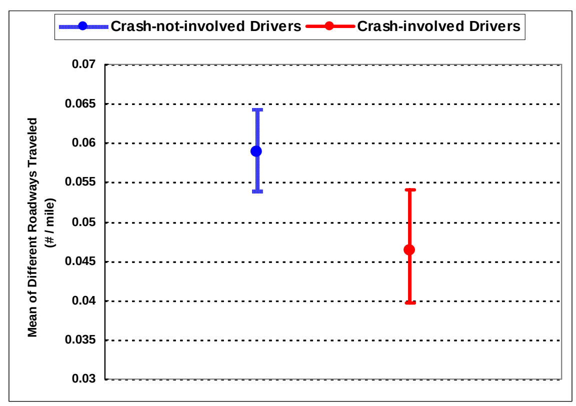
Unlike to the unfamiliar roadway exposure, the means test using the bootstrap technique and the Wilks' lambda test showed that numbers of different roadways traveled by the two driver-groups with and without crash involvements during the 6-months period were not significantly different (α = 0.05) (Figure 71 and Table 73). This result indicates that the two driver-groups generally did not use a statistically different numbers of roadways no matter they were involved or not involved in crashes during the 14-months period.



(A) Total Different Roadway Exposures (6-months)



(B) Average Different Roadway per Trip



(C) Average Different Roadway per Mile

Figure 71: Confidence Intervals of Numbers of Different Roadways Traveled Using the Bootstrap Technique

Table 73: Tests of Equality of Unfamiliar Roadway Exposures and the Numbers of Roadways Traveled between the Two Groups Using the Wilks' Lambda Test

	Total Different Roadways Traveled		
Normalization	Wilks' Lambda	F	Sig.
by 6-months	0.993	1.194	0.276
by trip	0.998	0.259	0.611
by mile	0.978	3.682	0.057

* indicates a significant mean difference (α = 0.05).

Table 74: Differences in Different Roadways Traveled between the Two Groups

Normalization	Numbers of Difference Roadway Exposures		Exposure Difference	% Difference
	Drivers who *were not involved* in crashes	Drivers who *were involved* in crashes		
by 6-months	260	288	28	10
by trip	0.64	0.68	0.04	6
by mile	0.06	0.05	-0.01	-20

* indicates a significant mean difference (α = 0.05).

The correlation analysis (Table 75) indicated that those two exposures, unfamiliar roadway exposure and the numbers of different roadways traveled, were highly correlated, which implies that drivers who traveled more have higher tendency to travel unfamiliar roadways. The linear discriminant analysis was not performed since those exposure metrics were not significantly different between the two driver-groups.

Table 75: Results of Correlation Analysis of Unfamiliar Roadway Exposures and the Numbers of Different Roadways Traveled

Correlation	Unfamiliar Roadway Exposures	Numbers of Different Roadways Traveled
Unfamiliar Roadway Exposures	1.000	-
Numbers of Different Roadways Traveled	**0.920**	1.000

Summary of Unfamiliar Roadway Exposures

This study evaluated differences in the numbers of unfamiliar roadways and of different roadways traveled by drivers with and without crash involvements with the 6-month GPS-collected trip data in order to verify whether those exposure metrics were significantly different or not. The results are summarized as follows:

1. The number of different roadways and the number of unfamiliar roadways traveled by drivers who had crash involvements during the 14-months period were larger (11 % and 8 %, respectively) than those of drivers who did not experience any crashes during the same period. However, those differences were not statistically significant.
2. Unfamiliar roadway exposures and the numbers of different roadways traveled of drivers who were involved in crashes were not significantly larger than drivers who were not involved in crashes.
3. This study suggests that driving behavior measures such as speed and acceleration patterns may be more useful for describing potential crash involvements of individual drivers.
4. This study suggests that researchers with lager samples in future assess difference in driving behavior between travels on familiar roadways and unfamiliar roadways and re-evaluate the relationships between unfamiliar roadway exposures and crash involvement rates.
5. Travel mileage exposure on unfamiliar roadways may be also one of potential measures and further study will be performed.

Left/Right Turns Exposure

Numbers of left and right turn movements may also be possible behavioral exposure measures since turning movements usually result in many conflict events, which are currently using as one of safety surrogate measures for intersection safety evaluation [3]. Thus, this study estimated the frequencies of left and right turns based on each trip made by individual drivers and evaluated if drivers who were involved in crashes had different frequencies of turning movements from drivers who were not involved in crashes (Figure 72).

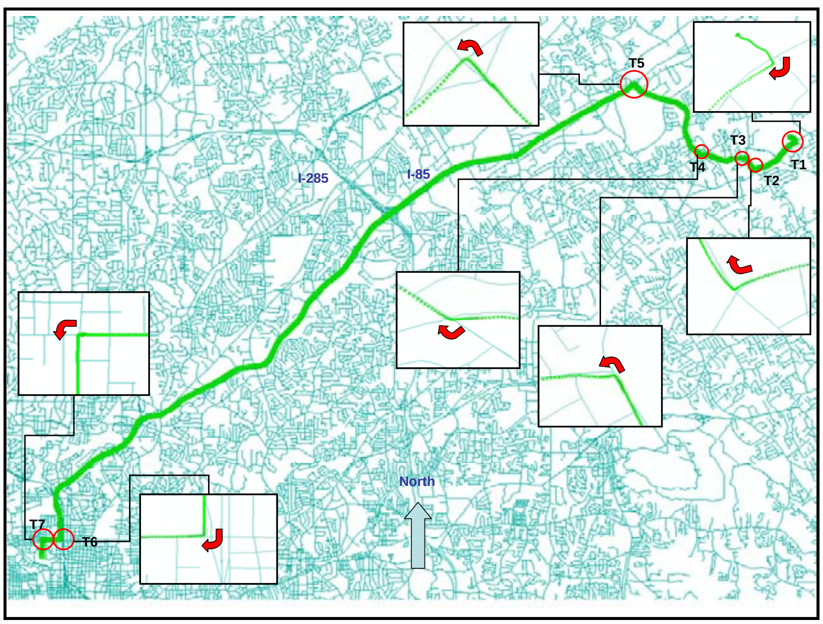


Figure 72: Example of Turning Movements based on Real Travel Route

To estimate the exposures to left and right turns, this study used roadway characteristics (RC) information in GIS roadway network database and heading data in GPS-observed data profiles. Figure 73 shows how left and right turns can be identified from the RC and heading information. When sequential GPS points were matched with the corresponding RC identification number (ID), the angle of each roadway could by estimated by the average of GPS heading values within each roadway [19]. If a heading change (the difference of angles between two roadways) was negative, the turn was considered as a left turn. If a heading change was positive, the turn was considered as a right turn (Figure 73).

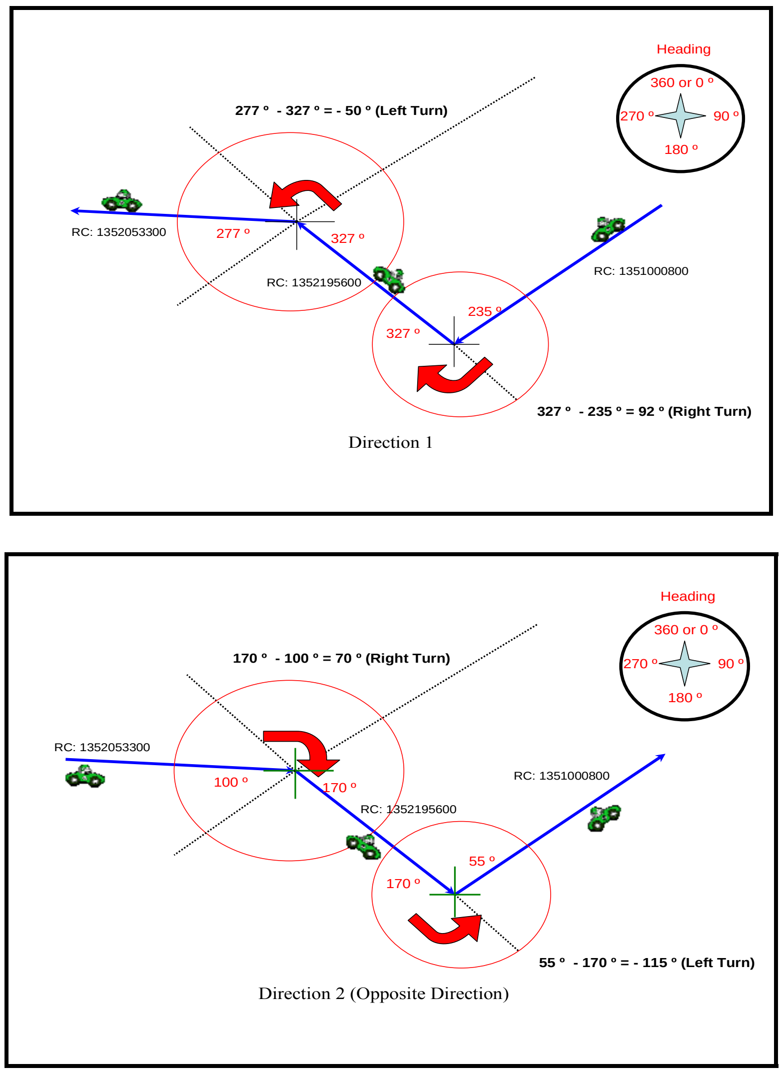


Figure 73: Turning Movement Identification

---

[19] To estimate the average of heading values, this study used heading profiles of GPS points having speeds greater than 10 mph.

Left-Turn Movement Exposure

This study first examined left turn exposures between the two driver-groups to verify if the left turn exposures of the two groups were significantly different during the study period. Figure 74 shows histograms and kernel density distributions of left turn exposures of them. Although the distribution of left turn exposures of drivers who were involved in crashes indicated a bimodal distribution and that of drivers who were not involved in crashes indicated a unimodal distribution, the Kruskal Wallis test indicated those distributions were not significantly different (p-value = 0.25).

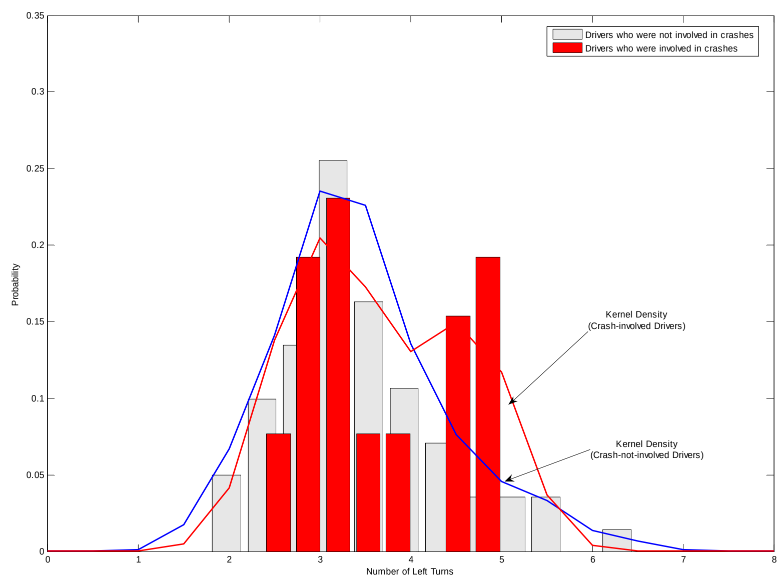
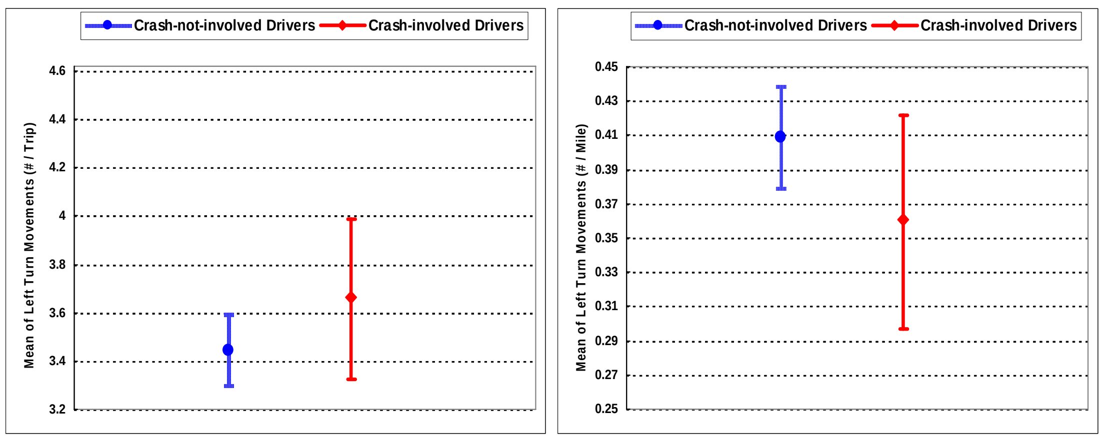


Figure 74: Distributions of Left Turn Exposures of Two Driver Groups

Using the bootstrap technique and the Wilks' lambda test, the left turn exposures per trip and per mile of drivers who were involved in crashes were not significantly different from drivers who were not involved in crashes (Figure 75 and Table 76). Nonetheless, the means of left turn exposures of drivers with and without crash involvements were 3.7 and 3.4 (the number of left turns per trip), a difference of 8 % (Table 77). Based on the unit of mile, the means of left turn exposures of drivers with and without crash involvements were 0.36 and 0.41 (the number of left turns per mile), a difference of -13 % (Table 77).



(A) Average Left Turn Exposure per Trip **(B) Average Left Turn Exposure per Mile**

Figure 75: Confidence Intervals and Distributions of Means of Left Turn Exposures

Table 76: Tests of Equality of Left Turn Movement Activities Using the Wilks' Lambda Test

Turn Movement	Wilks' Lambda	F	Sig.
Left-turn per Trip	0.992	1.317	0.253
Left-turn per Mile	0.991	1.557	0.214

* indicates a significant mean difference ($\alpha = 0.05$).

Table 77: Differences in Left Turn Exposures between the Two Groups

Normalization	Means of Left-Turn Activities		Exposure Difference	% Difference
	Drivers who <u>were not involved</u> in crashes	Drivers who <u>were involved</u> in crashes		
by trip	3.44	3.66	0.22	6
by mile	0.41	0.36	-0.05	-13

* indicates a significant mean difference ($\alpha = 0.05$).

Right-Turn Movement Exposure

In the case of right turn exposures, the distributions between the two driver-groups were significantly different (p-value = 0.05). While the mean of right turn exposures of drivers (4.2 times per trip) who were involved in crashes was larger (10 %) than drivers who were not involved in crashes (3.8 times per trip), the mode of crash-involved drivers was lower than crash-not-involved drivers (Figure 76).

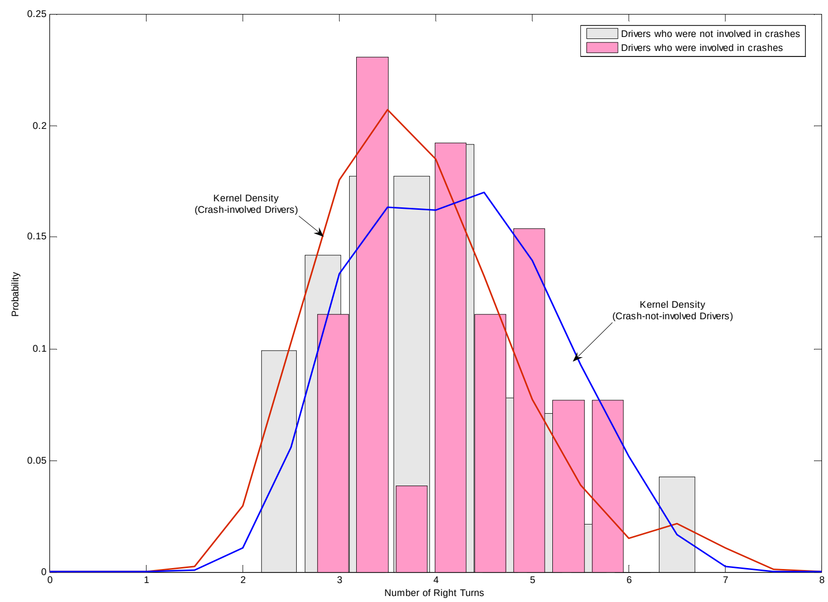
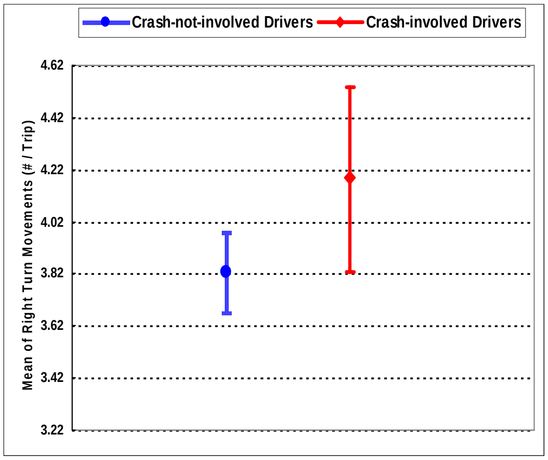
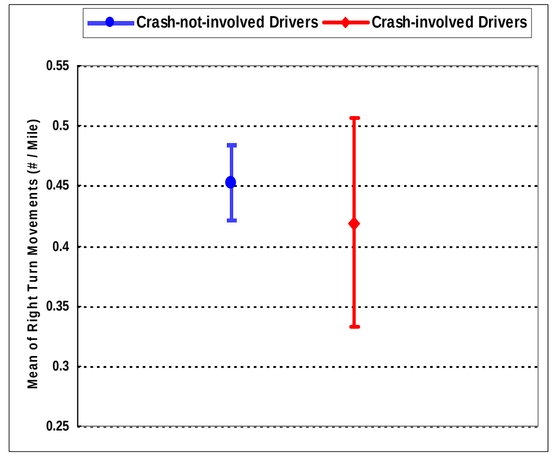


Figure 76: Distributions of Right Turn Exposures of Two Driver Groups

Similar to the left turn activities, the right turn exposures per trip and per mile of drivers who were involved in crashes were not significantly different from drivers who were not involved in crashes (Figure 77 and Table 78).





(A) Average Right Turn Exposure per Trip **(B) Average Right Turn Exposure per Mile**

Figure 77: Confidence Intervals of Means of Right Turn Exposures

Table 78: Tests of Equality of Right Turn Movement Activities Using the Wilks' Lambda Test

Turn Movement	Wilks' Lambda	F	Sig.
Right-turn per Trip	0.981	3.264	0.073
Right-turn per Mile	0.996	0.642	0.424

* indicates a significant mean difference ($\alpha = 0.05$).

