Evaluating the Impacts of Driver Behavior in the Speed Selection Process and the Related Outcomes

EVALUATING THE IMPACTS OF DRIVER BEHAVIOR IN THE SPEED SELECTION PROCESS AND THE RELATED OUTCOMES

A Dissertation Presented

by

COLE D. FITZPATRICK

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements of the degree of

DOCTOR OF PHILOSOPHY

FEBRUARY 2017

Civil and Environmental Engineering

EVALUATING THE IMPACTS OF DRIVER BEHAVIOR IN THE SPEED SELECTION PROCESS AND THE RELATED OUTCOMES

A Dissertation Presented

by

COLE D. FITZPATRICK

Approved as to style and content by:

_____
Michael A. Knodler, Chair

_____
Eleni Christofa, Member

_____
Eric Gonzales, Member

_____
Siby Samuel, Member

_____
Richard Palmer, Department Head
Civil & Environmental Engineering

ACKNOWLEDGEMENTS

This research was funded in part through fellowships from the National Science Foundation (NSF) I-CORPS program, Isenberg School of Management, the Dwight D. Eisenhower Transportation Fellowship Program, and Safety Research Using Simulation (SAFER SIM) University Transportation Center (UTC) at the University of Iowa. Funding for the UTC Program is provided by the Office of Assistant Secretary for Research and Innovation (OST-R) of the United States Department of Transportation (USDOT).

First and foremost, I would like to thank my advisor, Dr. Michael Knodler, for his support, encouragement and guidance through this process. From day one at UMass, Mike always had my best interests at heart and my success as a student would not have been possible without him. Additionally, I want to extend my gratitude to my committee members, Dr. Eleni Christofa, Dr. Eric Gonzales, and Dr. Siby Samuel, who helped guide my research and have offered mentorship, both professional and personal. I am grateful for the unique opportunity to have participated in I-Corps with my team members, Dr. John Collura and Dr. Paul Shuldiner. Thanks to all of the students who have worked with me on various projects including Ian McKinnon, Keith Wenners, Francis Tainter, Manny Pires, Radha Gomez, Saritha Rakasi, Curt Harrington, Daniel Abrams, and Yue Tang. I want to express my appreciation for the UMass Safe group members, Jennifer Gazzillo and Robin Riessman, who have helped me pursue projects related to crash data. Thank you to my parents for supporting me throughout my 20 plus years of schooling. I want to thank my wife, Katrina, for convincing me to apply to grad school and her support when I wanted to return to UMass for a Ph.D. Finally, thank you to my 5-month old daughter, Logan, for providing the ultimate motivation to finish my dissertation.

ABSTRACT

EVALUATING THE IMPACTS OF DRIVER BEHAVIOR IN THE SPEED SELECTION PROCESS AND THE RELATED OUTCOMES

FEBRUARY 2017

B.S.C.E. OREGON STATE UNIVERSITY

M.S.C.E. UNIVERSITY OF MASSACHUSETTS, AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS, AMHERST

Directed by: Michael A. Knodler Jr.

In the United States, traffic crashes claim the lives of 30,000 people every year and is the leading cause of death for 5-24 year olds. Driver error is the leading factor in over 90 percent of motor vehicle crashes, with the roadway and the vehicle each only accounting for about 2 percent of crashes. In the United States, nearly a third of fatal crashes are due to speeding and therefore, a critical step in improving traffic safety is research aimed to reduce speeding, such as crash data analysis, outreach campaigns, targeted enforcement, and understanding speed selection. In this dissertation, a multi-faceted approach was taken to improve roadway safety by examining the speeding-related crash designation, improving speed limit setting practices, and understanding the causes of speeding. Multiple experiments were conducted under this overarching goal. These experiments included an analysis of speeding-related crashes in Massachusetts, a naturalistic driving study, and a driving simulator study which investigated the causes of speeding. Collectively, the findings from these experiments can expand upon existing speed prediction models, improve crash data influence speed limit setting practices, guide speed management programs such as speed enforcement, and be used in public safety outreach campaigns.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1
EXECUTIVE SUMMARY

This dissertation consisted of four projects, each relating to traffic safety, specifically focused upon the topic of speed. The first project explored the commercialization potential of the UMass Safe Traffic Data Warehouse through the NSF I-Corps program. During this semester long course, over one hundred crash data stakeholders were interviewed. From these interviews, tremendous insights were gained as to how crash data was collected, utilized, and distributed within the Commonwealth of Massachusetts. During the I-Corps course, our team developed and iterated upon a business model for our hypothetical start-up company. While our team ultimately decided upon a "No-go" decision for incorporating our hypothetical startup, the information learned from these 100+ interviews provided a foundational understanding of how crash was used and what data quality deficiencies existed.

Using the knowledge gained from I-Corps, the second project was designed around one of these data quality issues, the speeding-related crash designation. In this project, logistic regression models were built to generate the probability of each crash being speeding-related. The accuracy of these models were then evaluated by conducting a double-blind crash narrative review of 600 crashes strategically sampled from the logistic regression models. The results of this review indicated that the model did perform well at identifying crashes which should have been designated as being speeding-related but were not originally designated as such. After this, a more detailed review of crashes with the "Driving too fast for conditions" (DTFFC) driver contributing code was conducted. This review indicated that the DTFFC code was being used to indicate driving too fast for: weather conditions, traffic conditions, or roadway geometry. From this finding, the

recommendation was made to split the DTFFC driver contributing code into three separate codes to give engineers more detail as to the nature of the safety problems on a roadway.

From I-Corps, we also learned that engineers use not only crash data when designing safety-oriented projects, but also make use of field data. One type of field data commonly used is assessing vehicle speeds. Based upon this knowledge, and the knowledge of known deficiencies in the speeding-related crash designation, the third project sought to develop a method to collect continuous speeds on a roadway. This method, which involved equipping trial drivers with a smartphone app, would allow engineers to target speed mitigation measures at the specific areas with the most extreme speed concerns. Additionally, this continuous speed method was compared to traditional spot speed methods of data collection and tested for use in USLimits2, an expert system for recommending rational speed limits. While only four locations were tested with USLimits2, ultimately a method could be designed to automate the system so that a continuous speed limit recommendation could be generated from continuous speed data. Although minimum segment lengths exist for the size of a speed zone, these continuous speed limit recommendations could be used as a tool by engineers to select the most appropriate speed limit for a roadway and to also place advisory speed signs.