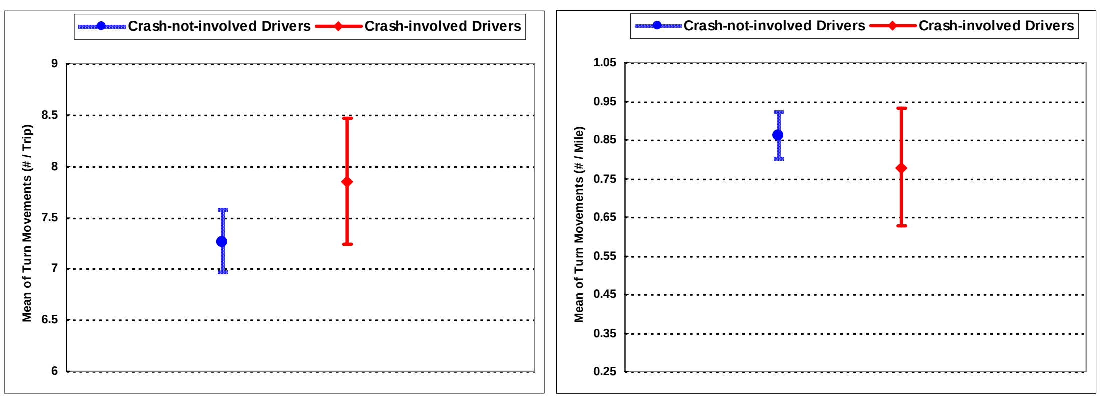
Based on the unit of mile (Table 79), the means of right turn exposures of drivers with and without crash involvements were 0.42 and 0.45 (the number of right turns per mile), a difference of -8 % (very small difference). Based on the unit of trip (Table 79), the means of right turn exposures of drivers with and without crash involvements were 4.19 and 3.83 (the number of right turns per trip), a difference of 9 %.

Table 79: Differences in Right Turn Exposures between the Two Groups

Normalization	Means of Right-Turn Activities		Exposure Difference	% Difference
	Drivers who *were not involved* in crashes	Drivers who *were involved* in crashes		
by trip	3.83	4.19	0.36	9
by mile	0.45	0.42	-0.03	-8

Left and Right-Turn Movement Exposure

The confidence test of means using the bootstrap technique and the Wilks' lambda test showed that the turn exposures per trip of drivers who were involved in crashes were not significantly different from drivers who were not involved in crashes (Figure 78). In addition, this study estimated the ratio of left and right turns to examine whether the choices of turn movements (left turn or right turn) between the two driver-groups were different or not, but this metric did not show any difference, either (Table 80).



(A) Average Turn Exposure per Trip **(B) Average Turn Exposure per Mile**

Figure 78: Confidence Intervals of Means of Turn Movement Exposures

Table 80: Tests of Equality of Turn Movement Activities Using the Wilks' Lambda Test

Turn Movement	Wilks' Lambda	F	Sig.
Turn Movement (Left & right) per Trip	0.986	2.370	0.126
Turn Movement (Left & right) per Mile	0.994	1.056	0.306
Turn movement Rate (Left/Right)	0.993	1.210	0.273

* indicates a significant mean difference ($\alpha = 0.05$).

Nonetheless, the means of turn exposures including left and right turns of drivers with and without crash involvements were 7.9 and 7.3 (number of turns per trip), a difference of 7 % (Table 81). Based on the unit of mile, the means of turn exposures of drivers with and without crash involvements were 0.8 and 0.9 (number of turns per mile), a difference of -11 % (Table 81).

Table 81: Differences in Turn Exposures between the Two Groups

Normalization	Means of Turn Movement Activities (Left + Right)		Exposure Difference	% Difference
	Drivers who *were not involved* in crashes	Drivers who *were involved* in crashes		
by trip	7.26	7.85	0.59	7
by mile	0.86	0.78	-0.08	-11

This study found that numbers of turn exposures were highly correlated with numbers of trips (0.9) (the correlation with travel mileage was 0.5). In addition, the correlation analysis (Figure 79 and Table 82) indicates that the exposures to left and right turns were highly correlated.

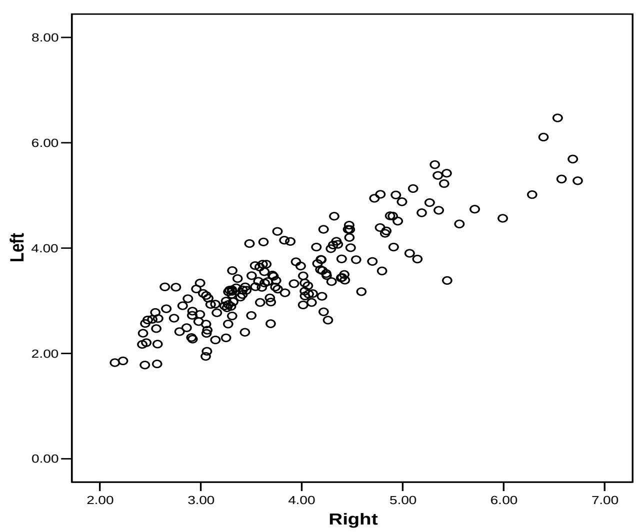


Figure 79: Results of Correlations between Left and Right Turns Exposures

Table 82: Correlation Analysis on Turn Movement Exposures

Correlation	Left Turn Exposure	Right Turn Exposure	Turning Movement Rate (Left/Right)
Left Turn Exposure	1.000	-	-
Right Turn Exposure	**0.869**	1.000	-

The linear discriminant analysis was not performed since those exposure metrics were not significantly different between the two driver-groups.

Summary of the Turning Movement Exposures

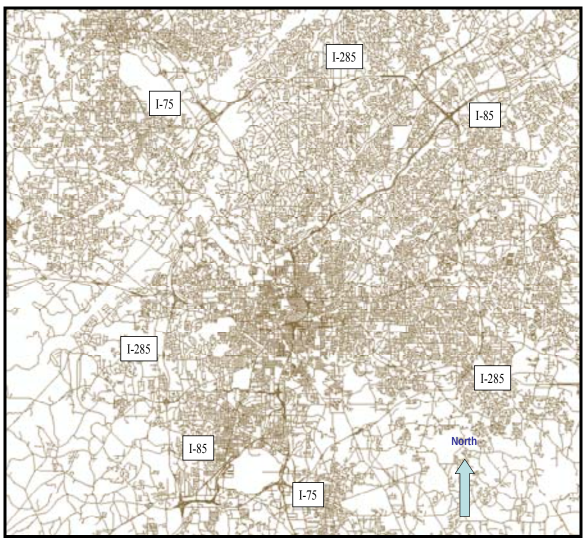
This study evaluated differences in the frequency of turning movement activities between the two driver-groups with and without crash involvements with the 6-month GPS collected trip data in order to verify whether those exposures were significantly different or not. The results are summarized as follows;

1. This study found that turn movement activities between drivers who were involved and not involved in crashes were not significantly different and that numbers of trips were highly correlated with frequency of turn movements (0.9) (the correlation with travel mileage was 0.5).
2. The average frequency of left turn activities of all drivers was 3.6 times per trip. The mean of left turn activities (3.7 times per trip) made by drivers who had crash involvements was larger (8 %) than that of drivers who did not experience any crashes (3.4 times per trip). Based on the unit of mile, the means of left turn exposures of drivers with and without crash involvements were 0.36 and 0.41 (the number of left turns per mile). However, the difference between the two driver-groups was not statistically significant.

3. The average frequency of right turning movements of all drivers was 4 times per trip. The mean of right turn activities (4.2 times per trip) made by drivers who had crash involvements was larger (10 %) than that of drivers who did not experience any crashes (3.8 times per trip). Based on the unit of trip, the means of right turn exposures of drivers with and without crash involvements were 4.19 and 3.83 (the number of right turns per trip). However, this difference between the two driver-groups was not statistically significant.
4. This study showed that the turn exposures per trip (and per mile) of drivers who were involved in crashes were not significantly different from drivers who were not involved in crashes. In addition, the ratio of left and right turns indicating the choices of turn movements (left turn or right turn) between the two driver-groups were not different or not. This result indicates that there are no differences in the choices of left or right turn activity between drivers with and without crash involvements.

Previous Crash Location Exposures

The last behavioral exposure metric evaluated in this study is a previous crash location exposure since the crash risks of individual drivers can be related with how long or how often they traveled or passed the previous crash locations (roadway segments or intersections). Figure 80 shows examples of roadway networks and locations of motor vehicle crashes that occurred on the roadway networks.



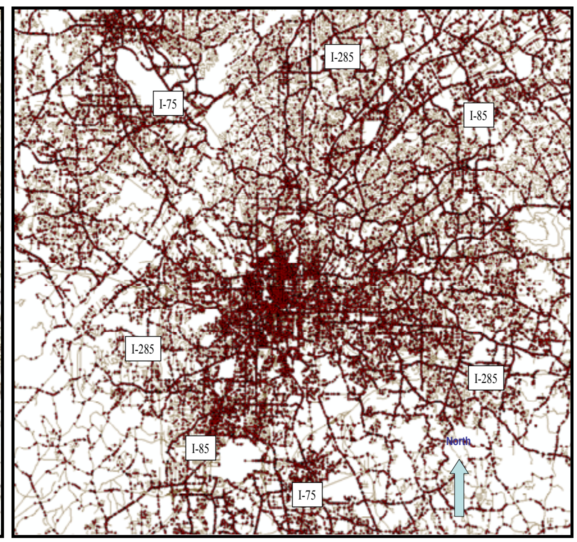


Figure 80: Roadway Network and the Corresponding Crash Locations, Atlanta

Figure 81 also shows crash locations along real travel routes and illustrates that each travel route can have different crash risks in terms of numbers of crashes occurred in the previous years. To estimate previous crash location exposure, this study counted numbers of crashes occurred between 2000 and 2002 along each travel route traveled by individual drivers using the GIS roadway characteristics (RC) and mile point information.

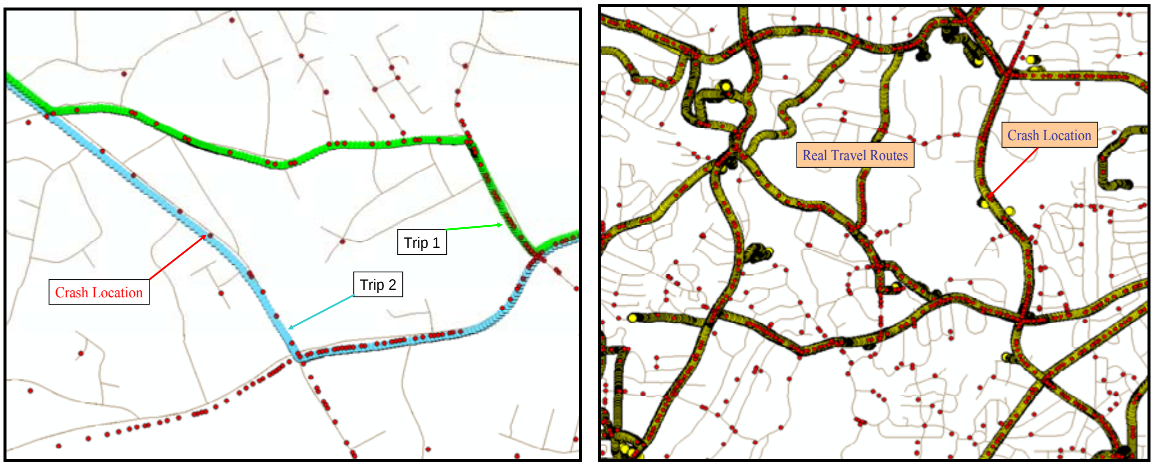


Figure 81: Real Travel Routes and Crashes Previously Occurred on Those Routes

Figure 82 shows distributions of previous crash location exposure per trip between the two driver-groups. On average, drivers who were involved in crashes had passed 309 crash locations (# of crashes along their travel routes) per trip, and drivers who were not involved in crashes had passed 166 crash locations (# of crashes along their travel routes) per trip.

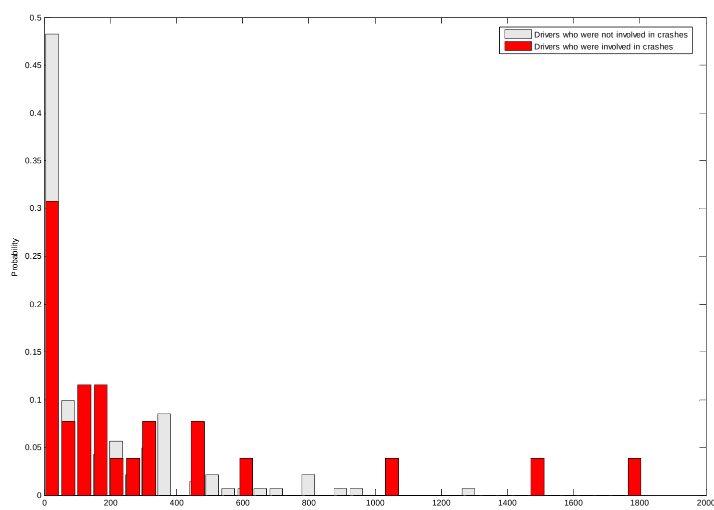
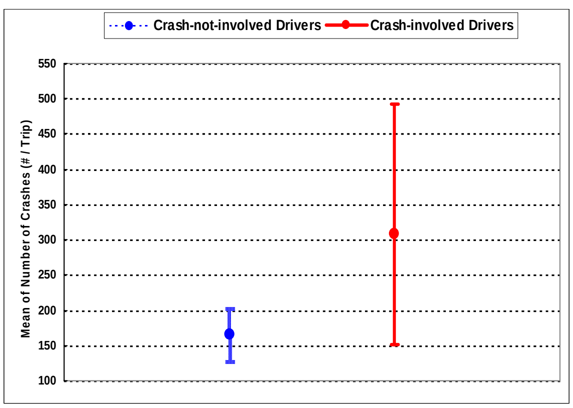
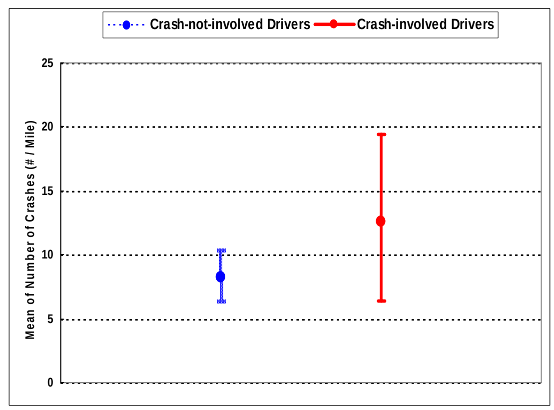


Figure 82: Distributions of Previous Crash Location Exposures between the Two Driver-Groups

Although the bootstrap technique did not show the difference in previous crash location exposure per trip between the two driver-groups, the Wilks' lambda test found that drivers who were involved in crashes had 46 % higher exposure to previous crash locations per trip than drivers who were not involved in crashes (Figure 83, Table 83, and Table 84). Based on the previous crash location exposure per mile, drivers who were involved in crashes also had 35 % higher exposure to previous crash locations, but the means test using the bootstrap technique and the Wilks' lambda test showed that the difference between them was not significant (Figure 83, Table 83, and Table 84).



(A) Average Exposure per Trip



(B) Average Exposure per Mile

Figure 83: Confidence Interval of the Means of Previous Crash Location Exposures Using the Bootstrap Technique

Table 83: Tests of Equality of Previous Crash Location Exposures Using the Wilks' Lambda Test

Exposure Metrics	Wilks' Lambda	F	Sig.
Previous Crash Location Exposures by Trip*	**0.966**	**5.822**	**0.017**
Previous Crash Location Exposures by Mile	0.984	2.692	0.103

* indicates a significant mean difference ($\alpha = 0.05$).

Table 84: Differences in Previous Crash Location Exposures between the Two Groups

Normalization	Means of Previous Crash Location Exposures		Exposure Difference	% Difference
	Drivers who *were not involved* in crashes	Drivers who *were involved* in crashes		
by trip *	**166**	**309**	**143**	**46**
by mile	8	13	4	35

* indicates a significant mean difference ($\alpha = 0.05$).

Weighted Previous Crash Location Exposure

Since the previous crash location exposure in above did not consider a "dwelling time" on the crash locations previously occurred, which implies how long a driver stays on the each crash location, this study tried to weight previous crash location exposure with the dwelling time. Figure 84 shows the distributions of weighted previous crash location exposure per trip between the two driver-groups.

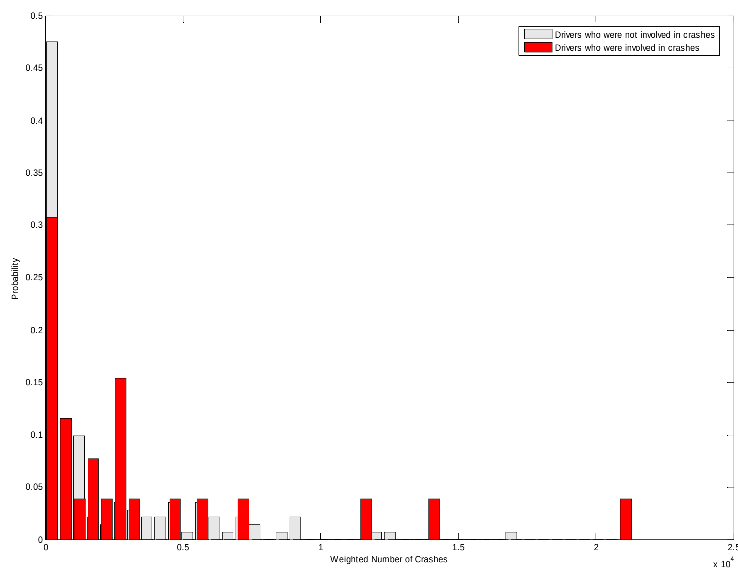
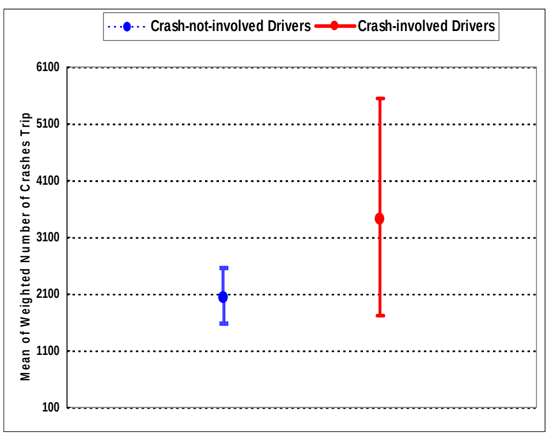
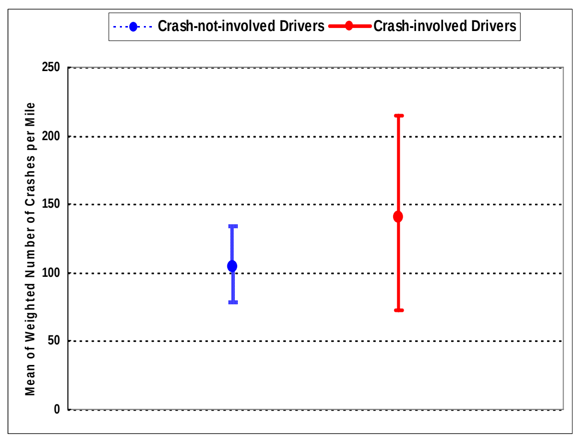


Figure 84: Distributions of Weighted Previous Crash Location Exposures between the Two Driver-Groups

On average, drivers who were involved in crashes obtained 3,412 (# of crashes at the certain roadway * dwelling time at that roadway) along their travel routes, and drivers who were not involved in crashes passed 2,044 (# of crashes at the certain roadway * dwelling time at that roadway) along their travel routes, a difference of 40 %. Based on the weighted previous crash location exposure per mile, drivers who were involved in crashes also had 26 % higher exposure to previous crash locations, but the means test using the bootstrap technique and the Wilks' lambda test showed that the difference between them was not significant (Figure 85, Table 85, and Table 86).



(A) Average Exposure per Trip



(B) Average Exposure per Mile

Figure 85: Confidence Interval of Means of Weighted Previous Crash Location Exposures between the Two Driver-Groups

Table 85: Tests of Equality of Weighted Previous Crash Location Exposures Using the Wilks' Lambda Test

Exposure Metrics	Wilks' Lambda	F	Sig.
Weighted Previous Crash Location Exposures by Trip	0.979	3.47	0.064
Weighted Previous Crash Location Exposures by Mile	0.994	0.985	0.322

* indicates a significant mean difference ($\alpha = 0.05$).

Table 86: Differences in Weighed Previous Crash Location Exposures between the Two Groups

Normalization	Means of Weighted Previous Crash Location Exposures		Exposure Difference	% Difference
	Drivers who *were not involved* in crashes	Drivers who *were involved* in crashes		
by trip	2,044	3,412	1,368	40
by mile	104	140	37	26

* indicates a significant mean difference ($\alpha = 0.05$).

The correlation analysis (Table 87) indicated that those two exposures, un-weighted previous crash location exposure and weighted previous crash location exposure, were highly correlated.

Table 87: Correlation Analysis on Previous Crash Location Exposure Metrics

Correlation	Previous Crash Location Exposures per Trip	Previous Crash Location Exposures per Mile	Weighted Previous Crash Location Exposures per Trip	Weighted Previous Crash Location Exposures per Mile
Previous Crash Location Exposures per Trip	1.000			
Previous Crash Location Exposures per Mile	**0.953**	1.000		
Weighted Previous Crash Location Exposures per Trip	**0.839**	**0.815**	1.000	
Weighted Previous Crash Location Exposures per Mile	**0.734**	**0.795**	**0.944**	1.000

Although the previous crash location exposure per trip provided the significant difference between the two driver-groups, this metric may be unstable because number of trips is strongly correlated with travel mileage. In addition, the unit of trip may not clearly support individual drivers who have different travel mileage per trip.

Summary of the Previous Crash Location Exposures

This study evaluated differences in numbers of previous crash location exposure between the two driver-groups with and without crash involvements with the 6-month GPS collected trip data in order to verify whether those exposures were significantly different or not. The results on the previous crash location exposure metrics are summarized as follows:

1. Based on the previous crash location exposure per mile, drivers who were involved in crashes had 35 % higher exposure to previous crash locations, but the means test using the bootstrap technique and the Wilks' lambda test showed that the difference between them was not significant. The previous crash location exposure per trip by drivers who had crash involvements during the 14-months period were also larger (46 %) than those of drivers who did not experience any crashes during the same period. Although the Wilks' lambda test showed the difference between them at the 0.05 significance level, this metric may not be appropriate for using classification process because the unit of trip is not independent from travel mileage.
2. The average weighted previous crash location exposure (# of crashes previously occurred at the roadway multiplied by dwelling times at that roadway) made by drivers (3,142 per trip) who had crash involvements were larger (40 %) than those of drivers who did not experience any crashes (2,044 per trip). Based on the weighted previous crash location exposure per mile, drivers who were involved in crashes also had 26 % higher exposure to

previous crash location exposure, but the means test using the bootstrap technique and the Wilks' lambda test showed that the difference between them was not significant.

3. Drivers who were involved in crashes more frequently tended to travel along previous crash locations (roadways and intersections) than drivers who were not involved in crashes, but this result may be a simple correlation to higher mileage traveled.
4. Although this study could evaluate only previous crash location exposures of individual drivers based on the crash frequency data due to the unavailability of annual average daily traffic (AADT) and total vehicle mile traveled (VMT) data on each facility type, further studies need to be performed to evaluate hazardous roadway exposure based on normalized crash rate.
5. Since this study showed that speed- and acceleration-related behavioral exposure metrics have positive relationships with the crash involvements and that drivers having crash involvements tend to more frequently travel at previous crash locations based on the trip-based previous crash location exposure, this study support the argument that those behavioral crash exposure metrics can be employed as the safety surrogate measures to select hazardous locations (roadways or intersections).

Chapter Eleven

MODELING PROCESS WITH COMBINED EXPOSURES

Linear Disrciminant Analysis (LDA)

From the previous chapters, this study found that 32 driving behavior activity exposure metrics could help classify high crash-involvement drivers from the driver population. However, since the previous chapters separately analyzed each behavioral exposure metric such as travel mileage, duration, speed, and acceleration, this chapter further investigated relationships among all 32 behavioral exposure metrics selected from the previous chapters. Although this study found that other activity metrics such as unfamiliar roadway exposure, left/right turn exposure, and previous crash location exposure were not significantly different between the two driver-groups, this study included those activity metrics into the modeling process because potential interactions among driving behavior activity metrics of individual drivers may be still possible (or relative importance of individual metrics in classification process may be changeable). Thus, this study totally employed 35 potential behavioral exposure metrics.

As mentioned in earlier chapters, this study investigates correlations between these selected exposure metrics since the correlation between variables statistically decreases the explanatory power of each variable as well as degrades the efficiency of the statistical model. In addition, the linear discriminant analysis requires avoiding correlations between variables before performing the modeling process. Thus, the correlation analysis with the 35 behavioral exposure metrics was conducted, and secondary variables were removed from the set of independent variables to exclude any impacts of correlated variables on the final model.

Modeling Process for All Drivers Using the Linear Discriminant Analysis

Furthermore, because the current insurance premium structures employ the information on demographic characteristics, this study examined the relationships of the gender and age with crash involvement rates and found that gender and age did not have any strong relationships with the crash involvements within the sample of this study (p-value: 0.619 (gender) p-value: 0.869 (age)). However, this study does not claim that demographic characteristics do not have relationships with crash involvement rate due to the relatively small sample data used in this study and suggests that driving behavior activity patterns may be one of possible information to classify drivers into different risk groups, instead. In addition, this study believes that fatality or severe injury rates have strong relationships with gender and age characteristics, but this study could not evaluate relationships between them due to the unavailability of information regarding crash type and severity.

Although there were no strong relationships between crash involvement rate and gender and age based on the driver sample in this study, interactions between driving behavior activity metrics and gender or age may be possible and those interactions can affect to the performance of classification process.

Because the linear discriminant analysis (LDA) requires only continuous independent variables, this study separately performs the modeling process by gender (male and female). For the impact of age (including also gender) on the classification, the logistics regression model and classification and regression tree analysis (CART) are utilized later in this chapter as discrete variables. Thus, this study developed three discriminant models using the linear discriminant analysis (LDA) for all drivers, male drivers only, and female drivers only, and this modeling process can help assess which of driving behavior activities related to crash involvement rate are relatively different between male and female drivers.

The result of correlation analysis for all drivers is shown at the Appendix B. At the same time, this study estimated structure coefficients (loading power or discriminating power) to assess the power of discrimination of each exposure metric and help select the most explainable variable among correlated variables. Based on the correlation result and the structure coefficient matrix, this study excluded all correlated variables based on the rank of structure coefficients.

The Stepwise Linear Discriminant Analysis for All Drivers

Table 88 shows the structure coefficients using 35 metrics for all drivers and selected exposure metrics based on correlation of variables (Appendix B), and Table 89 shows the structure coefficient matrix using 27 driving behavior activity metrics after removing correlation impacts.

Table 88: Variable Selection based on Structure Coefficient Matrix and Correlation Analysis for All Drivers

Exposure	Measure	Threshold	Facility/Time	Structure Coefficient	Rank	Variable Selection
Travel Duration	**Total**	**-**	**All**	**0.340**	**1**	√
Deceleration	**Frequency**	**4 mph/s**	**Local/Night**	**0.327**	**2**	√
Travel Duration	Total	-	Outside/Afternoon	0.322	3	-
Speed	**Frequency**	**15 mph**	**Arterial/Morning**	**0.321**	**4**	√
Deceleration	**Frequency**	**4 mph/s**	**Freeway/Morning**	**0.317**	**5**	√
Travel Mileage	**Total**	**-**	**Outside/Afternoon**	**0.281**	**6**	√
Deceleration	**Frequency**	**4 mph/s**	**Arterial/Morning**	**0.279**	**7**	√
Deceleration	**Frequency**	**8 mph/s**	**Freeway/Afternoon**	**0.271**	**8**	√
Travel Duration	**Total**	**-**	**Outside/AM Peak**	**0.270**	**9**	√
Speed	**Delta**	**-**	**Arterial/AM Peak**	**0.263**	**10**	√
Travel Mileage	Total	-	All	0.242	11	-
Acceleration	Frequency	10 mph/s	Freeway/Morning	0.233	12	-
Speed	**Frequency**	**10 mph**	**Freeway/Morning**	**0.223**	**13**	√
Acceleration	**Mean**	**-**	**Freeway/Afternoon**	**0.221**	**14**	√
Speed	**Frequency**	**10 mph**	**Local/Morning**	**0.214**	**15**	√
Unfamiliar Roadway	**-**	**-**	**-**	**-0.205**	**16**	√
Travel Duration	**Total**	**-**	**Freeway/PM Peak**	**0.203**	**17**	√
Previous Crash Location Exposure	**-**	**-**	**-**	**0.203**	**18**	√
Travel Mileage	**Total**	**-**	**Arterial/PM Peak**	**0.201**	**19**	√
Travel Mileage	Total	-	Freeway/PM Peak	0.198	20	-
Travel Duration	Total	-	Arterial/PM Peak	0.192	21	-
Speed	**Positive Delta**	**-**	**Arterial/AM Peak**	**0.186**	**22**	√
Speed	**Positive Delta**	**-**	**Freeway/PM Peak**	**0.185**	**23**	√
Travel Duration	**Total**	**-**	**Arterial/AM Peak**	**0.182**	**24**	√
Speed	**Mean**	**-**	**Freeway/Night**	**0.181**	**25**	√
Speed	**Frequency**	**20 mph**	**Arterial/Night**	**0.179**	**26**	√
Travel Duration	**Total**	**-**	**Outside/Night**	**0.145**	**27**	√
Travel Mileage	Total	-	Outside/AM Peak	0.144	28	-
Speed	**Mean**	**-**	**Freeway/Morning**	**0.131**	**29**	√
Speed	**Positive Delta**	**-**	**Local/AM Peak**	**0.102**	**30**	√
Travel Mileage	**Total**	**-**	**Freeway/Night**	**0.096**	**31**	√
Speed	**Positive Delta**	**-**	**Local/Night**	**0.095**	**32**	√
Travel Duration	Total	-	Freeway/Night	0.093	33	-
Turn Movement	**-**	**-**	**-**	**-0.090**	**34**	√

Table 89: Structure Coefficient Matrix of Variables after Removing Correlated Variables for All Drivers

Exposure	Measure	Threshold	Facility/Time	Structure Coefficient
Deceleration	Frequency	4 mph/s	Arterial/Morning	0.426
Deceleration	Frequency	4 mph/s	Local/Night	0.383
Speed	Frequency	15 mph	Arterial/Morning	0.376
Travel Duration	Total	-	All	0.373
Deceleration	Frequency	4 mph/s	Freeway/Morning	0.372
Travel Mileage	Total	-	Outside/Afternoon	0.330
Travel Duration	Total	-	Outside/AM Peak	0.316
Speed	Frequency	10 mph	Local/Morning	0.314
Unfamiliar Roadway	-	-	-	-0.301
Speed	Mean	-	Freeway/Morning	0.290
Travel Duration	Total	-	Freeway/PM Peak	0.274
Speed	Frequency	10 mph	Freeway/Morning	0.245
Speed	Positive Delta	-	Local/Night	0.241
Previous Crash Location Exposure	-	-	-	0.238
Travel Duration	Total	-	Outside/Night	0.203
Speed	Mean	-	Freeway/Night	0.202
Travel Duration	Total	-	Arterial/AM Peak	0.195
Speed	Positive Delta	-	Freeway/PM Peak	0.174
Speed	Frequency	20 mph	Arterial/Night	0.172
Deceleration	Frequency	8 mph/s	Freeway/Afternoon	0.157
Speed	Positive Delta	-	Local/AM Peak	0.150
Turn Movement	-	-	-	-0.134
Travel Mileage	Total	-	Arterial/PM Peak	0.127
Speed	Positive Delta	-	Arterial/AM Peak	0.115
Travel Mileage	Total	-	Freeway/Night	0.096
Acceleration	Mean	-	Freeway/Afternoon	0.081
Speed	Delta	-	Arterial/AM Peak	0.046

For the analysis of all drivers, the canonical correlation[20] of the final model was 0.7. As a result, 87.4 % of drivers who were not involved in crashes and 68.2 % of drivers who were involved in crashes were correctly classified. Overall performance of the model using this metric was 84.0 % (Table 90).

[20] The canonical correlation is a similar measure to the R-square in regression model.

Table 90: Classification Results Using All Selected Exposure Metrics for All Drivers

Crash Involvements	Predicted Group Membership	
	Drivers who were not involved in crashes	Drivers who were involved in crashes
Drivers who were not involved in crashes	87.4 %	12.6 %
Drivers who were involved in crashes	31.8 %	68.2 %

However, this study need to state that this modeling result may be over-specified due to the small sample size of crash-involved drivers (26 drivers) and relatively large numbers of explanatory variables (27 variables). In other words, each of 26 crash-involved drivers in the sample of this study might be classified by only specific exposure metrics. In addition, it is practically difficult to collect all relevant exposure metrics. Thus, this study utilized the stepwise linear discriminant analysis similar to the stepwise multiple regression model.

From the stepwise linear discriminant analysis using the forward approach, this study selects a subset of independent variables (driving behavior activity exposure metrics) explaining the most variations in the dependent variables (crash involvements). The process of variable selection using the stepwise method is shown in Appendix C. Table 91 shows the independent variables selected by the stepwise discriminant analysis and their structure coefficients. Based on the 0.05 and 0.1 significance levels, five and seven exposure metrics were selected, respectively.

Table 92 shows the performance of classification process by the linear discriminant analysis based on two significantly levels (0.05 and 0.1). Based on the 0.05 significance level, the overall performance with the five selected-variables was 81.6 %.

Table 91: Variable Selection and Structure Coefficient Matrix from the Stepwise Linear Discriminant Analysis for All Drivers

(A) 0.05 Significance Level

Exposure	Measure	Threshold	Facility/Time	Structure Coefficient
Speed	Frequency	15 mph	Arterial/Morning	0.493
Travel Mileage	Total	-	Outside/Afternoon	0.429
Deceleration	Frequency	4 mph/s	Freeway/Morning	0.395
Travel Duration	Total	-	Outside/AM Peak	0.306
Previous Crash Location Exposure	-	-	-	0.237

(B) 0.1 Significance Level

Exposure	Measure	Threshold	Facility/Time	Structure Coefficient
Travel Mileage	Total	-	Outside/Afternoon	0.383
Speed	Delta	-	Arterial/AM Peak	0.379
Speed	Frequency	15 mph	Arterial/Morning	0.375
Deceleration	Frequency	4 mph/s	Freeway/Morning	0.339
Travel Duration	Total	-	Outside/AM Peak	0.309
Deceleration	Frequency	4 mph/s	Local/Night	0.296
Previous Crash Location Exposure	-	-	-	0.202

Table 92: Final Classification Results Using a Subset of Selected Exposure Metrics for All Drivers

(A) 0.05 Significance Level

Crash Involvements	Predicted Group Membership	
	Drivers who were not involved in crashes	Drivers who were involved in crashes
Drivers who were not involved in crashes	84.7	15.3
Drivers who were involved in crashes	34.6	65.4

(B) 0.1 Significance Level

Crash Involvements	Predicted Group Membership	
	Drivers who were not involved in crashes	Drivers who were involved in crashes
Drivers who were not involved in crashes	89.9	10.1
Drivers who were involved in crashes	32.0	68.0

Figure 86 illustrates the difference in performance of discriminant analyses by a set of different exposure metrics, indicating the classification using the only five-exposure-metrics may provide useful information on crash involvement rate of individual drivers.

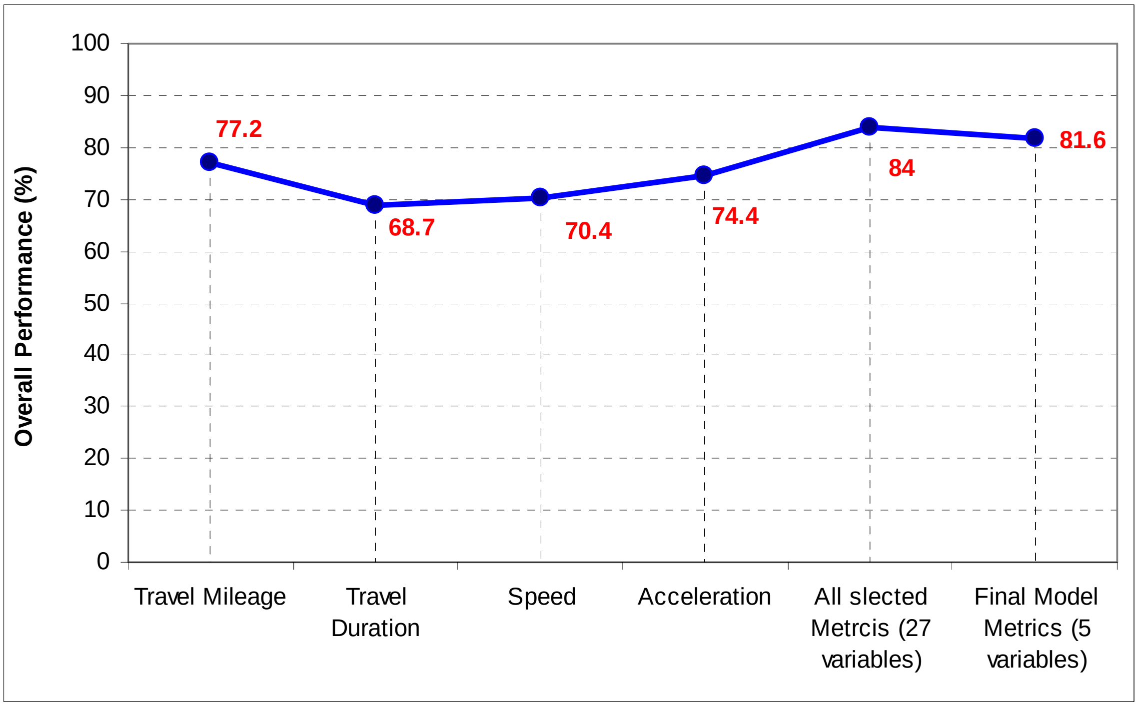


Figure 86: Overall Classification Performance by Different Behavior Metrics

Modeling Process for Male Drivers Using the Linear Discriminant Analysis

Similar to the classification process for all drivers, this study performed the same process for male drivers only. Table 93 shows the structure coefficients using 35 metrics for all drivers and selected exposure metrics based on correlation of variables (Appendix B), and Table 94 shows the structure coefficient matrix using the remained 22 exposure metrics after removing correlation impacts.