After reviewing speeding-related crash data to identify problems and make recommendations, we then looked to improve speed data collection by comparing traditional spot speed methods to continuous speed data collection. In the final project, we wanted to examine, in a laboratory setting, how one of the previously identified main causes of speeding, being late, influenced driver behavior. This investigation was achieved by conducting a driver simulator experiment with 36 participants. These participants were

split into three groups, a control group with no time pressures, an experimental group with an easily achievable time goal (*Hurried)*, and an experimental group with a difficult to achieve time goal (*Very Hurried)*. Consistent with the previous survey studies, the *Very Hurried* drivers selected higher speeds, accelerated faster, and made more aggressive maneuvers than the control group. Interestingly, *Hurried* drivers exhibited all of these differences as well, although unlike the *Very Hurried* group, these different were not significant from the control.

Ultimately, this dissertation has potential impacts on various transportation segments. Engineers could make use of a more accurate speeding-related designation and an improved method to collect speeds for targeting design locations and setting speed limits. Law enforcement officers could see value in the recommendations generated from the crash narrative review and could also apply a continuous speed collection technique to target enforcement to locations where the largest safety benefits could be achieved. Finally, vehicle manufacturers could make use of the findings from the driving simulator experiment. By understanding how time pressures impact driving behavior, autonomous vehicles could be programmed to understand the user's perception of time and select speeds which are likely to keep the user in autonomous mode rather than switching to manual driving because they are running late.

CHAPTER 2
INTRODUCTION

2.1 Problem Statement

In the United States, traffic crashes claim the lives of 30,000 people every year *(1)* and is the leading cause of death for 5-24 year olds *(2)*. Globally, 1.2 million people die in traffic crashes *(3)*. Driving consists of three components: the roadway, vehicle, and driver. Driver error is the leading factor in over 90 percent of motor vehicle crashes, with the roadway and the vehicle each only accounting for about two percent of crashes *(4)*. As such, designing roadways and vehicles that minimize the effects of driver error is a critical step in improving traffic safety.

Speed is one of the most important factors in traffic safety. In the United States, "the driver behavior of exceeding the posted speed limit or driving too fast for conditions" is designated as "speeding-related", as defined by the National Highway Traffic Safety Administration (NHTSA). As speed increases, the risk of a crash increases greatly in both rural and urban areas *(5)*. As does the severity of crashes involving pedestrians, *(6)* and not involving pedestrians *(7)*. Nearly a third of fatal crashes in the United States are designated as "speeding-related" *(8)*, highlighting the continued need to study crash data quality, speed capturing techniques, speed limit setting practices, and human factors in order to mitigate the frequency and severity of speeding-related crashes.

2.2 Overarching Dissertation Objectives

Based upon the identified problem statement, the overarching goal of this dissertation is to investigate speed selection and its impact on transportation. There are many specific methods aimed to reduce speeding, such as crash data analysis, outreach campaigns, targeted enforcement, and understanding speed selection. Within these existing

methods to reduce the safety impacts of speeding, there is a need for innovative approaches to data quality, speed data collection, and its utilization. In this dissertation, a multi-faceted approach was taken to improve roadway safety by examining the speeding-related crash designation, improving speed limit setting practices, and understanding the causes of speeding. Within the framework of this overarching goal, a series of research objectives has been developed. Background relating to each of the four objectives is found in Chapter 2.

Objective 1: Investigate how crash data is collected, distributed, and utilized within the state of Massachusetts. Data is the foundation of good decision making and it is hypothesized that crash data is utilized differently depending on the user. Understanding these different uses of crash data is an essential first step in studying transportation safety.

Objective 2: Improve the classification of speeding-related crashes. Classification of a crash as speeding-related or not speeding-related is at the discretion of the officer responding to the scene. The responding officer fills out a crash report which includes crash details, a narrative of what occurred, and a crash diagram. Previous analyses of speeding-related crashes show a need for better classification for these types of crashes. An improved classification of speeding-related crashes, would allow engineers to more accurately direct highway safety improvement funds and enable law enforcement to more efficiently target their speed-management campaigns.

Objective 3: Develop a method to capture continuous speed profiles. Current methods of speed data collection, while generally cost effective, are only accurate at a finite number of locations. MassDOT acknowledges this deficiency and states that “it would be ideal to have speed checks at an infinite number of locations so that the 85th percentile

speed could be computed at all points." *(9)* Previously, infeasible, recent advances in smartphone technology enable this type of data collection process for setting speed limits. Continuous speed profiles may improve speed limit setting practices and could also be an input for autonomous vehicle speed selection under free flow conditions.

Objective 4: Examine how a driver's "perception of time" influences their driving behavior. Previous research on hurried driving has indicated that drivers and pedestrian engage in riskier behavior when under time pressures. However, most of these insights have been qualitative. Driving simulation with positive and negative incentives would allow for the quantification of the effects from being late or in a hurry. In order to reduce speeding-related crashes, it is necessary to understand the psychological reasons behind why drivers consciously, or unconsciously, choose to speed.

2.3 Dissertation Organization

This dissertation focuses upon four projects which directly investigate speed's effect on traffic safety. Chapter 2, provides a background on previous work that is relevant to the four projects. Chapters 3-6 each contain one of the four projects. Within each chapter the specific motivation for that project is discussed, followed by the methods, results of the study, discussion of significant findings and limitations, and a conclusion. Chapter 7 contains the overall conclusions from this dissertation work along with possible areas of future work relating to each project.

CHAPTER 3
BACKGROUND

A literature review was conducted throughout the dissertation process. Any research regarding speeding-related crashes, speed data collection techniques, and the causes of speeding was reviewed to better understand the gaps in the literature to mold this research to maximize impact.