Table 93: Variable Selection based on Structure Coefficient Matrix and Correlation Analysis for Male Drivers

Exposure	Measure	Threshold	Facility/Time	Structure Coefficient	Rank	Variable Selection
Speed	**Frequency**	**20 mph**	**Arterial/Night**	**0.272**	**1**	√
Previous Crash Location	**-**	**-**	**-**	**0.250**	**2**	√
Deceleration	**Frequency**	**8 mph/s**	**Freeway/Afternoon**	**0.217**	**3**	√
Acceleration	**Frequency**	**10 mph/s**	**Freeway/Morning**	**0.210**	**4**	√
Speed	**Frequency**	**10 mph**	**Freeway/Morning**	**0.207**	**5**	√
Deceleration	Frequency	4 mph/s	Freeway/Morning	0.204	6	-
Speed	Frequency	15 mph	Arterial/Morning	0.183	7	-
Speed	**Delta**	**-**	**Arterial/AM Peak**	**0.166**	**8**	√
Deceleration	**Frequency**	**4 mph/s**	**Local/Night**	**0.161**	**9**	√
Speed	**Mean**	**-**	**Freeway/Morning**	**0.148**	**10**	√
Travel Mileage	**Total**	**-**	**Outside/AM Peak**	**0.134**	**11**	√
Speed	**Positive Delta**	**-**	**Arterial/AM Peak**	**0.110**	**12**	√
Speed	Frequency	10 mph	Local/Morning	0.107	13	-
Unfamiliar Roadway	**-**	**-**	**-**	**-0.104**	**14**	√
Travel Mileage	**Total**	**-**	**Freeway/Night**	**0.097**	**15**	√
Travel Duration	**Total**	**-**	**Arterial/AM Peak**	**0.094**	**16**	√
Travel Duration	Total	-	Freeway/Night	0.093	17	-
Travel Mileage	Total	-	Outside/Afternoon	0.084	18	-
Speed	Positive Delta	-	Freeway/PM Peak	0.082	19	-
Travel Mileage	**Total**	**-**	**Freeway/PM Peak**	**0.082**	**20**	√
Deceleration	Frequency	4 mph/s	Arterial/Morning	0.078	21	-
Travel Duration	Total	-	Freeway/PM Peak	0.077	22	-
Turn Movement	**-**	**-**	**-**	**-0.073**	**23**	√
Travel Duration	Total	-	Outside/AM Peak	0.071	24	-
Travel Mileage	**Total**	**-**	**Arterial/PM Peak**	**0.061**	**25**	√
Travel Mileage	**Total**	**-**	**All**	**0.058**	**26**	√
Speed	**Positive Delta**	**-**	**Local/Night**	**0.049**	**27**	√
Acceleration	**Mean**	**-**	**Freeway/Afternoon**	**0.044**	**28**	√
Speed	**Mean**	**-**	**Freeway/Night**	**0.038**	**29**	√
Travel Duration	**Total**	**-**	**Outside/Afternoon**	**0.037**	**30**	√
Travel Duration	Total	-	All	0.018	31	-
Travel Duration	**Total**	**-**	**Outside/Night**	**0.012**	**32**	√
Travel Duration	Total	-	Arterial/PM Peak	0.004	33	-
Speed	Positive Delta	-	Local/AM peak	0.001	34	-

Table 94: Structure Coefficient Matrix of Variables after Removing Correlated Variables for Male Drivers

Exposure	Measure	Threshold	Facility/Time	Structure Coefficient
Speed	Frequency	20 mph	Arterial/Night	0.413
Previous Crash Location Exposure	-	-	-	0.351
Acceleration	Frequency	10 mph/s	Freeway/Morning	0.331
Travel Mileage	Total	-	All	0.331
Deceleration	Frequency	8 mph/s	Freeway/Afternoon	0.331
Unfamiliar Roadway	-	-		-0.322
Speed	Frequency	10 mph	Freeway/Morning	0.304
Speed	Positive Delta	-	Local/Night	0.304
Speed	Mean	-	Freeway/Morning	0.269
Travel Mileage	Total	-	Freeway/PM Peak	0.255
Speed	Mean	-	Freeway/Night	0.237
Travel Mileage	Total	-	Freeway/Night	0.235
Travel Mileage	Total	-	Outside/AM Peak	0.206
Acceleration	Mean	-	Freeway/Afternoon	0.166
Speed	Delta	-	Arterial/AM Peak	0.152
Deceleration	Frequency	4 mph/s	Local/Night	0.144
Speed	Positive Delta	-	Arterial/AM Peak	0.090
Turn Movement	-	-	-	0.089
Travel Mileage	Total	-	Arterial/PM Peak	-0.084
Travel Duration	Total	-	Outside/Afternoon	0.065
Travel Duration	Total	-	Arterial/AM Peak	0.035
Travel Duration	Total	-	Outside/Night	0.012

For the analysis of all male drivers, the canonical correlation of the final model was 0.73. As a result, 95.8 % of drivers who were not involved in crashes and 90.0 % of drivers who were involved in crashes were correctly classified. Overall performance of the model using this metric was 94.8 %.

Table 95: Classification Results Using All Selected Exposure Metrics for Male Drivers

Crash Involvements	Predicted Group Membership	
	Drivers who were not involved in crashes	Drivers who were involved in crashes
Drivers who were not involved in crashes	95.8 %	4.2 %
Drivers who were involved in crashes	10.0 %	90.0 %

The Stepwise Linear Discriminant Analysis for Male Drivers

Due to the same possible limitations to the classification for all drivers in above, this study selected a subset of independent variables (driving behavior activity exposure metrics) explaining the most variations in the dependent variables (crash involvements) from the 22 potential behavior exposures. Table 96 shows the independent variables selected by the stepwise discriminant analysis and their structure coefficients. Based on the 0.05 significance level, four exposure metrics were selected.

Table 97 shows that the overall performance of classification process for male drivers only with the four selected-variables was 88.7 %. This study also performed the discriminant analysis based on the 0.1 significance level, but the result was the same as that of 0.05 significance level.

Table 96: Variable Selection and Structure Coefficient Matrix from the Stepwise Linear Discriminant Analysis for Male Drivers

Exposure	Measure	Threshold	Facility/Time	Structure Coefficient
Speed	Frequency	20 mph	Arterial/Night	0.407
Travel Mileage	Total	-	Outside/AM Peak	0.372
Hazardous Roadway	-	-	-	0.367
Deceleration	Frequency	8 mph/s	Freeway/Afternoon	0.284

Table 97: Final Classification Results Using a Subset of Selected Exposure Metrics for Male Drivers Only

Crash Involvements	Predicted Group Membership	
	Drivers who *were not involved* in crashes	Drivers who *were involved* in crashes
Drivers who *were not involved* in crashes	93.1	6.9
Drivers who *were involved* in crashes	30.8	69.2

Modeling Process for Female Drivers Using the Linear Discriminant Analysis

This study also performed the same process for female drivers only. Table 98 shows the structure coefficients using 35 metrics for all drivers and selected exposure metrics based on correlation of variables (Appendix B), and Table 99 shows the structure coefficient matrix using the remained 23 exposure metrics after removing correlation impacts.

Table 98: Variable Selection based on Structure Coefficient Matrix and Correlation Analysis for Female Drivers

Exposure	Measure	Threshold	Facility/Time	Structure Coefficient	Rank	Variable Selection
Travel Duration	**Total**	**-**	**All**	**0.397**	**1**	√
Travel Duration	**Total**	**-**	**Outside/Afternoon**	**0.340**	**2**	√
Speed	**Frequency**	**15 mph**	**Arterial/Morning**	**0.236**	**3**	√
Travel Mileage	Total	-	Outside/Afternoon	0.227	4	-
Travel Duration	**Total**	**-**	**Outside/AM Peak**	**0.226**	**5**	√
Travel Mileage	Total	-	All	0.222	6	-
Deceleration	**Frequency**	**4 mph/s**	**Arterial/Morning**	**0.222**	**7**	√
Deceleration	**Frequency**	**4 mph/s**	**Local/Night**	**0.214**	**8**	√
Travel Duration	**Total**	**-**	**Arterial/PM Peak**	**0.213**	**9**	√
Acceleration	**Mean**	**-**	**Freeway/Afternoon**	**0.186**	**10**	√
Travel Mileage	Total	-	Arterial/PM Peak	0.176	11	-
Deceleration	Frequency	4 mph/s	Freeway/Morning	0.167	12	-
Speed	**Mean**	**-**	**Freeway/Night**	**0.162**	**13**	√
Travel Duration	**Total**	**-**	**Freeway/PM Peak**	**0.159**	**14**	√
Travel Mileage	Total	-	Freeway/PM Peak	0.143	15	-
Speed	**Frequency**	**10 mph**	**Local/Morning**	**0.140**	**16**	√
Speed	**Delta**	**-**	**Arterial/AM Peak**	**0.140**	**17**	√
Unfamiliar Roadway	**-**	**-**	**-**	**-0.139**	**18**	√
Travel Duration	Total	-	Outside/Night	0.131	19	-
Speed	**Positive Delta**	**-**	**Freeway/PM Peak**	**0.129**	**20**	√
Speed	**Positive Delta**	**-**	**Local/AM Peak**	**0.125**	**21**	√
Travel Duration	Total	-	Arterial/AM Peak	0.124	22	-
Deceleration	**Frequency**	**8 mph/s**	**Freeway/Afternoon**	**0.114**	**23**	√
Speed	**Positive Delta**	**-**	**Arterial/AM Peak**	**0.111**	**24**	√
Speed	**Frequency**	**10 mph**	**Freeway/Morning**	**0.062**	**25**	√
Speed	Positive Delta	-	Local/Night	0.061	26	-
Travel Mileage	Total	-	Outside/AM Peak	0.051	27	-
Speed	**Frequency**	**20 mph**	**Arterial/Night**	**0.039**	**28**	√
Turn Movement	**-**	**-**	**-**	**-0.034**	**29**	√
Speed	Mean	-	Freeway/Morning	0.023	30	-
Travel Duration	**Total**	**-**	**Freeway/Night**	**0.021**	**31**	√
Travel Mileage	Total	-	Freeway/Night	0.021	32	-
Previous Crash Location	**-**	**-**	**-**	**-0.015**	**33**	√
Acceleration	**Frequency**	**10 mph/s**	**Freeway/Morning**	**0.001**	**34**	√

Table 99: Structure Coefficient Matrix of Variables after Removing Correlated Variables for Female Drivers

Exposure	Measure	Threshold	Facility/Time	Structure Coefficient
Travel Duration	Total	-	All	0.695
Travel Duration	Total	-	Outside/Afternoon	0.594
Deceleration	Frequency	4 mph/s	Local/Night	0.374
Deceleration	Frequency	4 mph/s	Arterial/Morning	0.283
Travel Duration	Total	-	Arterial/PM Peak	0.281
Travel Duration	Total	-	Outside/AM Peak	0.271
Travel Duration	Total	-	Freeway/PM Peak	0.248
Unfamiliar Roadway	-	-	-	-0.245
Travel Duration	Total	-	Freeway/Night	0.205
Acceleration	Frequency	10 mph/s	Freeway/Morning	0.203
Speed	Positive Delta	-	Freeway/PM Peak	0.192
Speed	Delta	-	Arterial/AM Peak	0.155
Acceleration	Mean	-	Freeway/Afternoon	0.142
Speed	Frequency	15 mph	Arterial/Morning	0.135
Speed	Frequency	10 mph	Local/Morning	0.091
Speed	Positive Delta	-	Arterial/AM Peak	0.072
Speed	Mean	-	Freeway/Night	0.064
Speed	Frequency	20 mph	Arterial/Night	0.061
Speed	Positive Delta	-	Local/AM Peak	0.057
Previous Crash Location Exposure	-	-	-	-0.051
Speed	Frequency	10 mph	Freeway/Morning	0.044
Deceleration	Frequency	8 mph/s	Freeway/Afternoon	0.037
Turn Movement	-	-	-	-0.029

For the analysis of all female drivers, the canonical correlation of the final model was 0.75. As a result, 94.6 % of drivers who were not involved in crashes and 83.3 % of drivers who were involved in crashes were correctly classified. Overall performance of the model using this metric was 92.6 %.

Table 100: Final Classification Results Using All Selected Exposure Metrics for Female Drivers

Crash Involvements	Predicted Group Membership	
	Drivers who were not involved in crashes	Drivers who were involved in crashes
Drivers who were not involved in crashes	94.6 %	5.4 %
Drivers who were involved in crashes	16.7 %	83.3 %

The Stepwise Linear Discriminant Analysis for Female Drivers

Due to the same possible limitations to the classification for all drivers and male driver only in above, this study also selected a subset of independent variables from the 23 potential behavior exposures. Table 101 shows the independent variables selected by the stepwise discriminant analysis and their structure coefficients. Based on the 0.05 and 0.1 significance levels, three and five exposure metrics were selected, respectively.

Table 101: Variable Selection Using the Stepwise Linear Discriminant Analysis for Female Drivers

(A) 0.05 Significance Level

Exposure	Measure	Threshold	Facility/Time	Structure Coefficient
Travel Duration	Total	-	All	0.862
Travel Duration	Total	-	Outside/Afternoon	0.707
Deceleration	Frequency	4 mph/s	Local/Night	0.510

(B) 0.1 Significance Level

Exposure	Measure	Threshold	Facility/Time	Structure Coefficient
Travel Duration	Total	-	All	0.722
Travel Duration	Total	-	Outside/Afternoon	0.628
Deceleration	Frequency	4 mph/s	Local/Night	0.357
Speed	Mean	-	Freeway/Night	0.313
Speed	Delta	-	Arterial/AM Peak	0.267

Table 102 shows that the overall performances of classification process for female drivers with the three selected-variables (α = 0.05) was 80.2 % and with the five selected variables (α = 0.05) was 87.7 %.

Table 102: Final Classification Results Using a Subset of Selected Exposure Metrics for Female Drivers Only

(A) 0.05 Significance Level

Crash Involvements	Predicted Group Membership	
	Drivers who *were not involved* in crashes	Drivers who *were involved* in crashes
Drivers who *were not involved* in crashes	79.5	20.5
Drivers who *were involved* in crashes	15.4	84.6

(B) 0.1 Significance Level

Crash Involvements	Predicted Group Membership	
	Drivers who *were not involved* in crashes	Drivers who *were involved* in crashes
Drivers who *were not involved* in crashes	88.5	11.5
Drivers who *were involved* in crashes	16.7	83.3

Crash History and Speed Citation Record

This study also investigated whether the crash and citation histories could explain potential crash involvements (The crash and citation histories are also employed for the current insurance premium estimates). Based on the Wilks' lambda test (Table 103), the numbers of crash involvements in previous years were significantly different between the two driver-groups at the 0.05 significance level, which indicates drivers who had crash involvements in previous had higher possibility of being involved in crashes again.

The number of tickets received in the last 5 years was significantly different between the two driver-groups at the 0.1 significance level, implying drivers having more citation histories had more changes to being involved in crashes in future. However, the number of tickets received in the lifetime was not significantly different between the two driver-groups.

Table 103: Tests of Equality of Crash and Citation Histories Using the Wilks' Lambda Test

	Wilks' Lambda	F	Sig.
The numbers of tickets in the last 5 years	0.979	3.288	0.072
The numbers of tickets in the life time	0.987	2.036	0.156
The numbers of crashes in the previous 4 years *	**0.963**	**5.804**	**0.017**
The numbers of crashes in the life time (except the study period) *	**0.947**	**8.433**	**0.004**

* indicates a significant mean difference ($\alpha = 0.05$).

Table 104 shows the correlation between age, crash history, and citation history and indicates that there were no strong correlations among them.

Table 104: Correlation Analysis on Crash and Citation Histories and Age Factors

	age	The numbers of tickets in the last 5 years	The numbers of tickets in the life time	The numbers of crashes in the previous 4 years	The numbers of crashes in the life time (except the study period)
age	1.00	-	-	-	-
The numbers of tickets in the last 5 years	-0.17	1.00	-	-	-
The numbers of tickets in the life time	0.11	0.49	1.00	-	-
The numbers of crashes in the previous 4 years	-0.08	0.24	0.20	1.00	-
The numbers of crashes in the life time (except the study period)	0.06	0.12	0.22	0.42	1.00

Finally, this study repeated the linear discriminant analysis including two crash variables (the numbers of crashes in the previous 4 years and the numbers of crashes in the lifetime except the 14-months study period). Table 105 shows the performance of the discriminant analysis using all selected behavioral exposure metrics and crash histories. As a result, 88.3 % of drivers who were not involved in crashes and 76.2 % of drivers who were involved in crashes were correctly classified. Overall performance of the model using this metric was 86.3 %. The small improvement of classification performance was found compared to the performance result of all drivers using only driving behavior activity exposure metrics (overall 2.3 %). This result indicates that although previous crash involvement history is related to the future crash involvements, this impact may be explained by driving behavior activity metrics.

Table 105: Classification Results Including Crash History Variables

Crash Involvements	Predicted Group Membership	
	Drivers who *were not involved* in crashes	Drivers who *were involved* in crashes
Drivers who *were not involved* in crashes	88.3 %	11.7 %
Drivers who *were involved* in crashes	23.8 %	76.2 %

Logistic Regression Model

From the linear discriminant analyses (LDA), this study selected numerous potential driving behavior activity exposure metrics that had strong powers of discriminating two driver-groups with and without crash involvements and found that those behavioral exposure metrics were different based on gender of drivers. Unlike to the linear discriminant analysis, the logistic regression model can employ both discrete and continuous variables for modeling process. Thus, in addition to gender, this study examines the impact of age on relationships between behavior exposure metrics and crash involvements using the logistic regression models. Table 106 shows the sample size of drivers by age, gender, and crash involvements.

Table 106: Sample Size of Drivers by Age, Gender, and Crash Involvement

	Drivers who were not involved in crashes		Drivers who were involved in crashes		Total
Age	Male	Female	Male	Female	
15 - 24	2	5	0	1	8
25 - 34	6	8	2	1	17
35 - 44	13	14	0	3	30
45 - 54	10	18	6	3	37
55 - 64	20	26	2	3	51
65 +	12	7	3	2	24
Subtotal	63	78	13	13	167
Total	141		26		167

From 27 behavioral exposure metrics selected from the linear discriminant analysis and the age variable, this study finally selected four exposure metrics having significant coefficient values form the logistic regression model (Table 107). This study found that outside-regional travel mileage during afternoon, frequency of 15 mph over-speed activity on arterials during morning, frequencies of hard decelerations greater than

8 mph/s on freeways during afternoon, frequency of hard decelerations greater than 4 mph/s on freeways during morning, and previous crash location exposure were significant in the logistic regression model for all drivers. However, the variables of gender and age were not significant for classification process using the logistic regression model. The result of the logistic regression model was similar to that of the stepwise linear discriminant analysis.

Table 107: Significances of Each Independent Variable in the Model for All Drivers

Behavior Activity Exposure	B	S.E.	Wald	df	Sig.
Travel mileage outside regional area during afternoon *	0.001	0.000	10.476	1	0.001
Frequency of 15 mph over-speed activity per mile on arterials during morning *	0.302	0.110	7.496	1	0.006
Frequency of hard deceleration per mile on freeways during afternoon (8 mph/s) **	80.767	51.663	2.444	1	0.118
Frequency of hard deceleration per mile on freeways during morning (4 mph/s) *	0.895	0.356	6.327	1	0.012
Previous crash location exposure per mile *	0.070	0.024	8.795	1	0.003
Constant	-7.587	1.506	25.398	1	0.000

* indicates a significant mean difference (α = 0.05).
** indicates a significant mean difference (α = 0.1).

Albeit with relatively small R-square values (Table 108), the omnibus test in Table 109 showed the developed model using five selected behavioral metrics was statistically significant for classifying the two driver-groups (less than 0.05). The goodness-of-fit test using the Hosmer and Lemeshow test in Table 109 indicated that the developed statistical model fitted well a set of observation (greater than 0.05).

Table 108: Model Pseudo R-Square Values for All Drivers for All Drivers

-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
88.865	0.197	0.326

Table 109: Significance of Model Fitting and Goodness-of-Fit Test for All Drivers

	Chi-square	Degree of Freedom	Significant level
Omnibus Test	27.454	5	0.000
Hosmer and Lemeshow Test	4.037	8	0.854

Using the logistic regression model, this study tried to interpret the relationships between each behavioral exposure measure and crash involvement rates. For the multiple independent variables, the logistic regression model represents

$$\text{Probability}\,(Event) = \frac{1}{1+e^{-Z}}$$

where, Z is the linear combination such as $Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_P X_P$,

p is the number of independent variables used in the model.

Thus, the final logistic regression model for all drivers using the coefficients of independent variables in Table 106 is

$$\text{Probability}\,(Event) = \frac{1}{1+e^{-(-7.587+0.001X_1+0.302X_2+80.767X_3+0.895X_4+0.07X_5)}}$$

where,

X_1 is the travel mileage outside regional area during afternoon,

X_2 is the frequency of 15 mph over-speed activity on arterials during morning,

X_3 is the frequency of hard deceleration on freeways during afternoon (8 mph/s),

X_4 is the frequency of hard deceleration on freeways during morning (4 mph/s),

X_5 is the previous crash location exposure per mile,

The "Event" indicates either crash-involvement or crash-not-involvement.

As a result, 94.2 % of drivers who were not involved in crashes and 54.5 % of drivers who were involved in crashes were correctly classified. Overall performance of the model using those driving behavior activity metrics in the logistic regression model was 87.2 %.

Classification and Regression Tree (CART) Analysis

In addition to the linear discriminant analysis (LDA) and logistic regression model, this study utilized the classification and regression tree (CART) analysis, commonly referred to as hierarchical tree-based regression (HTBR) [61] to visualize the classification process and verify the results of the previous two models. This technique is a non-parametric statistical method, so it does not require any distribution assumptions [61].

Similar to the logistic regression model, the classification and regression tree (CART) analysis was performed with the 27 behavioral exposure metrics and demographic data such as age and gender. Of those independent variables, the CART will also decide the most important variables for clustering drivers into the two different crash-involvement groups. Figure 87 shows the results of classification tree analysis.

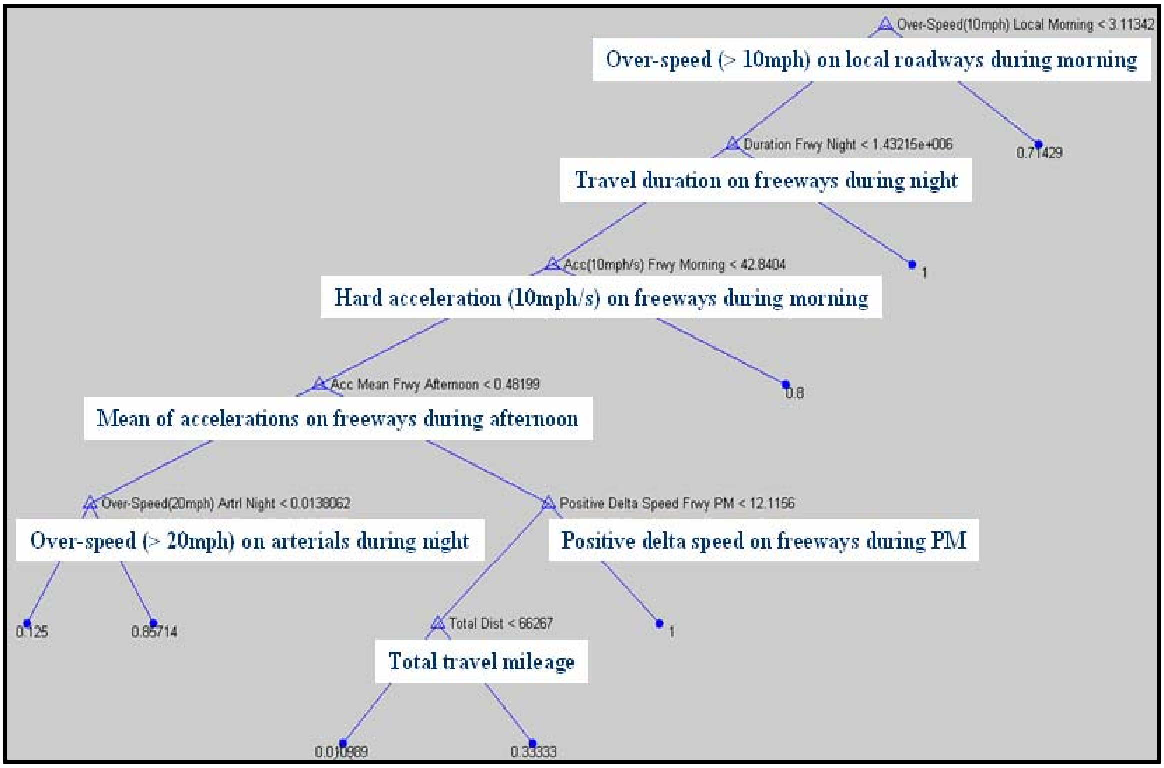


Figure 87: Result of the Classification and Regression Tree Analysis (All Drivers)

Finally, this study proposes driving behavior activity exposure metrics of individual drivers for possible safety surrogate measures as well as for driver re-training and education programs in next chapter.

Chapter Twelve

PRACTICAL APPLICATIONS

Surrogate Safety Measure

Because identifying roadways or intersections where crashes will frequently occur and understanding the causes of those possible crashes are very important, various crash prediction models and surrogate safety measures have been examined over several decades.

Traffic volume, driving speed, and speed variation are mainly used for predicting crash rate and severity on freeways as surrogate safety measures [4, 5, 62]. Conflict events, deceleration rates, braking power distributions, speed variations, number of vehicles caught in dilemma zone, and numbers of signal violations are useful surrogate measures especially for arterials [3].

However, those safety surrogate measures require conducting a field survey during a specific period and spending significant labor and time. Using the GPS-observed behavioral exposure metrics, potentially high crash risk locations (roadway segments or intersections) can be easily identified without spending much time and high cost. In addition, the GPS-observed driving behavior activity metrics can help safety engineers and policy makers determine hazardous locations that need to be improved safety conditions and to be implemented additional safety devices in advance before crashes occurred.

As this study already showed that the frequency of hard decelerations is strongly related with the crash involvement rates, this study tried to find locations where hard deceleration behaviors frequently occurred from individual GPS-instrumented vehicle data. Figure 88 shows the potential hazardous locations based on hard deceleration (8 mph/s) behavior and crash locations in previous occurred between 2000 and 2002.

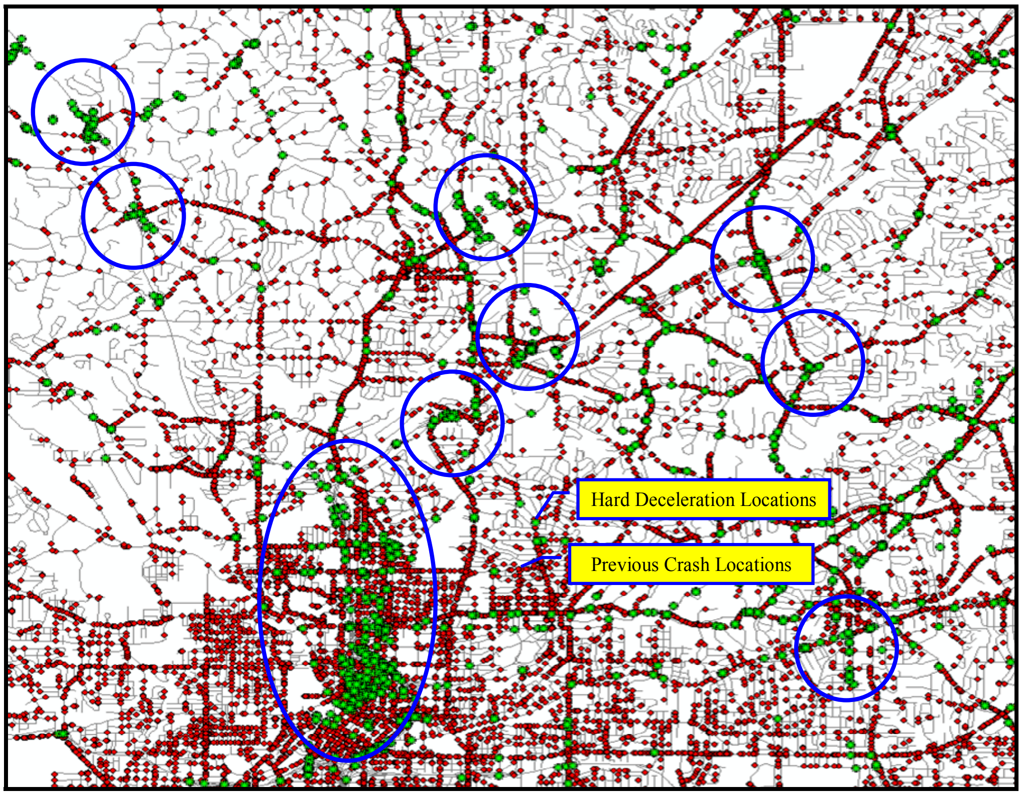


Figure 88: Potential Hazardous Locations based on Hard Deceleration Behavior

Driver Safety Evaluation

The findings of this study can be incorporated into driver education campaigns and driver evaluation or monitoring programs since this study found that crash-involved drivers traveled more, at high speeds, and with larger accelerations than non-crash-involved drivers. If driving behavior of individual drivers is regularly monitored and evaluated, drivers will be able to recognize and modify their driving activities linked with potential crash involvements. For this purpose of driver safety education and performance monitoring programs, this study provides the examples of application showing how these driving behavior-activity crash exposure metrics can help drivers recognize their crash-related driving behaviors. All examples show driving behavior on freeways especially during the morning period.

Figure 89 shows the driving behavior of one female driver who were actually involved in a crash during the 14-months study period. She was 33 years old, had 18 years driving experience, received one speeding tickets in the life time, and was involved in four crashes in the life time (except one crash in the study period). Based on her driving behavior activity patterns during the 6-months period, this study found that she had the high frequency of hard deceleration activities (4 mph/s: 2 times per mile and 8 mph/s: 6 times per 1000-mile), exhibited 16 times of 10 mph over-speeding every mile, and traveled on average at speed of 10 mph over than speed limit. Overall, her crash involvement risk was 0.98. Thus, this study can suggest her to reduce driving speeds and to travel with sufficient safety distance.

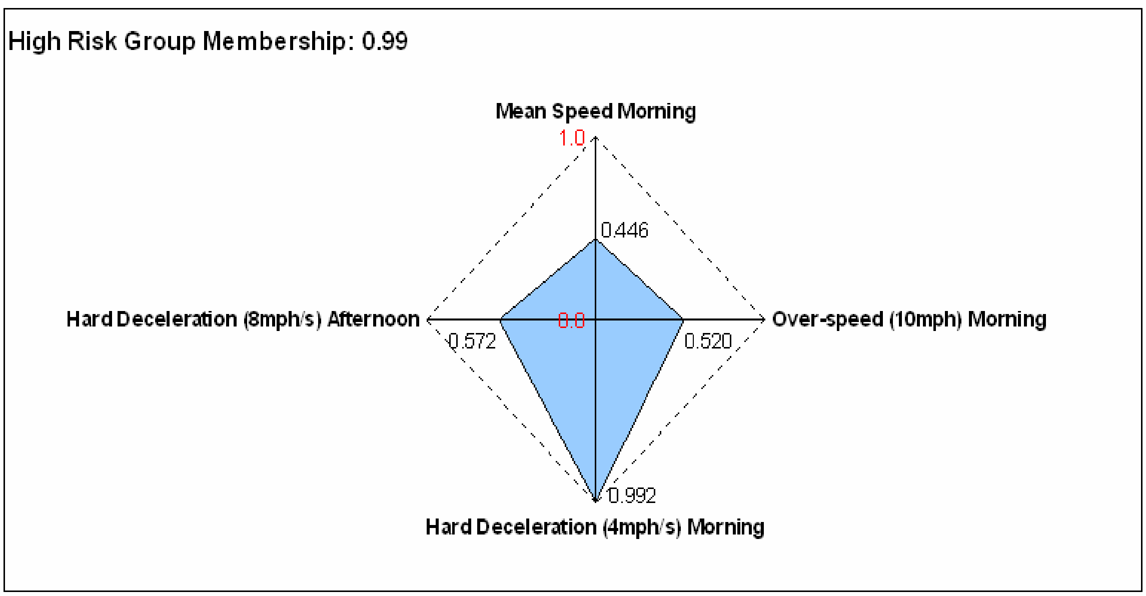


Figure 89: Potential Crash Risk Levels of Driver 2210

Figure 90 shows the driving behavior of one older male driver who had not a crash involvement during the study period. He was 72 years old, had 55 years driving experience, received no speeding tickets in the life time, and was involved in three crashes (no fault) in the life time (but no crash involvements in the last 4 years). Based on his driving behavior activity patterns during the 6-months period, this study found that he had very low frequency of hard acceleration (4 mph/s: 0.06 times per mile and 8 mph/s: zero times per mile), exhibited only three times of 10 mph over-speeding every mile, and traveled on average at speeds of 4.7 mph over than speed limits. Overall, her crash involvement risk was 0.24. Thus, this study can suggest that he maintain his driving patterns.

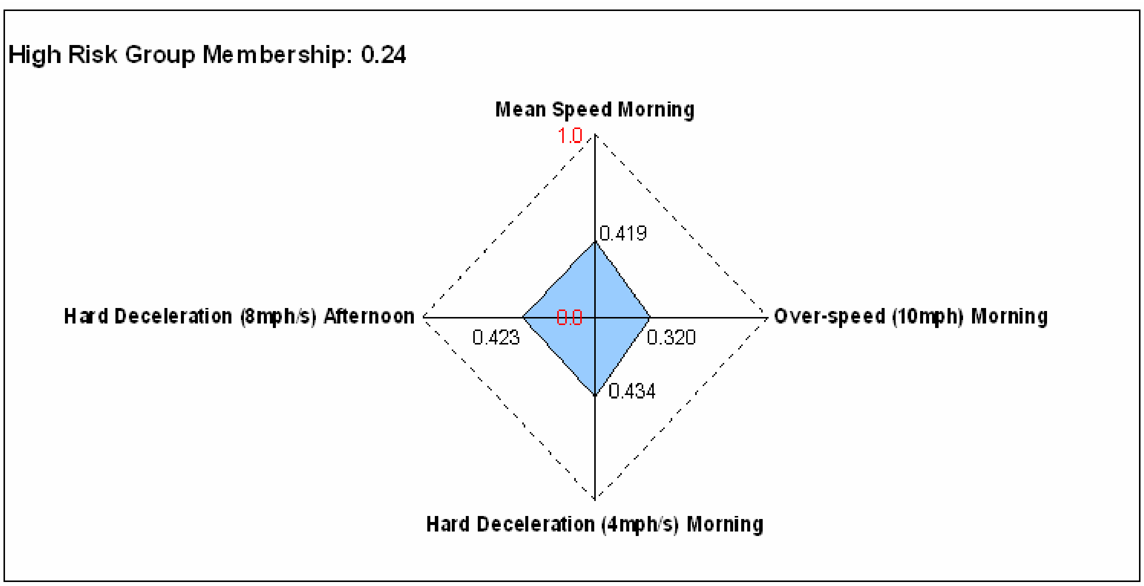


Figure 90: Potential Crash Risk Levels of a Driver 2021

Figure 91 shows the driving behavior of one male driver who had no crash involvement during the study period. He was 56 years old, had 41 years driving experience, received three speeding tickets in the life time (one in the last 5 years), and was involved in three crashes (one fault) in the life time. Based on his driving behavior activity patterns during the 6-months period, this study found that he had the high frequency of hard decelerations (4 mph/s: 0.14 times per mile and 8 mph/s: 2 times per 1000-mile), exhibited 19 times of 10 mph over-speeding every mile, and traveled on average at speed of 13 mph over than speed limit. Overall, her crash involvement risk was 0.58. Thus, this study can suggest him to reduce driving speeds and hard accelerations and decelerations.