3.1 Speeding-Related Crash Data

Nearly a third of fatal crashes in the United States are designated as "speeding-related", which is defined by the National Highway Traffic Safety Administration (NHTSA) as "the driver behavior of exceeding the posted speed limit or driving too fast for conditions." *(8)*. This speeding-related crash designation is critical as the American Association of State Highway Transportation Officials (AASHTO) Strategic Highway Safety Plan recommends the use of targeted conventional speed enforcement as a strategy to reduce speeding-related crashes *(10)*. This type of strategy requires accurate data related to roadways with a high frequency of speeding-related crashes. However, an inherent challenge with the speeding-related designation is the manner in which it is derived. The law enforcement officer who responds to a crash and completes the subsequent crash report must select one or more *Driver Contributing Codes* (DCCs) which are supposed to explain why the crash occurred. This discretionary decision is often made following an investigation of the scene and interviews with the motor vehicle operator(s) and any witnesses.

Numerous studies have investigated speeding-related crashes, and while none investigated the reliability of the speeding-related designation, each acknowledged the limitations of the designation. For example, the Oregon Department of Transportation

conducted a study where high speeding-related crash locations were identified for possible mitigation. In their discussion they note, "the analysis relies on crash reports, which are subject to the interpretations of a variety of individuals completing the crash report form. Specifically, the fact that a crash has been identified as speeding-related is not based on a scientific analysis, and may be the result of opinion or best judgment" *(11)*.

The Federal Highway Administration (FHWA) funded a study which developed a speeding-related typology and compared data from two different states which used differing definitions for speeding-related crashes. The study noted several crash characteristics which were more commonly found in crashes designated as speeding-related. Additionally, they concluded that the NHTSA definition was most appropriate for the speeding-related classification. Finally, the report cautioned against the type of analyses which was conducted in Oregon stating, "it is difficult to know whether an identified variable shows a true higher association with speed or whether the association shown is partially due to an officer bias" and "treatment programs oriented to these factors may not be as successful as if oriented to other characteristics where such a bias is not expected" *(12)*.

In 2014, a Speed Management Plan was developed jointly by NHSTA, FHWA, and the Federal Motor Carrier Safety Administration (FMCSA). The plan sought to reduce speeding-related fatalities and injuries and improve the safety experience for all road users. While the plan recommends a data driven approach using the speeding-related designation, it also cautioned that "the precise role of speeding in crashes can be difficult to ascertain, as speeding is often defined in broad terms. Further, the determination of whether speeding

was involved in a fatal crash is often based on the judgment of the investigating law enforcement officer." *(13)*

The crash narrative is the responding officer's written account of what occurred before, during, and after the crash. Crash narratives can be used to more thoroughly investigate the cause of a crash as crash narratives often provide information beyond what is captured in the pre-defined fields of the crash report. Examples highlighting the utility of crash narratives are present throughout the traffic safety related literature. In one of the more in depth studies, McKnight and McKnight reviewed 2,000 crash narratives to determine if crashes involving younger drivers were due to carelessness or inexperience *(14)*. Crash narratives have also been utilized previously to conduct in-depth investigations of crashes involving military vehicles *(15)*, work zone crashes *(16)*, helmet status in motorcycle crashes *(17)*, and distraction-related crashes *(18)*

3.2 Speed Data Collection Techniques

There are many ways to conduct a speed study, each with its own strengths and weaknesses. An objective of this dissertation was to compare a new data collection technique with some existing methods. Existing methods of speed data collection include:

- Pneumatic Tubes with Automated Traffic Recorders (ATRs)
- RADAR/LiDAR Speed Guns
- Probe Vehicles
- Inductive Loops
- Side-fire RADAR Units
- On Board Diagnostic (OBD) Black Boxes
- GPS Smartphone Apps

ATRs capture volume, vehicle class, gap and speed data over long time periods. ATRs are commonly used to capture speed data over one week and to measure average annual daily traffic (AADT). ATRs can accurately capture vehicle speeds *(19)* and do not

influence driver behavior *(20)*, but cannot easily distinguish whether or not a vehicle is traveling in free-flow conditions. As mentioned above, an ATR is installed in a single location. If multiple data collection locations are desired, then multiple ATR installations are required, which can be costly.

RADAR and LiDAR speed sensors are the preferred method of speed detection by law enforcement as they can provide the speed of a selected vehicle. They differ in that a RADAR gun can be easily used while moving, while a LiDAR gun functions more effectively while stationary *(21)*. However, LiDAR guns are more effective at longer ranges and can be more accurate as a laser sight allows the user to know exactly which vehicle is being captured. While other states stipulate larger samples, in Massachusetts a spot speed study using a RADAR or LiDAR gun involves an inconspicuous observer capturing a sample set of 100 vehicle speeds in free-flow conditions *(9)*. On rural roads with low volumes this can often take several hours to collect. If more locations are needed, speed studies using a RADAR or LiDAR gun can be costly in terms of person-hours. Additionally, the LiDAR gun itself costs $2000-$3000.

Inductive loops installed consecutively in a roadway provide a more permanent method to capture vehicle speeds. Loops use magnetic fields to detect the presence of passing vehicles and typically cost $1000 per installation before traffic control expenses *(22)*. A single inductive loop can be used to calculate vehicle speeds but require algorithms to be installed on the traffic signal controller *(23)*.

Side-fire RADAR units are portable devices which can be installed on utility poles and can capture multiple lanes of bi-directional traffic speeds. The units are easy to install and capture speeds accurately, but require a clear line of sight and measuring the geometry

of the roadway prior to installation. Additionally, the high cost of the unit, $4000-$5000, may make this form of data collection prohibitive for smaller agencies *(24)*.

Trial runs, or probe drives, are usually conducted in addition to one of the methods described above. MassDOT's guidelines for probe drives stipulate that three drivers are to drive the portion of roadway being studied with an observer seated directly behind them recording their speed every 1/10th of a mile *(9)*. Probe drives are conducted in order to provide a more complete speed profile than the spot speed observations. However, the effect of the passenger observer is significant on the driver's performance as they feel like they are being studied. This effect is lessened when the probe drive is monitored via vehicle instrumentation. The 100-Car Naturalistic Driving Study found that participants had a lower incident rate in the first hour of the study, but quickly forgot they were being monitored and resumed normal driving behavior *(25)*. Probe drives provide more granular data than the previous methods but are not as granular as the following two methods.