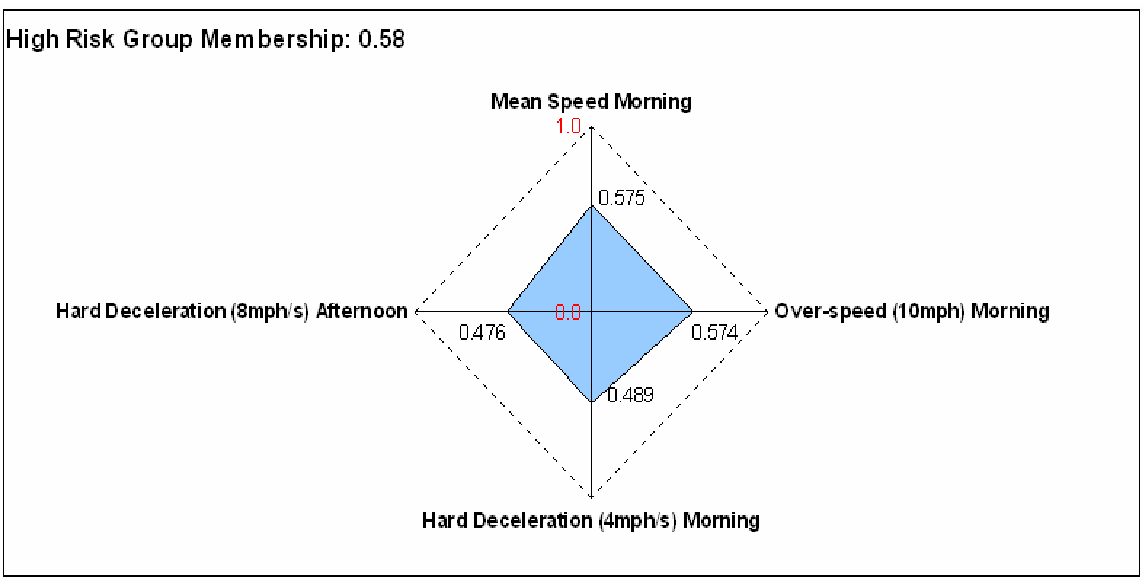


Figure 91: Potential Crash Risk Levels of a Driver 2090

Chapter Thirteen

CONCLUSIONS AND FUTURE RESEARCH

Summary of the Findings

Identifying and understanding the relationships between observed driving behavior activity over long-term periods and corresponding crash involvement rate is paramount to enhancing safety improvement programs and providing useful insights for transportation safety engineers, policy makers, insurance industries, and the public.

Unlike previous data collection methods, recent advancement in mobile technology and accuracy of global positioning systems (GPS) allow researchers to monitor driving activities of large fleets of vehicles, for long-time study periods, at great detail. The GPS-measured travel data provide abundant reliable information to help identify relationships between various driving behavior activities and crash involvement rates (crash risks) under varying conditions (i.e., where, when, how, and under what conditions the driving occurs) of facility type and time of day. Coupling the detailed travel information with known driver, household, and vehicle characteristics, activities (operations) can then be tied back to a wide variety of socio-demographic parameters. Furthermore, GPS-measured data can evaluate how driving behavior patterns change during a trip in response to changes in roadway operating conditions. In this respect, to identify and substantiate specific driving behaviors linked with crash involvements is the challenge.

This study investigates the driving patterns of 141 drivers who have not experienced crashes and 26 drivers who have experienced crashes during the 14-month study period using the longitudinally collected GPS data during a six-month Commute Atlanta study. This study allows an empirical investigation to assess whether drivers with recent crash experiences have exhibited different driving or activity patterns (travel mileage, travel duration, speed, acceleration, speed stability pattern, frequency of unfamiliar roadway activities, frequency of turn movement activities, and previous crash location exposure). To analyze driving behavior and activity patterns, this study also discussed various techniques of implementing GPS data streams in safety analyses. The summary on GPS data processing techniques is as follows:

- Of three acceleration calculation methods, the central difference method provides more stable and moderate acceleration rates. Researchers should note which acceleration computation method is implemented in their research, understand the differences of acceleration computation methods, and be careful in particular when the estimates are obtained from speed-acceleration combinations.
- General GPS data streams contain random errors that have the potential to affect speed, acceleration, and travel mileage estimates. Of various smoothing techniques, this study developed and found that the modified Kalman filter provided the smallest differences from the VSS-derived speed, acceleration, and travel mileage estimates across all statistical metrics.
- Because tire diameter fluctuates as a function of air pressure inside tires related to internal temperature, the vehicle speed sensor (VSS) or a distance measurement instrument (DMI) can produce systematic errors caused by tire

inflation if the tire diameters are not taken into account. This study showed relationships between tire inflation and average speeds based on travel mile.

Regarding the potential driving behavior-activity crash exposure measures using the GPS-observed data, this study found the following relationships:

- While the choices on trip time by drivers between who were involved and not involved in crashes were not different, distributions of activity exposure by facility types were significantly different. Drivers who had experienced crashes during the 14-month study period were more likely to use freeways for the 6-months behavior observation period.
- Travel mileage and duration have positive relationships with the crash involvement rate since this study showed that exposures to mileage and duration of crash-involved drivers were much higher (32% and 21%, respectively) than those who were not involved in crashes. This result supports the exposure theory where higher exposure means higher opportunity of being involved in a crash. Especially, Crash-involved drivers traveled much higher during peak periods and nighttime on freeways than non-crash-involved drivers.
- Drivers who had experienced crashes were more likely to travel at high speeds than drivers who had not experienced crashes based on both "amount" and "frequency" measures of speed activities. Again, this study supports the conventional theory, "higher speed is linked to higher opportunity of crash

involvements". Speeding activities of drivers with crash involvements were larger at most times than drivers who without crash involvements. Based upon the sample data, morning and afternoon provided the largest differences in speeding behaviors between the two driver-groups.

- Similar to the speed patterns, drivers who were involved in crashes frequently tended to produce large acceleration and deceleration activities. This study provided that larger acceleration levels and more frequent hard acceleration and deceleration events (larger speed changes) indicate higher opportunity of being involved in crashes. Although the amount of accelerations and the frequency of hard acceleration and deceleration events on arterials or local roadways were larger than those on freeways, significant behavioral differences between the two driver-groups were found from the activities on freeways instead of on arterials and local roadways. Significant differences on freeways, especially during morning and afternoon periods between the two driver-groups imply that drivers who were involved in crashes may be more likely to conduct tailgating behavior or frequently use cellular phones. Further study will be needed to confirm the impacts of tailgating behavior and cellular phone use on hard acceleration and deceleration activities using the more sophisticated instruments
- To better understand driver behavior, this study proposed a new metric indicating a speed stability pattern with the longitudinally observed speed trajectories of individual drivers, which could not be investigated in previous studies due to the difficulty of data collection. Although this study did not find

any statistical differences in speed stability patterns between the two driver-groups, this study provided that drivers with crash involvements tended to more frequently change their driving speeds than drivers without crash involvements during the study period

- The number of different roadways and the number of unfamiliar roadways inside 13-counties area traveled by drivers who had crash involvements during the 14-months period were larger than those of drivers who did not experience any crashes, but statistical significances were not found. This study suggests that outside-regional unfamiliar roadway exposure also needs to be evaluated in future research.
- The average left-turn exposure per mile of drivers with and without crash involvements were 0.36 and 0.41, and the average right-turn exposure per mile of drivers with and without crash involvements were 4.19 and 3.83. However, the differences between the two driver-groups were not statistically significant. The turn exposures per mile including left and right turns of drivers who were involved in crashes were not significantly different from drivers who were not involved in crashes, either. This result indicates that there may be no differences in the choices of left or right turn activity between drivers with and without crash involvements
- Although drivers who were involved in crashes more frequently tended to travel along previous crash locations (roadways and intersections) than drivers who were not involved in crashes, the previous crash location exposure metric between the two driver-groups was not significantly different. However, for

the final combined modeling process, this metric significantly improved the performance of classification.

- Of the numerous driving behavior activity exposure metrics, this study found that driving behavior-activity exposure metrics strongly related with crash involvement rates were travel mileage, travel duration, speed pattern, acceleration activities, and previous crash location exposure. In addition, based on the result of classification modeling process, this study examined the impact of interactions among potential driving behavior-activity exposure metrics on the performance of model prediction. Thus, to obtain high performance of classification, the classification process need to be performed with combined potential behavior exposure metrics rather than employing them individually.
- This study found that critical behavior activity exposure metrics related to crash involvements between male and female drivers were different. For male drivers, the frequency of 20 mph over-speed activity, the frequency of hard deceleration greater than 8 mph/s, the frequency of hard acceleration greater than 10 mph/s, and the previous crash location exposure were found as important exposure metrics. For female drivers, the frequency of 15 mph over-speed activity, the frequency of hard deceleration greater than 4 mph/s, and travel duration were critical.
- Based on findings in this study, safety engineers and policy makers can develop some strategies (or campaigns), and drivers can modify their activity patterns and reduce unnecessary travels. Insurance companies simply using

the total mileage estimate for insurance premium structures as one of behavior activity measures may enhance and further refine current insurance classification decision rules with the more disaggregated behavioral exposures such as speeding, hard acceleration/deceleration, and exposure to previous crash locations.

- This study found that that behavior activity metrics may be employed as the safety surrogate measures to select hazardous roadways or intersections where crash-related behavior activities such as over-speed activities or hard deceleration events frequently occur.

Contributions to Transportation Safety

This study contributes to the safety in the transportation field in numerous ways. First, this study is one of the first attempts to evaluate crash involvement rate of individual drivers using driving behavior activity data longitudinally collected from GPS-instrumented vehicles. Thus, this study points out the limitations of the existing methods used in various transportation research and provides useful approaches showing how the GPS technology can be used for safety research especially requiring the large-scale data collection process.

Second, unlike the previous research efforts that employ aggregate exposure measures, this study proposes numerous driving behavior-activity exposure metrics to evaluate the probability of being involved in a crash. Based on the proposed behavior activity metrics and the developed models, the cause-effect relationships between driving behaviors and crash involvements can be evaluated in detail. However, more sample data will be very useful in refining those relationships.

Third, this study discusses differences in driving behavior of drivers who were involved and not involved in crashes and examines that drivers within the same age and gender group may not produce the same behavior activity patterns. As a result, driving behavior can be more strongly related to the crash involvement rates than driver age and gender. Thus, this study may support the analytical framework used by insurance industries to estimate insurance premium since this study indicates that insurers may provide more reasonable and equitable premium structure to customers.

Fourth, this study expects that driving behavior-activity exposure metrics of individual drivers can be utilized as one of safety surrogate measures to identify potentially high crash-risk roadways where hard deceleration events or high speeding patterns frequently occur. Thus, the GPS-observed driving behavioral metrics can help safety engineers and policy makers determine hazardous locations that need to be improved safety conditions and to be implemented additional safety devices to mitigate risk and reduce the number of future crashes.

Fifth, this study provides more detailed and effective techniques that evaluate potential crash-risk of individual drivers and expects that current driver education programs and other safety campaigns can be much improved.

Finally, it is expected that this study provides useful guidance for researchers who plan to evaluate the relationships between driver behavior and crash risk with larger sample data in future.

Limitations and Future Study Suggestions

This study tried to evaluate differences in driving behavior activity patterns between crash-involved and non-crash-involved drivers using the GPS-observed activity data prior to crash involvements. Thus, this study can be considered as an observational research, not an experimental design research. In the experimental design research, researchers can recruit large number of drivers, crash-involved and non-crash-involved drivers, and can evaluate their driving behavior patterns, but those driving behavior patterns may not be normal driving behavior since drivers may modify their driving behavior patterns after being involved in a crash. On the other hand, the observational research such as this study tries to randomly recruit participants and observe their normal driving behavior activity patterns and crash involvements. Thus, this observational study may have relatively small sample size regarding crash-involved drivers due to the rarity of crash involvement. Although this study collected crash data from the participants during the 14-month period, due to the rare event characteristics of crash involvements, this study employed the small sample size (only 26 drivers) for drivers who were involved in crashes. Thus, this study suggests that researchers in future who have larger sample data and longer period of data collection need to re-examine driving behavior activity patterns between crash-involved and non-crash-involved drivers.

In addition to the sample size issue, this study used only the self-reported crash data for clustering drivers into the two different groups. As mentioned earlier, it may be possible that this source of crash data is under-estimated. Thus, this study also suggests that researchers need to compare at least two different sources of crash data such as the official crash database and the self-reported crash data and verify them for the future research.

The self-reported crash data used in this study did not contain other important information such as where, when, what conditions crashes occurred, crash type, and severity of a crash. Due to this limitation, this study evaluated driving behavior activity patterns based on only crash involvement status. If researcher can evaluate relationships between driving behavior activity patterns and fatality or severity of crashes, researchers may find other important behavior metrics and strong relationships with demographic characteristics such as gender and age. Such information may provide the relationships between behavioral metrics and crash situations in detail.

Due to the limitation regarding roadway characteristics in the GIS database, this study evaluated only driving behavior activity patterns inside 13-counties area such as speeding patterns, unfamiliar roadway exposure, left/right turn exposure, and previous crash location exposure. Outside-regional driving behavior activity exposures may have significant differences in crash involvement rate, so they need to be performed in future research.

Although this study evaluated disaggregated driving behavior-activity exposures based on time of day and facility type, further investigations regarding exposures to roadways having different geometric designs (grade and curvature) and operational designs (speed limit and traffic volume) need to be performed. In addition, speed difference between individual driving speed and surrounding traffic speed may be one of potential behavioral crash-related exposure measures.

This study also suggests that the relationships between crash involvement rate and activities based on trip purpose (commuting or shopping) may be a potential behavior exposure measure. Due to the small sample size, this study was not able to evaluate impacts of vehicle types on crash involvements, so future research also need to be investigated this issue. Finally, this study suggests that researchers may need to evaluate ability to modify crash-related behavior by driver safety evaluation programs.

APPENDIX A

Background of the OBD

An onboard diagnostics (OBD) system initially developed for reducing vehicle emissions is a sophisticated electronic monitoring system. Vehicle manufacturers started to develop electronic control systems depending on their technologies to meet emissions control regulation mandated by U.S. EPA during the 70's and 80's. California required electronic emissions control systems in 1985 for 1988 and later model year vehicles in order to control components and systems related to emissions.

The first regulation of onboard system (OBD I) using oxygen sensors for maintaining a stoichiometric air/fuel ratio (14.7:1) during closed-loop operations was applied to only California certified vehicles. The onboard diagnostics (OBD I) system was first introduced by General Motors in 1981. At that time, the design of systems varied from vehicle manufacture to manufacturer. OBD I mentioned in Title 13 California Code 1968 filed on November 15, 1985 [63] was not designed to detect fully emissions related failures.

This electronic emissions control systems (OBD I systems) provided minimal monitoring requirements. General OBD I requirements were fuel metering and delivery system, exhaust recirculation system, Powertrain Control Modules (PCM) / Electronic Control Module (ECM) and other emissions related electrical components, Malfunction Indicator Light (MIL), and Diagnostics Trouble Code (DTC) [64]. The OBD I systems did not monitor engine misfires, converter failures, and evaporative system problems [64]. There was also no standardization throughout the vehicle industry, so each vehicle manufacturer used a different term for the warning light illuminated when a fault was determined [65], used different locations of Data Link Connector (DLC), and provided different DTCs. Another limitation of the OBD I systems was that it could not detect certain kinds of problems such as a bad catalytic converter. Furthermore, the OBD I systems would only illuminate the MIL after a system failure had occurred because it had no way of monitoring progressive deterioration of emissions-related components [66].

Thus, the Society of Automotive Engineers (SAE) had set standardizations for the onboard diagnostics (OBD II) systems in terms of electric terms (J1930), diagnostic connector (J1962), scan tool (J1978), diagnostic test modes (J1979), diagnostic trouble code definition (J2012), data network interface (J1850) in order to provide fundamental and common requirements when vehicle manufactures and the automotive industry implemented the increasingly complex electronic systems on new vehicles [67].

The state of California began requiring the more strict emissions control regulation on the monitoring engine-related systems and detecting failures of the emissions-related systems. This regulation was mandated all Light-Duty Vehicles (LDV) on 1994 and later model year to equip onboard diagnostics (OBD II) systems, but this regulation was waived for the next two years (1994 and 1995 model year vehicles) since vehicle manufacturers were not able to comply with the federal regulation.

The onboard diagnostics (OBD II) systems controlled by engine computer can be used for new vehicles certification as well as the verification of old vehicles in terms of the inspection and maintenance (I/M) program. As the electronic technologies were rapidly developed and computer network systems were improved, OBD II systems became more sophisticated and complicated than previous OBD I systems. The OBD II system-equipped vehicles typically have twice the number of oxygen sensors than previous vehicles, operate more powerful PCM, and use electronically erasable programmable read only memory (EEPROM). Thus, the onboard diagnostics (OBD II) systems can handle up to 15,000 new calibration variables with 16 bit (Chrysler) or 32 bit (Ford and GM) and allow the PCM to be reprogrammed with revised or updated software [66]. The onboard diagnostics (OBD II) systems are required to monitor catalyst system, engine misfire, evaporative system, secondary air system, air conditioning system, fuel system, oxygen sensors, Exhaust Gas Recirculation (EGR) system, thermostat, and other comprehensive components [68].

Connection the GT-TDC with the OBD System

All OBD II system-equipped vehicles (1996 and later model year vehicles) have a SAE J1962 data link connector (DLC). The Society of Automotive Engineers (SAE) defines definition of the data link connector (DLC) in terms of locations, access, and visibility, so the data link connector (DLC) containing 16 pins is usually located under the dashboard on the driver's side (Figure 92). After connecting between the onboard diagnostics (OBD II) system and the OBD connector with RS 232 interface (Figure 92), the GT-TDC begins communicating with the engine computer.

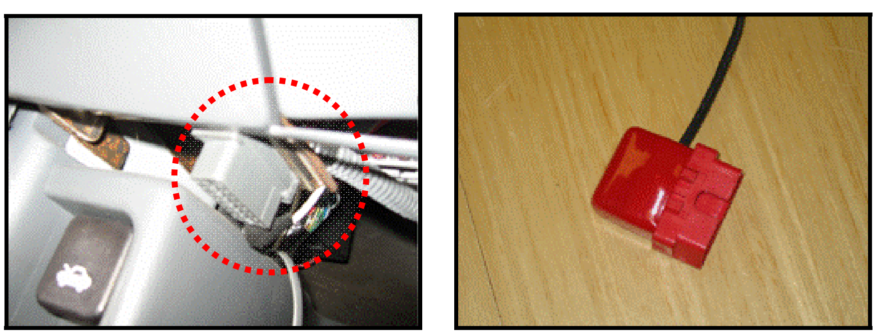


Figure 92: Data Link Connector (DLC) and RS-232 Cable

To communicate between the GT-TDC (data receiver) and engine computer embedded in a vehicle (data sender), a specific communication interface (or protocol) must be established. The three types of communication protocols exist such as International Standards organization (ISO) 9141, SAE J1850 Variable Pulse Width Modulation (VPW), and SAE J1850 Pulse Width Modulation (PWM). In general, Fords uses SAE J1850 PWM, General Motors uses SAE J1850 VPW, and other manufactures of imported vehicles use ISO 9141 protocol interfaces.