There are various devices which plug into a vehicle's OBD port and function similar to an airplane's black box. An OBD black box can capture the vehicle's GPS position, speed, steering wheel position and RPM one to three times per second *(26)*. The data is a large step up from trial runs in terms of accuracy and OBD devices have less of an impact on driver behavior. However, these devices are similar in cost to LiDAR guns and require after-market installation in vehicles. Additionally, these devices cannot distinguish when the vehicle is traveling in free-flow conditions.

Smartphone apps can have similar functionality to an OBD black box by recording a user's GPS position and speed using the phone's built-in location services. Specifically in this study, we used Ubipix, a smartphone app that captures speed and position every

second and combines that with video captured from the smartphone *(27)*. Users upload captured data to a cloud-based database where it can be shared publicly or kept private, depending on the users' preference. The data is displayed graphically on their platform and the user can place tags at certain locations such as when free-flow conditions are or are not present. Ubipix is significantly less expensive than OBD devices as there is no cost for the app and the user pays on an as needed basis. Ubipix implementation is cost-effective and has a minimal learning curve associated with data processing, which will be described further in the methodology.

3.3 Speed Limit Setting Practices

Engineers use an assortment of traffic control devices to communicate simple messages to vehicle drivers, with speed limit signage being the primary mechanism for conveying appropriate roadway speeds to the motoring public. More specifically, speed limits are the front lines of speed management and serve as a valuable tool in promoting roadway safety. Speed limits that are too low lead to high non-compliance rates *(28)*. By comparison, speed limits violate driver expectancy if they are set above safe operating speeds. Speed limits should reflect the roadway environment and driver expectation. In 1998, the American Association of State Highway and Transportation Officials (AASHTO) published its Strategic Highway Safety Plan which set a target of halving fatalities within two subsequent decades. Within the AASHTO plan, "Setting Appropriate Speed Limits" was identified as an objective to reduce speed-related crashes *(29)*.

To set appropriate speed limits it is important to understand the differences in the designated design speed, inferred design speed, and operating speeds. The designated design speed is defined by AASHTO as "a selected speed used to determine the various geometric design features of the roadway" *(30)*. The inferred design speed differs from the

designated design speed in segments of roadway where all design elements exceed criterion-limiting values *(31)*. For example, if the designated design speed on a roadway sets a minimum sight distance requirement, the inferred design speed would exceed the designated design speed when longer sight distance is present. The inferred design speed could, in theory, be less than the design speed if the road was improperly designed. Often, speed limits are set to the critical inferred design speed, or the segment of roadway where the inferred design speed is at a minimum and most near the designated design speed. This results in operating speeds on the adjacent segments that greatly exceed the posted speed limit, leading to challenges for law enforcement as to how to set a threshold for enforcement.

Over the course of the past decade the concept of rational speed limits has evolved while being promoted on a national level. Rational speed limits are based upon speed data analysis to establish a speed limit that is clear to motorists, provides logical enforcement, and creates a safe roadway environment *(32)*. By this logic, the speed limits on some roadways may be increased or decreased in the effort to improve safety. Various studies have shown that an increased speed limit, combined with enforcement, can lead to fewer speeders, a decrease in standard deviation of speeds, and decreases in crash frequency *(33)*. Education is also critical to implementation, as rational speed limits are more effective when motorists are aware of the increased enforcement *(34, 35)*.

NCHRP Report 500 which provides guidance on the AASHTO Strategic Highway Safety Plan states that a speed limit should depend on four factors: design speed, crash frequencies and outcomes, speed tolerance and enforcement threshold, and finally vehicle operating speed measured as "a range of 85th percentile speeds taken from spot speed

surveys of free-flowing vehicles at representative locations along the highway" *(10)*. Free-flowing conditions exist when drivers are able to choose their desired speed without constraints from other vehicles on the road.

The Federal Highway Administration (FHWA) has taken this a step further with the development of USLimits2, a "web-based expert advisor system designed to assist practitioners in determining appropriate speed limits in speed zones" (Srinivasan et al., 2006, 2008). The inputs include: type of surrounding development, access frequency, road function, crash history, pedestrian activity, and existing vehicle operating speeds. The system takes 85th and 50th percentile speeds from segments that do not have adverse alignments. System guidance suggests that speed data should be taken from a 24-hour weekday period, which differs from many states' guidelines which require a spot speed study of 100-200 free-flow vehicles *(38)*. With either method, the location(s) of data collection is subject to engineering judgement as time, equipment, and cost restraints limit the amount of data collection points.

3.4 Causes of Speeding and Risky Behaviors

A variety of survey studies have been performed to try to determine why people speed. In 2011, NHTSA conducted a nationwide survey of 6,144 households to ask people the reasons why they did, or did not, speed. The survey results included 30% of people admitted to being "speeders" with an additional 40% classifying themselves as "sometime speeders". When asked the reason as to why people sped, the most common response was "I'm Late", which accounted for 35% of all responses. "Emergency/illness" was the next most common, which tallied to 31% of all responses. "In a hurry" and "traffic flow" each accounted for 7% of the responses *(39)*.

Beck et al. conducted a telephone survey of 796 licensed drivers to compare hurried drivers to unhurried drivers, they found that hurried drivers were more likely to admit to risky behaviors such as speeding and not wearing a seat belt *(40)*. This work was followed up by another survey of 769 college students. The results of this survey indicated that hurried drivers were more likely to be frustrated with other drivers, more impatient, more aggressive, and take more risks. Additionally, drivers who self-reported a ticket in the previous month were more likely to be hurried drivers *(41)*. While these surveys point to reasons for speeding, there is a need to quantify how perception of time impacts driver performance.

Additional research involving pedestrians in a hurry has provided further evidence of how time impacts risky behaviors. Zhang et al. built a model to predict pedestrians' likelihood to "red-light-run", or cross when they did not have a crossing signal, in China. One of the significant inputs into their models was whether the pedestrian was in a hurry and was thus unlikely to accept the delay of waiting for the crossing signal *(42)*. Similarly, Charron et al. utilized a pedestrian simulator to see if children perform unsafe crossing maneuvers when they are in hurry. The study, with 80 ten-year-old participants found that the children who were in a hurry more frequently exhibited the risky behaviors of running across the street or not using the pedestrian crossing *(43)*. A driving simulator study would enable a similar analysis of how perception of time impacts drivers' willingness to engage in speeding and other risky behaviors.

To date, only one driving simulator has been conducted which investigates this time phenomenon. Bertola et al. constructed a study which investigated how driver inattention, familiarity, and time pressure affected driving performance on rural two-lane horizontal

curves. The study consisted of 14 total participants, of which 6 of them were subjected to two different time pressure methods. The first was simply a scenario where drivers were to imagine that they were running late for a doctor's appointment. The second method added to that scenario a timer and small ($4) financial incentives for meeting goal completion times. The results indicated that the drivers with the time pressure had a higher mean average speed than the control group. However, possibly due to the small sample size, there was no difference between the methods. The lack of penalties, either for crashes or excessive speeding within the scenarios, may have resulted in a biased result as speeds would go unchecked. Additionally, the only aggressiveness metric that was evaluated was mean average speed across the drive *(44)*. While the results of this study began to quantify how time pressure, or drivers' perception of time, impacts speed choice, there is a need for a more robust driving simulator study which can investigate speed in more detail along with additional driver aggressiveness measures.

CHAPTER 4
INVESTIGATING THE USE OF CRASH DATA AND ITS COMMERCIALIZATION POTENTIAL

4.1 Summary

The UMassSafe Traffic Safety Data Warehouse contains over 15 years of crash data that can be utilized in safety research. In addition to academia, these datasets might be of interest to transportation engineers, insurance companies and police departments. To fully understand crash data in the Commonwealth of Massachusetts and determine the potential commercialization of the UMassSafe Traffic Safety Data Warehouse, our team of John Collura, Principal Investigator (PI); Michael Knodler, Co-PI; Paul Shuldiner, Business Mentor; and Cole Fitzpatrick, Entrepreneurial Lead participated in the National Science Foundation (NSF) I-Corps program. As a result, we gained an understanding of the strengths and deficiencies regarding how crash data is collected, accessed, and utilized in the Commonwealth of Massachusetts. Additionally, a hypothetical business model was developed, which highlighted how key crash data stakeholders could be served by the UMass Safety Data Warehouse.

4.2 Project Motivation

The initial task of this dissertation research was to thoroughly investigate how crash data is collected, distributed and utilized. When a crash occurs, the law enforcement officer who responds to the crash is responsible for completing a crash report. How that crash report reaches the state database varies depending on the municipality and how that crash report is analyzed depends on the user. The objective of this initial project was to fully understand the crash data environment so that future projects could directly target known deficiencies with the data.

4.3 Methods

The main objective of this initial research step was to become familiar with crash data in Massachusetts and discover any deficiencies which might exist before further work was conducted using the data. The NSF Innovation Corps (I-Corps) program was utilized to gain familiarity with crash data. The NSF I-Corps program seeks to "prepare scientists and engineers to extend their focus beyond the laboratory and broaden the impact of select, NSF-funded, basic-research projects." *(45)* This program involved conducting over 100 interviews with key crash data stakeholders and iteratively developing a hypothetical business model.

The business model focused upon the commercialization potential of the MassSafe Data Warehouse, **Figure 1**. The UMassSafe Traffic Safety Data Warehouse has been developed as a tool for maximizing the use of highway safety data. The data warehouse includes "administrative" datasets collected by state agencies and other organizations; including crash, citation, roadway inventory, etc. Currently, 14 such datasets are housed in the UMassSafe Traffic Safety Data Warehouse, with over 15 years of data available. Crash, citation, hospital, death certificate, and roadway inventory data have been linked using advanced statistical methodologies to create a single dataset that allows analysts to consider the comprehensive crash experience; including driver behavior, crash characteristics, roadway environment, and crash outcomes such as injuries and costs. Researchers in the UMassSafe research group have successfully used the Data Warehouse for many years on projects for the Massachusetts State Police, Executive Office of Public Safety and the Massachusetts DOT, among others.

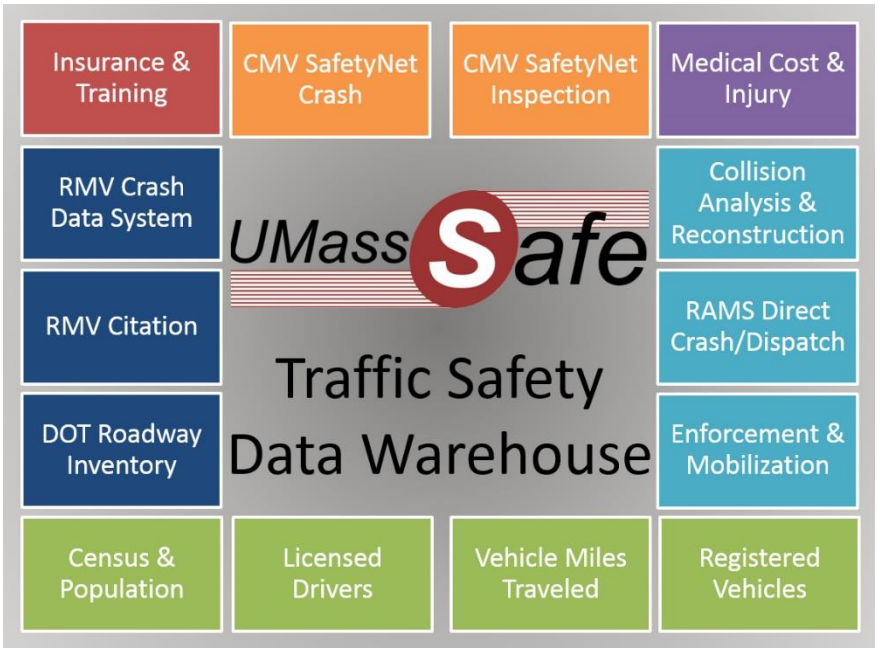


Figure 1. UMass Safety Data Warehouse.

The I-Corps program involved two intensive workshops that focused on the development of our business model. The two workshops, which were held at University of Southern California in Los Angeles, were separated by eight weeks during which we conducted 100 stakeholder interviews and presented weekly webinar updates on the progress of our hypothetical company "Safety Data Express". The two workshops, and the online weekly updates, were attended by: John Collura, Principal Investigator (PI); Michael Knodler, Co-PI; Paul Shuldiner, Business Mentor; and Cole Fitzpatrick, Entrepreneurial Lead.