APPENDIX B

Results of Correlation Analysis Using the Selected Behavioral Exposure Measures

	frwy pm distance	frwy night distance	artrl pm distance	outside am distance	outside afternoon distance	total distance	frwy pm duration	frwy night duration	artrl am duration	artrl pm duration	outside am duration	outside afternoon duration
frwy pm distance	1.0	0.4	0.3	0.1	0.0	0.6	1.0	0.4	0.2	0.4	0.2	0.0
frwy night distance	0.4	1.0	0.1	-0.1	-0.1	0.3	0.3	1.0	-0.1	0.1	0.0	0.0
artrl pm distance	0.3	0.1	1.0	0.1	-0.2	0.5	0.3	0.1	0.4	0.9	0.3	0.0
outside am distance	0.1	-0.1	0.1	1.0	0.3	0.0	0.1	-0.1	0.1	0.1	0.7	0.2
outside afternoon distance	0.0	-0.1	-0.2	0.3	1.0	-0.2	-0.1	-0.1	-0.2	-0.2	0.0	0.8
total distance	0.6	0.3	0.5	0.0	-0.2	1.0	0.5	0.3	0.4	0.4	0.4	0.1
frwy pm duration	1.0	0.3	0.3	0.1	-0.1	0.5	1.0	0.3	0.3	0.4	0.2	0.0
frwy night duration	0.4	1.0	0.1	-0.1	-0.1	0.3	0.3	1.0	-0.1	0.1	0.0	0.0
artrl am duration	0.2	-0.1	0.4	0.1	-0.2	0.4	0.3	-0.1	1.0	0.5	0.4	-0.1
artrl pm duration	0.4	0.1	0.9	0.1	-0.2	0.4	0.4	0.1	0.5	1.0	0.3	0.0
outside am duration	0.2	0.0	0.3	0.7	0.0	0.4	0.2	0.0	0.4	0.3	1.0	0.2
outside afternoon duration	0.0	0.0	0.0	0.2	0.8	0.1	0.0	0.0	-0.1	0.0	0.2	1.0
outside night duration	0.1	0.5	0.3	0.1	0.3	0.2	0.1	0.5	0.0	0.3	0.2	0.4
total duration	0.4	0.1	0.5	0.1	0.2	0.8	0.3	0.1	0.4	0.5	0.5	0.6
frwy morning mean speed	0.0	0.2	0.0	0.1	0.0	0.2	0.0	0.2	0.0	-0.1	0.1	0.0
frwy night mean speed	0.1	0.2	0.0	0.1	0.0	0.2	0.1	0.2	-0.1	-0.1	0.0	0.0
artrl am delta speed	-0.1	-0.2	0.3	0.1	-0.1	0.2	-0.1	-0.2	0.1	0.1	0.1	-0.1
frwy pm positive delta speed	0.2	0.2	0.0	-0.1	-0.1	0.2	0.2	0.2	0.0	0.0	-0.1	-0.1
artrl am positive delta speed	0.2	-0.1	0.1	-0.1	0.0	0.2	0.2	-0.1	0.1	0.0	0.0	0.1
local am positive delta speed	0.3	0.1	0.1	0.0	0.0	0.2	0.3	0.1	0.0	0.1	0.1	-0.1
local night positive speed	0.2	0.4	0.0	-0.1	-0.1	0.2	0.2	0.4	0.1	0.0	0.0	-0.2
frwy morning overspeed (10mph)	0.2	0.2	0.1	0.1	-0.1	0.2	0.2	0.2	0.1	0.0	0.0	-0.3
local morning overspeed (10mph)	0.1	0.1	0.1	-0.1	-0.1	0.1	0.1	0.1	0.0	0.0	-0.1	-0.2
artrl morning overspeed (15mph)	0.1	0.2	0.1	0.0	-0.2	0.3	0.1	0.2	0.0	0.0	0.0	-0.1
artrl night pverspeed (f20mph)	0.0	0.1	0.0	-0.1	0.0	0.1	0.0	0.1	-0.1	0.0	0.0	0.1
frwy afternoon mean acceleration	0.2	0.1	0.1	-0.1	-0.1	0.1	0.2	0.1	0.1	0.1	0.0	0.0
frwy morning hard acceleration (10mph/s)	0.3	0.1	0.0	0.0	-0.1	0.0	0.3	0.1	0.0	0.0	0.0	-0.2
frwy morning hard deceleration (4mph/s)	0.3	0.0	0.0	-0.1	-0.2	0.1	0.4	0.0	0.2	0.0	0.0	-0.1
artrl morning hard deceleration (4mph/s)	0.2	0.0	0.2	-0.1	-0.1	0.1	0.2	0.0	0.2	0.2	0.1	0.0
local night hard deceleration (4mph/s)	0.0	0.0	0.2	-0.1	-0.2	0.0	0.1	-0.1	0.2	0.2	0.0	-0.2
frwy afternoon hard deceleration (8mph/s)	0.0	-0.1	0.0	0.0	-0.1	0.0	0.0	-0.1	0.2	0.0	0.0	-0.1
unfamiliar roadway exposure	-0.2	0.0	-0.2	-0.2	-0.2	-0.5	-0.2	0.0	-0.2	-0.1	-0.3	-0.3
turn movememt exposure	0.0	0.1	0.0	-0.4	-0.5	0.0	0.0	0.0	0.0	0.0	-0.2	-0.3
Previous crash location exposure	0.0	0.2	0.0	-0.1	-0.2	0.2	0.0	0.1	0.1	-0.1	-0.1	-0.2

Cont'd

	outside night duration	total duration	frwy morning mean speed	frwy night mean speed	artrl am delta speed	frwy pm positive delta speed	artrl am positive delta speed	local am positive delta speed	local night positive speed	frwy morning overspeed (10mph)	local morning overspeed (10mph)
frwy pm distance	0.1	0.4	0.0	0.1	-0.1	0.2	0.2	0.3	0.2	0.2	0.1
frwy night distance	0.5	0.1	0.2	0.2	-0.2	0.2	-0.1	0.1	0.4	0.2	0.1
artrl pm distance	0.3	0.5	0.0	0.0	0.3	0.0	0.1	0.1	0.0	0.1	0.1
outside am distance	0.1	0.1	0.1	0.1	0.1	-0.1	-0.1	0.0	-0.1	0.1	-0.1
outside afternoon distance	0.3	0.2	0.0	0.0	-0.1	-0.1	0.0	0.0	-0.1	-0.1	-0.1
total distance	0.2	0.8	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.1
frwy pm duration	0.1	0.3	0.0	0.1	-0.1	0.2	0.2	0.3	0.2	0.2	0.1
frwy night duration	0.5	0.1	0.2	0.2	-0.2	0.2	-0.1	0.1	0.4	0.2	0.1
artrl am duration	0.0	0.4	0.0	-0.1	0.1	0.0	0.1	0.0	0.1	0.1	0.0
artrl pm duration	0.3	0.5	-0.1	-0.1	0.1	0.0	0.0	0.1	0.0	0.0	0.0
outside am duration	0.2	0.5	0.1	0.0	0.1	-0.1	0.0	0.1	0.0	0.0	-0.1
outside afternoon duration	0.4	0.6	0.0	0.0	-0.1	-0.1	0.1	-0.1	-0.2	-0.3	-0.2
outside night duration	1.0	0.4	0.0	0.0	-0.1	0.0	-0.1	0.0	0.1	-0.1	0.0
total duration	0.4	1.0	-0.1	0.0	0.1	0.0	0.2	0.0	0.0	-0.1	-0.1
frwy morning mean speed	0.0	-0.1	1.0	0.5	0.1	0.4	0.0	0.2	0.3	0.4	0.3
frwy night mean speed	0.0	0.0	0.5	1.0	0.1	0.3	0.0	0.2	0.2	0.2	0.1
artrl am delta speed	-0.1	0.1	0.1	0.1	1.0	0.0	0.3	0.1	0.1	0.0	0.1
frwy pm positive delta speed	0.0	0.0	0.4	0.3	0.0	1.0	0.1	0.3	0.3	0.5	0.4
artrl am positive delta speed	-0.1	0.2	0.0	0.0	0.3	0.1	1.0	0.4	0.2	0.1	0.1
local am positive delta speed	0.0	0.0	0.2	0.2	0.1	0.3	0.4	1.0	0.5	0.3	0.4
local night positive speed	0.1	0.0	0.3	0.2	0.1	0.3	0.2	0.5	1.0	0.3	0.5
frwy morning overspeed (10mph)	-0.1	-0.1	0.4	0.2	0.0	0.5	0.1	0.3	0.3	1.0	0.5
local morning overspeed (10mph)	0.0	-0.1	0.3	0.1	0.1	0.4	0.1	0.4	0.5	0.5	1.0
artrl morning overspeed (15mph)	-0.1	0.1	0.2	0.2	0.4	0.2	0.5	0.3	0.3	0.3	0.4
artrl night pverspeed (f20mph)	0.0	0.1	0.1	0.0	0.2	0.1	0.5	0.2	0.2	0.2	0.1
frwy afternoon mean acceleration	0.1	0.2	-0.2	0.1	-0.1	0.1	0.0	0.0	0.1	0.0	0.0
frwy morning hard acceleration (10mph/s)	0.0	-0.1	0.0	0.0	-0.2	0.0	0.0	0.1	0.2	0.2	0.0
frwy morning hard deceleration (4mph/s)	0.0	0.1	-0.2	0.0	-0.1	0.1	0.0	0.0	0.1	0.1	0.0
artrl morning hard deceleration (4mph/s)	0.2	0.2	-0.1	-0.1	-0.1	0.1	-0.1	0.1	0.1	0.1	0.3
local night hard deceleration (4mph/s)	0.0	0.0	0.0	0.0	-0.1	0.1	-0.2	0.1	0.1	0.1	0.3
frwy afternoon hard deceleration (8mph/s)	-0.2	-0.1	-0.1	-0.1	0.0	0.1	0.0	0.0	0.0	0.2	0.0
unfamiliar roadway exposure	-0.1	-0.4	-0.3	-0.1	-0.1	0.0	-0.1	-0.1	0.0	-0.2	-0.1
turn movememt exposure	-0.2	-0.1	-0.2	0.0	0.0	0.1	-0.1	-0.2	-0.1	0.0	0.0
Previous crash location exposure	-0.1	-0.1	0.4	0.3	0.1	0.2	-0.1	-0.1	0.1	0.2	0.2

Cont'd

	artrl morning overspeed (15mph)	artrl night pverspeed (f20mph)	frwy afternoon mean acceleration	frwy morning hard acceleration (10mph/s)	frwy morning hard deceleration (4mph/s)	artrl morning hard deceleration (4mph/s)	local night hard deceleration (4mph/s)	frwy afternoon hard deceleration (8mph/s)	unfamiliar roadway exposure	turn movememt exposure	Previous crash location exposure
frwy pm distance	0.1	0.0	0.2	0.3	0.3	0.2	0.0	0.0	-0.2	0.0	0.0
frwy night distance	0.2	0.1	0.1	0.1	0.0	0.0	0.0	-0.1	0.0	0.1	0.2
artrl pm distance	0.1	0.0	0.1	0.0	0.0	0.2	0.2	0.0	-0.2	0.0	0.0
outside am distance	0.0	-0.1	-0.1	0.0	-0.1	-0.1	-0.1	0.0	-0.2	-0.4	-0.1
outside afternoon distance	-0.2	0.0	-0.1	-0.1	-0.2	-0.1	-0.2	-0.1	-0.2	-0.5	-0.2
total distance	0.3	0.1	0.1	0.0	0.1	0.1	0.0	0.0	-0.5	0.0	0.2
frwy pm duration	0.1	0.0	0.2	0.3	0.4	0.2	0.1	0.0	-0.2	0.0	0.0
frwy night duration	0.2	0.1	0.1	0.1	0.0	0.0	-0.1	-0.1	0.0	0.0	0.1
artrl am duration	0.0	-0.1	0.1	0.0	0.2	0.2	0.2	0.2	-0.2	0.0	0.1
artrl pm duration	0.0	0.0	0.1	0.0	0.0	0.2	0.2	0.0	-0.1	0.0	-0.1
outside am duration	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	-0.3	-0.2	-0.1
outside afternoon duration	-0.1	0.1	0.0	-0.2	-0.1	0.0	-0.2	-0.1	-0.3	-0.3	-0.2
outside night duration	-0.1	0.0	0.1	0.0	0.0	0.2	0.0	-0.2	-0.1	-0.2	-0.1
total duration	0.1	0.1	0.2	-0.1	0.1	0.2	0.0	-0.1	-0.4	-0.1	-0.1
frwy morning mean speed	0.2	0.1	-0.2	0.0	-0.2	-0.1	0.0	-0.1	-0.3	-0.2	0.4
frwy night mean speed	0.2	0.0	0.1	0.0	0.0	-0.1	0.0	-0.1	-0.1	0.0	0.3
artrl am delta speed	0.4	0.2	-0.1	-0.2	-0.1	-0.1	-0.1	0.0	-0.1	0.0	0.1
frwy pm positive delta speed	0.2	0.1	0.1	0.0	0.1	0.1	0.1	0.1	0.0	0.1	0.2
artrl am positive delta speed	0.5	0.5	0.0	0.0	0.0	-0.1	-0.2	0.0	-0.1	-0.1	-0.1
local am positive delta speed	0.3	0.2	0.0	0.1	0.0	0.1	0.1	0.0	-0.1	-0.2	-0.1
local night positive speed	0.3	0.2	0.1	0.2	0.1	0.1	0.1	0.0	0.0	-0.1	0.1
frwy morning overspeed (10mph)	0.3	0.2	0.0	0.2	0.1	0.1	0.1	0.2	-0.2	0.0	0.2
local morning overspeed (10mph)	0.4	0.1	0.0	0.0	0.0	0.3	0.3	0.0	-0.1	0.0	0.2
artrl morning overspeed (15mph)	1.0	0.5	0.0	0.0	0.0	0.1	0.0	-0.1	-0.1	0.0	0.0
artrl night pverspeed (f20mph)	0.5	1.0	0.0	0.3	0.1	0.0	-0.1	0.0	-0.1	0.0	-0.1
frwy afternoon mean acceleration	0.0	0.0	1.0	0.0	0.4	0.4	0.1	0.0	0.1	0.2	0.0
frwy morning hard acceleration (10mph/s)	0.0	0.3	0.0	1.0	0.6	0.1	0.2	0.2	0.0	0.2	0.0
frwy morning hard deceleration (4mph/s)	0.0	0.1	0.4	0.6	1.0	0.4	0.1	0.2	0.0	0.2	0.1
artrl morning hard deceleration (4mph/s)	0.1	0.0	0.4	0.1	0.4	1.0	0.5	0.1	0.0	0.2	-0.1
local night hard deceleration (4mph/s)	0.0	-0.1	0.1	0.2	0.1	0.5	1.0	0.2	0.0	0.2	-0.2
frwy afternoon hard deceleration (8mph/s)	-0.1	0.0	0.0	0.2	0.2	0.1	0.2	1.0	0.0	0.1	0.0
unfamiliar roadway exposure	-0.1	-0.1	0.1	0.0	0.0	0.0	0.0	0.0	1.0	0.2	-0.1
turn movememt exposure	0.0	0.0	0.2	0.2	0.2	0.2	0.2	0.1	0.2	1.0	0.1
Previous crash location exposure	0.0	-0.1	0.0	0.0	0.1	-0.1	-0.2	0.0	-0.1	0.1	1.0

APPENDIX C

The Process of Stepwise Linear Discriminant Analysis for All Drivers

Step	Variables	Sig. of F value to be selected
1	totalduration	0.001
2	totalduration	0.002
	local_night_dec4	0.004
3	totalduration	0.002
	local_night_dec4	0.001
	Previous_crash_location	0.012
4	totalduration	0.028
	local_night_dec4	0.000
	Previous_crash_location	0.002
	outside_afternoon_dist	0.003
5	totalduration	0.096
	local_night_dec4	0.000
	Previous_crash_location	0.003
	outside_afternoon_dist	0.001
	artrl_morning_f15	0.005
6	totalduration	0.219
	local_night_dec4	0.001
	Previous_crash_location	0.005
	outside_afternoon_dist	0.000
	artrl_morning_f15	0.003
	frwy_morning_dec4	0.004
7	local_night_dec4	0.000
	Previous_crash_location	0.005
	outside_afternoon_dist	0.000
	artrl_morning_f15	0.001
	frwy_morning_dec4	0.002
8	local_night_dec4	0.001
	Previous_crash_location	0.004
	outside_afternoon_dist	0.000
	artrl_morning_f15	0.002
	frwy_morning_dec4	0.003
	outside_am_peakduration	0.042

Cont'd

Step	Variables	Sig. of F to Enter	Step	Variables	Sig. of F to Enter	Step	Variables	Sig. of F to Enter
	frwy_night_dist	0.356		frwy_night_dist	0.640		frwy_night_dist	0.558
	artrl_pm_peak_dist	0.055		artrl_pm_peak_dist	0.659		artrl_pm_peak_dist	0.845
	outside_afternoon_dist	0.008		outside_afternoon_dist	0.063		outside_afternoon_dist	0.015
	frwy_pm_peak_duration	0.052		frwy_pm_peak_duration	0.388		frwy_pm_peak_duration	0.498
	artrl_am_peakduration	0.081		artrl_am_peakduration	0.558		artrl_am_peakduration	0.959
	outside_am_peakduration	0.010		outside_am_peakduration	0.292		outside_am_peakduration	0.354
	outside_nightduration	0.164		outside_nightduration	0.943		outside_nightduration	0.961
	totalduration	0.001		f_morning_mean	0.165		f_morning_mean	0.220
	f_morning_mean	0.209		f_night_mean	0.078		f_night_mean	0.093
	f_night_mean	0.084		a_am_peak_delta	0.026		a_am_peak_delta	0.020
	a_am_peak_delta	0.012		f_pm_peak_pos	0.087		f_pm_peak_pos	0.192
	f_pm_peak_pos	0.076		a_am_peak_pos	0.222		a_am_peak_pos	0.100
	a_am_peak_pos	0.076		l_am_peak_pos	0.342		l_am_peak_pos	0.449
0	l_am_peak_pos	0.329	1	l_night_pos	0.335	2	l_night_pos	0.512
	l_night_pos	0.364		frwy_morning_f10	0.024		frwy_morning_f10	0.071
	frwy_morning_f10	0.033		local_morning_f10	0.017		local_morning_f10	0.112
	local_morning_f10	0.042		artrl_morning_f15	0.009		artrl_morning_f15	0.014
	artrl_morning_f15	0.002		artrl_night_f20	0.188		artrl_night_f20	0.143
	artrl_night_f20	0.087		frwy_afternoon_mean	0.121		frwy_afternoon_mean	0.190
	frwy_afternoon_mean	0.035		frwy_morning_dec4	0.007		frwy_morning_dec4	0.018
	frwy_morning_dec4	0.003		artrl_morning_dec4	0.040		artrl_morning_dec4	0.611
	artrl_morning_dec4	0.008		local_night_dec4	0.004		frwy_afternoon_dec8	0.047
	local_night_dec4	0.002		frwy_afternoon_dec8	0.008		unfamiliar	0.427
	frwy_afternoon_dec8	0.010		unfamiliar	0.501		Turns	0.184
	unfamiliar	0.050		Turns	0.526		Previous_crash_location	0.012
	Turns	0.389		Previous_crash_location	0.033		-	-
	Previous_crash_location	0.052		-	-		-	-

Cont'd

Step	Variables	Sig. of F to Enter	Step	Variables	Sig. of F to Enter	Step	Variables	Sig. of F to Enter
3	frwy_night_dist	0.875	4	frwy_night_dist	0.716	5	frwy_night_dist	0.999
	artrl_pm_peak_dist	0.630		artrl_pm_peak_dist	0.603		artrl_pm_peak_dist	0.678
	outside_afternoon_dist	0.003		frwy_pm_peak_duration	0.316		frwy_pm_peak_duration	0.346
	frwy_pm_peak_duration	0.565		artrl_am_peakduration	0.482		artrl_am_peakduration	0.395
	artrl_am_peakduration	0.778		outside_am_peakduration	0.179		outside_am_peakduration	0.147
	outside_am_peakduration	0.311		outside_nightduration	0.640		outside_nightduration	0.810
	outside_nightduration	0.871		f_morning_mean	0.744		f_morning_mean	0.250
	f_morning_mean	0.853		f_night_mean	0.443		f_night_mean	0.811
	f_night_mean	0.356		a_am_peak_delta	0.015		a_am_peak_delta	0.132
	a_am_peak_delta	0.039		f_pm_peak_pos	0.463		f_pm_peak_pos	0.805
	f_pm_peak_pos	0.447		a_am_peak_pos	0.055		a_am_peak_pos	0.557
	a_am_peak_pos	0.065		l_am_peak_pos	0.402		l_am_peak_pos	0.954
	l_am_peak_pos	0.396		l_night_pos	0.632		l_night_pos	0.665
	l_night_pos	0.790		frwy_morning_f10	0.194		frwy_morning_f10	0.651
	frwy_morning_f10	0.214		local_morning_f10	0.274		local_morning_f10	0.838
	local_morning_f10	0.277		artrl_morning_f15	0.005		artrl_night_f20	0.612
	artrl_morning_f15	0.021		artrl_night_f20	0.063		frwy_afternoon_mean	0.054
	artrl_night_f20	0.094		frwy_afternoon_mean	0.077		frwy_morning_dec4	0.004
	frwy_afternoon_mean	0.186		frwy_morning_dec4	0.008		artrl_morning_dec4	0.707
	frwy_morning_dec4	0.034		artrl_morning_dec4	0.466		frwy_afternoon_dec8	0.061
	artrl_morning_dec4	0.555		frwy_afternoon_dec8	0.093		unfamiliar	0.810
	frwy_afternoon_dec8	0.069		unfamiliar	0.992		Turns	0.909
	unfamiliar	0.579		Turns	0.859		-	-
	Turns	0.117		-	-		-	-

Cont'd

Step	Variables	Sig. of F to Enter	Step	Variables	Sig. of F to Enter	Step	Variables	Sig. of F to Enter
6	frwy_night_dist	0.996	7	frwy_night_dist	0.864	8	frwy_night_dist	0.913
	artrl_pm_peak_dist	0.401		artrl_pm_peak_dist	0.167		artrl_pm_peak_dist	0.443
	frwy_pm_peak_duration	0.807		frwy_pm_peak_duration	0.852		frwy_pm_peak_duration	0.794
	artrl_am_peakduration	0.621		artrl_am_peakduration	0.323		artrl_am_peakduration	0.888
	outside_am_peakduration	0.105		outside_am_peakduration	0.042		outside_nightduration	0.823
	outside_nightduration	0.814		outside_nightduration	0.776		totalduration	0.837
	f_morning_mean	0.511		totalduration	0.219		f_morning_mean	0.297
	f_night_mean	0.790		f_morning_mean	0.437		f_night_mean	0.945
	a_am_peak_delta	0.057		f_night_mean	0.831		a_am_peak_delta	0.066
	f_pm_peak_pos	0.920		a_am_peak_delta	0.045		f_pm_peak_pos	0.770
	a_am_peak_pos	0.605		f_pm_peak_pos	0.939		a_am_peak_pos	0.435
	l_am_peak_pos	0.970		a_am_peak_pos	0.520		l_am_peak_pos	0.838
	l_night_pos	0.408		l_am_peak_pos	0.916		l_night_pos	0.365
	frwy_morning_f10	0.850		l_night_pos	0.358		frwy_morning_f10	0.909
	local_morning_f10	0.706		frwy_morning_f10	0.952		local_morning_f10	0.633
	artrl_night_f20	0.896		local_morning_f10	0.526		artrl_night_f20	0.773
	frwy_afternoon_mean	0.324		artrl_night_f20	0.866		frwy_afternoon_mean	0.215
	artrl_morning_dec4	0.392		frwy_afternoon_mean	0.234		artrl_morning_dec4	0.400
	frwy_afternoon_dec8	0.241		artrl_morning_dec4	0.493		frwy_afternoon_dec8	0.274
	unfamiliar	0.788		frwy_afternoon_dec8	0.286		unfamiliar	0.667
	Turns	0.828		unfamiliar	0.852		Turns	0.822
	-	-		Turns	0.889		-	-

REFERENCES

1. Subramanian, R., "Motor Vehicle Traffic Crashes as a Leading Cause of Death in the U.S., 2002", NHTSA Technical Report, DOT HS 809 843, 2005.

2. Georgia Department of Motor Vehicle Safety, "Crash Analysis, Statistics, and Information Notebook 1996-2003", CASI Report. The State of Georgia, 2004.

3. Gettman, D. and L. Head, "Surrogate Safety Measures from Traffic Simulation Models", Final Report. Federal Highway Administration (FHWA), 2003.

4. Crettenden, A. and A. Drummond, "The Young Driver Problem versus the Young Problem Driver - A Review and Crash Data Analysis", Federal Office of Road Safety, 1994.

5. Ogle, J., "Quantitative Assessment of Driver Speeding Behavior Using Instrumented Vehicles", Doctoral Dissertation, The Department of Civil Engineering, The Georgia Institute of Technology, Atlanta, 2005.

6. Jonah, B.A., "Age Differences in Risky Driving. Health Education Research - Theory and Practice", Vol. 5, No. 2, pp. 139-149, 1990.

7. Litman, T. and S. Fitzroy, "Safe Travels - Evaluating Mobility Management Traffic Safety Impacts", Victoria Transport Policy Institute, Canada, 2005.

8. Solomon, D., "Accident on Main Rural Highways Related to Speed, Driver, and Vehicle", US Department of Commerce & Bureau of Public Road, 1964.

9. Janke, M., "Accidents, Mileage, and the Exaggeration of Risk. Accident Analysis and Prevention", Vol. 23, No. 3, pp. 183-188, 1991.

10. Greenshields, B.D. and F.N. Platt, "Development of a Method of Predicting High-Accident and High-Violation Drivers", Journal of Applied Psychology, Vol. 51, No. 3, pp. 205-210, 1967.

11. Horberry, T., L. Hartley, K. Gobetti, F. Walker, B. Johnson, S. Gersbach, and J. Ludlow, "Speed choice by drivers: The issue of driving too slowly", Ergonomics, Vol. 47, No. 14, pp. 1561-1570, 2004.

12. Progressive Insurance, website: https://tripsense.progressive.com. Accessed April 2005.

13. Guensler, R., A. Amekudzi, J. Williams, S. Mergelsberg, and J. Ogle, "Current States Regulatory Support for Pay-As-You-Drive Automobile Insurance Options", Journal of Insurance Regulation, Vol. 21, No. 3, 2003.

14. Massie, D., K. Campbell, and A. Williams, "Traffic Accident Involvement Rates by Driver Age and Gender", Accident Analysis and Prevention, Vol. 27, No. 1, pp. 73-87, 1995.

15. Cooper, P.J., "The relationship between speeding behavior (as measured by violation convictions) and crash involvement", Journal of Safety Research, Vol. 28, No. 2, pp. 83-95, 1997.

16. Li, H., "Investigating Morning Commute Route Choice Behavior Using Global Positioning Systems and Multi-day Travel Data", Doctoral Dissertation, The Department of Civil Engineering, The Georgia Institute of Technology, Atlanta, 2004.

17. Jun, J., R. Guensler, and J. Ogle, "Smoothing Methods Designed to Minimize the Impacts of GPS Random Errors on Travel Distance, Speed, and Acceleration Profile Estimates" Transportation Research Record, 2006. (In press).

18. Jun, J., J. Ko, S. Yoon, and R. Guensler, "Impacts of Acceleration Calculation Methods on Onroad Vehicle Engine Power Estimation", CD-ROM, Air & Waste Management Association Annual Meeting (AWMA), New Orleans, 2006.

19. Hines, W.W., D.C. Montgomery, D.M. Goldsman, C.M. Borror, "Probability and Statistics in Engineering", John Wiley & Sons, Inc., 2002.

20. Wolfe, A.C., "The Concept of Exposure to the Risk of a Traffic Accident and an Overview of Exposure Data Collection Methods", Accident Analysis and Prevention, Vol. 14, No. 5, pp. 337-340, 1982.

21. Drummond, A., "Young Driver Research Program: A Technical and Strategic Overview of Exposure Reduction Measures as a Means of Reducing Young Driver Crashes", Federal Office of Road Safety: Canberra, ACT., 1994.

22. Gebers, M., "The Contribution of Driving Exposure in the Prediction of Traffic Accidents", Research Notes, Department of Motor Vehicles, California, 2003.

23. Kirk, A. and N. Stamatiadis, "Evaluation of the Quasi-Induced Exposure", US DOT Transportation Center: Southeastern Transportation Center, 2001.

24. Kim, K., L. Lawrence, J. Richardson, L. Li, "Personal and Behavioral Predictors of Automobile Crash and Injury Severity", Accident Analysis and Prevention, Vol. 27, No. 4, pp. 469-481, 1995.

25. White, S. and A. Nelson, "Some Effects of Measurement Errors in Estimating Involvement Rate As a Function of Deviation From Mean Traffic Speed", Journal of Safety Research, Vol. 2, No. 2, pp. 67-72, 1970.

26. O'Day, J., "Accident Data Quality. National Cooperative Highway Research Program: Synthesis of Highway Practice 192", Transportation Research Board: Washington D.C., 1993.

27. Shinar, D., J. Treat, and S.T. McDonald, "The validity of police reported data", Accident Analysis and Prevention, Vol. 15, pp. 175-191, 1983.

28. Wahlberg, A.A., "The Stability of driver acceleration behavior, and a replication of its relation to bus accident", Accident Analysis and Prevention, Vol. 36. pp. 83-92, 2000.

29. Klauer, S.G., T.A. Dingus, V.L. Neale, J.D. Sudweeks, D.J. Ramsey, "The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data", NHTSA, April 2006.

30. Ogle, J., R. Guensler, W. Bachman, M. Koutsak, J. Wolf, "Accuracy of Global Positioning System for Determining Driver Performance Parameters", Transportation Research Board, No. 1818, pp. 12-24, 2002.

31. Trimble, "AgGPS 214 High-Accuracy Receiver, Centimeter Precision in Agriculture", Technical Note, 2000.

32. Lin, L., "Application of GPS RTK and Total Station System on Dynamic Monitoring Land Use", National Science Council: Taiwan, Republic of China, NSC 92-2415-H-004-025.

33. Hauer, E. and A.S. Hakkert, "Extent and Some Implications of Incomplete Accident Reporting", Transportation Research Record, Vol. 1185, pp. 1-10, 1988.

34. Zito, R., G. D'Este, and M.A.P. Taylor, "Global Positioning Systems in the Time Domain: How Useful a Tool for Intelligent Vehicle-Highway Systems", Transportation Research C, Vol. 3, No. 4, pp. 193-209, 1995.

35. Jackson, E., L. Aultman-Hall, B.A. Holmen, and J. Du, "Evaluating the Ability of Global Positioning System Receivers to Measure Real-World Operating Mode for Emissions Research", CD-ROM, Transportation Research Board. Washington, D.C., 2005.

36. Benz, R.J. and M.A. Ogden, "Development and Benefits of Computer-Aided Travel Time Data Collection", Transportation Research Record, No. 1551, pp. 1-7, 1996.

37. Hallmark, S.L. and R. Guensler, "Comparison of Speed-Acceleration Profiles from Field Data with NETSIM Output for Modal Air Quality Analysis of Signalized Intersections", Transportation Research Record, No. 1664, pp. 40-46, 1999.

38. Burden, R.L. and J.D. Faires, "Numerical Analysis", PWS-Kent Publishing Company, 1989.

39. Sin, H.G., "Field Evaluation Methodology for quantifying Network-Wide Efficiency, Energy, Emission, and Safety Impacts of Operational-Level Transportation Projects", School of Civil and Environmental Engineering, Virginia Polytechnic Institute and State University: Virginia, 2001.

40. Rakha, H., I. Lucic, S. Demarchi, J. Setti, and M.V. Aerde, "Vehicle Dynamics Model for Predicting Maximum Truck Acceleration Levels", Journal of Transportation Engineering, Vol. 127, No. 5, pp. 418-425, 2001.

41. Rakha, H., F. Dion, and H.G. Sin, "Field Evaluation of Energy and Emission Impacts of Traffic Flow Improvement Projects using GPS Data: Issues and Proposed Solutions", CD-ROM, the 80th Annual Meeting of the Transportation Research Board, Washington, D.C., 2001.

42. Yoon, S., H. Li, J. Jun, J. Ogle, R. Guensler, M. Rodgers, "A Methodology for Developing transit Bus Speed-Acceleration Matrices to be Used in Load-Based Mobile Source Emissions Models", Transportation Research Record, No. 1941, 2005.

43. Jun, J., J. Ogle, and R. Guensler, "Trade-off between the Accuracy Level and the Data Size Caused by Filtering with Number of Satellites and PDOP Values of GPS Data", Manuscript, Georgia Institute of Technology, 2006.

44. Gates, T.J., S.D. Schrock, and J.A. Bonneson, "Comparison of Portable Speed Measurement Devices", CD-ROM, Transportation Research Board, Washington D.C., 2004.

45. Hastie, T., R. Tibshirani, and J. Friedman, "The Elements of Statistical Learning: Data Mining, Inference, and Prediction", Springer, 2001.

46. Martinez, W.L. and A.R. Martinez, "Computational Statistics Handbook with Matlab", Chapman and Hall/CRC, 2002.

47. Welch, G. and G. Bishop, "An Introduction to the Kalman Filter", The University of North Carolina and ACM, Inc., 2001.

48. Bashi, A.S., "A Practitioner's Short Guide to Kalman Filtering", website: http://www.uno.edu/~SAGES/publications/PractitionersKalman.PDF, Accessed May 2005.

49. Simon, D., "Kalman Fitering", Embedded Systems Programming, 2001.

50. SiRF. Technical Specification for SiRF Star II Architecture. website: http://www.tdc.co.uk/gps/gps_receivers_leadtek.htm. Accessed July 2005.

51. TireRack. Online Performance Source, The Tire Rack, website: http://www.tirerack.com/tires/tiretech. Accessed April 2005.

52. Miller, I., J.E. Freund, and R.A. Johnson, "Probability and Statistics for Engineers", Prentice Hall, 1990.

53. Washington, S., M. Karlaftis, and F. Mannering, "Statistical and Econometric Methods For Transportation Data Analysis", Chapman & Hall/CRC, 2003.

54. Hair, J., R. Anderson, R. Tatham, W. Black, "Multivariate Data Analysis", Prentice Hall, 1998.

55. Tabachnick, B.G. and L.S. Fidell, "Using Multivariate Statistics", 2nd Edition, HarperCollins, 1989.

56. Lachenbruch, P.A., "Discriminant Analysis", Hafner, 1975.

57. Klecka, W.R., "Discriminant Analysis. Quantitative Applications in the Social Sciences Series", No. 19, Sage Publications, 1980.

58. Cohen, J., P. Cohen, S.G. West, L.S. Aiken, "Applied Multiple Regression/Correlation Analysis for the Behavioral Science", 2nd Edition, Lawrence Erlbaum, 2003.

59. Brookhuis, K.A., G. De Vries, and D. De Waard, "The effect of mobile telephoning on driving performance", Accident Analysis and Prevention, Vol. 23, No. 4, pp. 309-316, 1991.

60. Jun, J., R. Guensler, and J. Lee, "Potential Use of Operating Mode Durations and Dynamic Ratios in Analyzing Driver Behavior Using Instrumented Vehicle Data", CD-ROM, The 85th Annual Meeting of Transportation Research Board, Washington, D.C., 2006.

61. Hallmark, S.L., "Analysis and Prediction of Individual Vehicle Activity for Microscopic Traffic Modeling", Doctoral Dissertation, Civil and Environmental Engineering, The Georgia Institute of Technology, Atlanta, 1999.

62. GAO, "Highway Safety - Research Continues on a Variety of Factors That Contribute to Motor Vehicle Crashes", United States General Accounting Office, GAO-03-436, 2003.

63. Peper, A.C., Automotive History and Information, website: http://www.obd-2.com, Accessed May 2003.

64. NCVEC, National Center for Vehicle Emissions Control and Safety, website: http:www.obdiicsu.com, Accessed May 2003.

65. Bordoff, F., Onboard Diagnostics (OBD II), website: http://www.familycar.com/OBD.htm, Accessed May 2003.

66. Carley, L., "Understanding OBD II: Past, Present & Future", website: http://members.aol.com/carleyware/library/us796obd.htm, Accessed April 2003.

67. SAE, "On-Board Diagnostics for Light and Medium Duty Vehicle Standards Manual", SAE-HS-3000, Society of Automotive Engineers, 1999.

68. ERG, "On-Board Diagnostics Outreach and Communication Plan", Lexington, 2001.