4.4 Findings

Over 100 interviews were conducted with transportation engineers, researchers, insurance agents, police officers, personal injury attorneys, and transportation expert witnesses. This section outlines the key findings from the interviews with these crash data stakeholders.

Transportation engineers are one of the most frequent users of crash data. Within the Commonwealth of Massachusetts, any safety improvement project requires crash data analysis. This analysis consists of gathering all the reported crashes on that intersection or section of roadway. In the case of an intersection, the crashes are then compiled into an intersection collision diagram which visually depicts the common crash types and locations within an intersection, **Figure 2**. Finally, a road safety audit is conducted by a team of engineers and other stakeholders to identify other safety concerns that may not have been revealed by the intersection collision diagram.

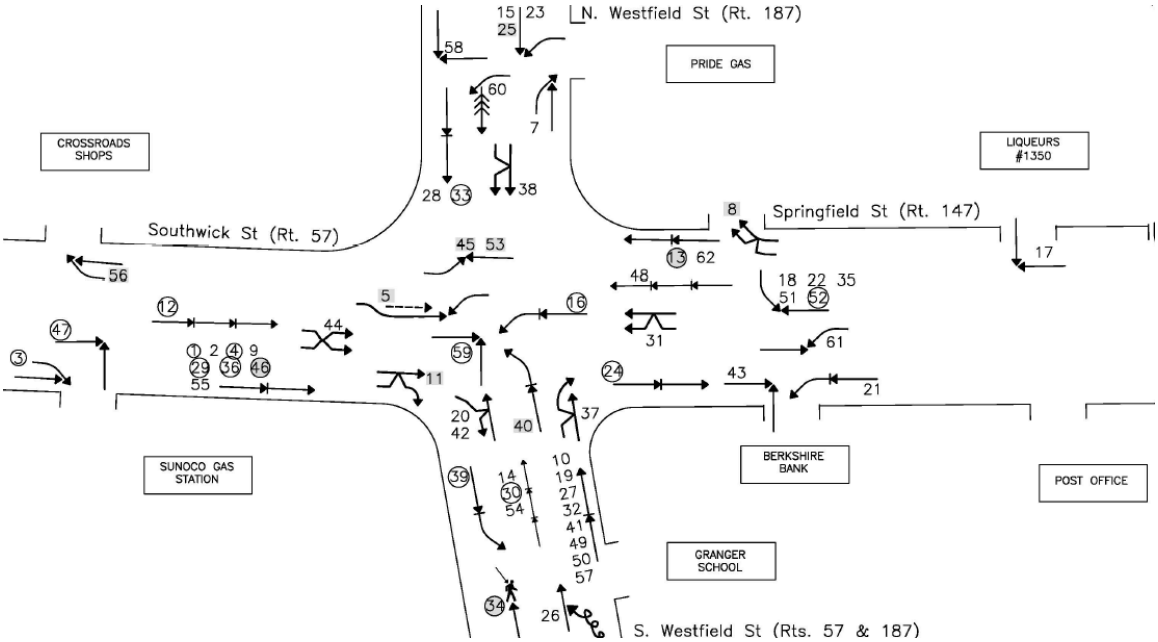


Figure 2. Example intersection collision diagram provided during an interview with a transportation engineer.

The common complaint made by transportation engineers was with respect to the first step of the process, gathering all the reported crashes. Engineers began the crash report gathering process by obtaining reports from the Massachusetts Department of Transportation. However, these data were only in the form of a summary rather than individual crash reports. Next, engineers would contact the local police station to request

full crash reports. This request often took multiple months to fulfill. After conversations with records clerks with police departments, we learned that this delay is because most departments do not log reports by location but rather just by the year. Additionally, some smaller towns only have one records clerk, making larger requests even more challenging. Many of the police chiefs we interviewed expressed a desire for a system that would ease the burden on their records clerks.

Data quality and timeliness of the data were also cited as concerns by engineers. From experience, many engineers expressed a distrust of certain fields within the crash report or a general mistrust of data collected in a specific municipality. Timeliness of the data caused frustration as often projects are initiated after a fatal crash, or string of fatal crashes. However, it often takes up to two years for a crash to become part of the accessible database. From conversations with the Registry of Motor Vehicles (RMV), who manages the crash data repository for Massachusetts, and police departments we learned the causes of this delay. When a police officer completes a crash report, it is then stored at their local police department. Then, periodically, police departments send their crash data to the RMV. This frequency varies between police departments and can be as frequent as weekly or as in-frequently as once every three months. Next, the RMV has to add this data to their database. This process is not challenging when police departments digitally send their records, as many do. However, many departments still send the RMV paper copies of their crash records, although most said that they are working on transitioning to digital reporting.

The primary concern of insurance companies was related to the aforementioned time in obtaining a crash report and the cost for acquiring an individual report, for most police departments this cost $10 to insurance companies. Unlike transportation engineers,

personal injury attorneys and transportation expert witnesses were unable to obtain historical data. This was because the provider of the data, MassDOT or police departments, were often the possible defendant in a lawsuit from the attorney and thus did not have the same motivation to share data as they would with engineers.

4.5 Conclusion

A specific objective during the program was to make a "go/no-go" decision on whether or not to incorporate our hypothetical start-up company after engaging in this intensive process comprised of many customer interviews. A benefit that resulted from this specific objective, and the entire I-Corps program, was a thorough understanding of the strengths and deficiencies regarding how crash data is collected, accessed, and utilized in the Commonwealth of Massachusetts.

As described previously, the UMass Safety Data Warehouse, is a collection of 14 datasets from a multitude of data suppliers. The diverse data sets allows researchers and practitioners to investigate traffic safety questions that are otherwise unanswerable. This Data Warehouse was the initial focus of our business model, and by the end of the program we had determined that a commercial market for the resale of crash reports and crash data was limited. However, there was a potential market for a service that provided on-demand intersection collision diagrams. The progression of the business model throughout the course is depicted in **Figure 3**.

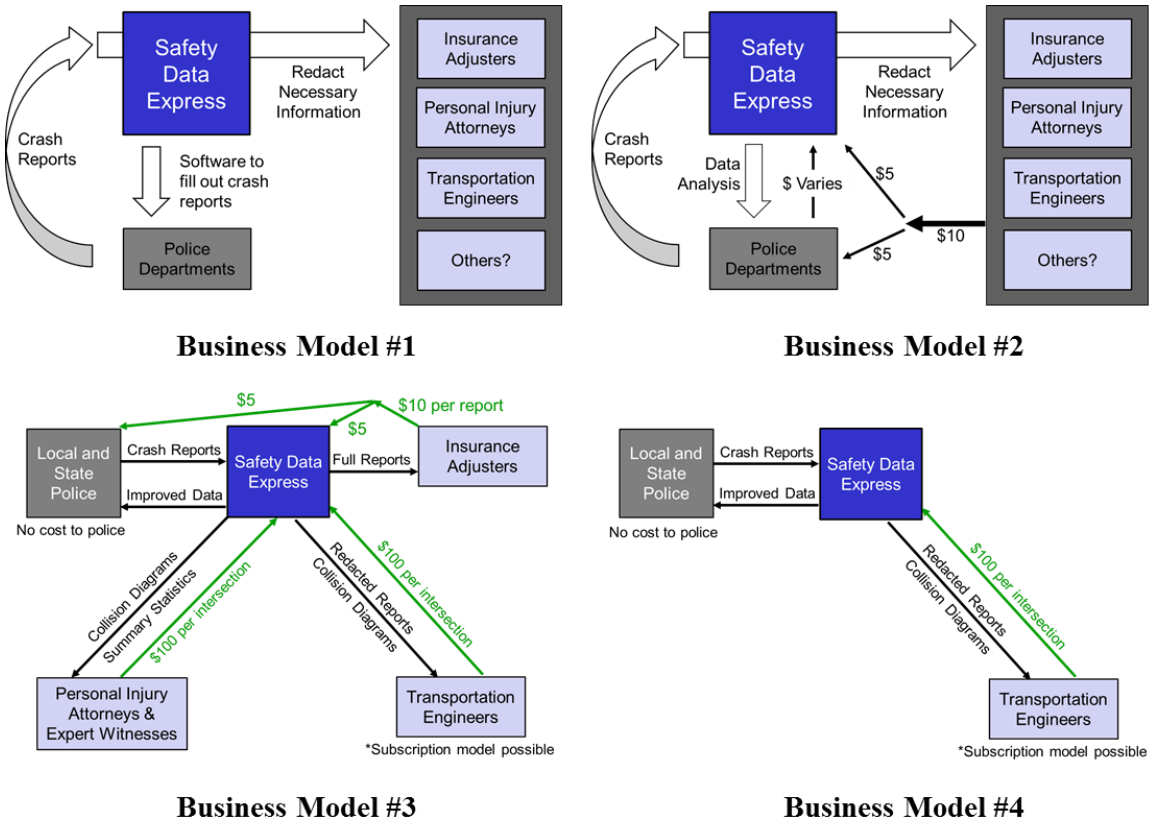


Figure 3. Business models developed during NSF I-Corps program.

During the project, our team wanted to investigate if there was a demand for our product outside of these large public agencies. Through our initial conversations with transportation engineers we found that there was limited demand for our additional datasets as engineers only needed one of them for their projects and they were able to obtain it, upon request, from the State DOT.

We shifted the investigation to insurance companies to examine how they use crash data and whether our product would be of interest to them. We discovered that insurance companies routinely request the police reports from the jurisdiction that responded to the crash. While this data exists within the UMass Safety Data Warehouse, the individual reports are anonymized and would thus be of limited use to insurance companies. Identifying the potential opportunity that lied with insurance companies and the frequent

need for crash reports, we spoke with police officers and police chiefs specifically to gauge their interest in a partnership where our company would provide the service of responding to these requests for reports in exchange for a portion of the fee charged. Most police chiefs were very receptive to the idea as their records clerks are frequently overburdened by requests for crash reports. However, we also discovered that a company, Appriss, provides this very service in the Northeast US through getcrashreports.com.

Identifying Appriss' lead and the fact that our idea to partner with police was not novel, we shifted back to transportation engineers with the thought that we could provide a service using our Safety Data Warehouse. Similar to how transportation consulting firms subcontract traffic data collection, we asked consultants if they would consider subcontracting safety analyses. Results were mixed as some engineers felt that conducting these analyses in-house led to a more complete understanding of the problem they were trying to fix. Others felt that subcontracting the often tedious task of safety analysis could save them money and free up time to focus on other aspects of the project.

While our team decided on a no-go decision as we still needed to better understand the size of the market opportunity, the program revealed some key flaws in the crash data environment that would need to be hashed out before a commercial effort could take place. Not only did I-Corps provide a crash course in business development, but it also resulted in a foundational understanding of how crash data is collected, distributed, and analyzed within the Commonwealth of Massachusetts. This underlying understanding was crucial during successive research projects, specifically Chapter 4, which investigated the speeding-related designation.

CHAPTER 5
AN INVESTIGATION OF THE SPEEDING-RELATED CRASH DESIGNATION THROUGH CRASH NARRATIVE REVIEWS SAMPLED VIA LOGISTIC REGRESSION

5.1 Summary

While many studies have utilized the speeding-related designation in safety analyses, no studies have examined the underlying accuracy of this designation. Herein, we investigate the speeding-related crash designation through the development of a series of logistic regression models that were derived from the established speeding-related crash typologies and validated using a blind review, by multiple researchers, of 604 crash narratives. The developed logistic regression model accurately identified crashes which were not originally designated as speeding-related but had crash narratives that suggested speeding as a causative factor. Only 53.4% of crashes designated as speeding-related contained narratives which described speeding as a causative factor. Further investigation of these crashes revealed that the driver contributing code (DCC) of "driving too fast for conditions" was being used in three separate situations. Additionally, this DCC was also incorrectly used when "exceeding the posted speed limit" would likely have been a more appropriate designation. Finally, it was determined that the responding officer only utilized one DCC in 82% of crashes not designated as speeding-related but contained a narrative indicating speed as a contributing causal factor. The use of logistic regression models based upon speeding-related crash typologies offers a promising method by which all possible speeding-related crashes could be identified.

5.2 Project Motivation

The primary objective of this study was to improve the identification of speeding-related crashes by investigating commonalities in the types of crashes that are routinely

misclassified as either speeding-related or not speeding-related. Logistic regression models based upon established speeding-related crash typologies were developed to predict the probability that a specific crash would be designated as speeding-related. The model outputs were then used to strategically sample crash narratives in order to identify potential crashes where the model prediction disagreed with officer's recorded crash causation (i.e. driver contributing code). The resulting evaluation of crash narratives was based upon two hypotheses that were tested:

Hypothesis 1: model predictions correlate with crash causation determinations resulting from crash narrative reviews.

Hypothesis 2: commonalities exist among the crashes with a misclassified speeding-related designation as determined through the crash narrative reviews.

The resulting output of the hypothesis testing would be an improved methodology to identify speeding-related crashes and any crash commonalities identified from misclassified crashes would be used to improve the classification of speeding-related crashes.

5.3 Methods

This study consisted of three primary phases. First, a series of logistic regression models were developed to assign a probability that a crash was, or was not, designated via the crash report as being speeding-related. Second, these models were utilized to sample crash reports for subsequent crash narrative reviews by multiple researchers that were unaware of the crash designation (i.e. a double blind narrative review). Finally, based upon the crash narrative review, specific crashes which had crash narratives that did not align with the officer's speeding-related designation were manually reviewed to identify shared characteristics. This section will describe the methods for the three phases of this study.

5.3.1 Logistic Regression Model

Three years of crash data from the state of Massachusetts from 2012-2014 were obtained. The roadway inventory database, maintained by the Massachusetts Department of Transportation (MassDOT) was utilized in order to link the crash to the roadway on which it occurred. Initially, 373,205 unique crashes were included in the database with an individual entry for each driver involved in the crash. Next, any crashes with an improperly coded driver age (e.g., driver age > 110) or driver sex (driver sex ≠ male or female) were removed from the database. For interstate crashes, entries were removed which had recorded speed limits which differed between the crash report and the roadway inventory. This was not conducted on other functional classifications as the speed limits reported on the crash report were inconsistent with those from the roadway inventory file. Instead, speed limit was not included in these models due to the low confidence in the data accuracy. Finally, only entries involving "Person Number: 1", also known as motor vehicle operator #1 (MV1), were included in the model development. This decision was made to conform to one of the fundamental assumptions of logistic regression models which states that all observations must be independent from one another. MV1 was selected for inclusion in the model as MV1 was more commonly at fault for exceeding the posted speed limit or driving too fast for conditions (DTFFC). Specifically, in 4.2% of all crashes MV1 was at fault due to speeding, compared to only 1.1% of crashes being the fault of MV2-5 for speeding. The crashes were then filtered by the functional classification of the roadway on which they occurred in order to create five logistic models. Multiple models were developed in order to improve the prediction capabilities of the model. The grouping of functional classifications and sample size for each model is presented in **Table 1**. Altogether, 161,419

crashes, both injury and property-damage crashes, were used to develop five different logistic regression models.

Table 1. Sample Sizes of the Five Functional Classification Models

Federal Function Classification Number	Functional Classification	No. of Crashes	No. of Speeding-Related Crashes
1,2	Interstate, Principal Arterial (Freeways and Expressways)	28667	2353
3	Principal Arterial (Other)	45966	1235
4	Minor Arterial	43458	1774
5, 6	Major Collector, Minor Collector	17670	1306
7	Local	25658	1851

The five logistic regression models were developed based upon the speeding-related crash typology from *(12)*. Two crash characteristics were expressed in different ways in order for the model to better fit the data. First, a crash occurring at night can be identified either by the time at which the crash occurred or the light conditions. Second, the crash type input was either single vehicle crash or first harmful event occurring outside of the roadway. The data field which resulted in a better model fit was selected. It was not possible to use both as the fields described are highly correlated. Including multiple correlated variables violates one of the main assumptions of logistic regression modeling which cautions against multicollinearity *(46)*. **Table 2** displays the coefficients for the variables included in each model, when a coefficient is not present, that variable was not included in the model. The constant and significant variables are used to calculate Y' which is the principal component of the logistic regression equation which calculates the probability (P) of the event occurring: $P(1) = e^{Y'}/(1 + e^{Y'})$.

Table 2. Variable Coefficients for the Logistic Regression Models to Calculate Y'

Variable	Federal Functional Classification Model				
	1/2	**3**	**4**	**5/6**	**7**
Driver Age (continuous)	-.026	-.032	-0.034	-0.038	-0.045
Speed Limit (continuous)	-.021	-	-	-	-
Driver Sex: (0: Female 1: Male)	0.347	.270	0.652	0.576	0.558
Road Surface: (0: Dry, 1: Not Dry)	2.354	1.691	1.558	1.411	1.224
Light Conditions: (0: Light, 1: Not Light)	-	-	0.209	0.344	-
Time of Crash: (0: 6am-10pm 1: 10pm to 6am)	-	0.450	-	-	0.303
Injury Severity (0: Not Fatal or Incapacitating 1: Fatal or Incapacitating)	0.664	0.523 (p = 0.018)	1.068	0.777	-
First Harmful Event: (0: Within Roadway 1: Outside Roadway)	-	-	1.668	-	1.365
# of Vehicles Involved: (0: More than one 1: One)	1.305	1.893	-	-	-
Crash Location (0: Not at Intersection 1: At Intersection)	-	-	-	-0.991	-
Constant	-2.409	-4.046	-3.79	-2.134	-2.634
Hosmer-Lemeshow Model P-value	0.340	0.714	0.122	0.307	0.168
Hosmer-Lemeshow Model Chi-square	9.03	5.40	12.71	9.43	11.64

Note: All variable p-values < 0.01 unless otherwise noted. Df = 8 for all five models.

The goodness of fit for each model was evaluated using the Hosmer-Lemeshow test. The test compares the number of observed events to the expected number of events in equally sized subgroups *(47)*. A p-value of 0.05 or less signifies that the hypothesis that the model fits the data can be rejected, thus p-values above 0.05 are acceptable with higher p-values implying a better model fit. Hosmer-Lemeshow test p-values for each of the five models are shown at the bottom of **Table 2**.

5.3.2 Crash Narrative Sampling

Crash narratives were sampled from six groups based on the logistic regression models. The six groups were based on the two officer-designations: Speeding-Related, Not Speeding-Related, and the three model outputs: high probability, medium probability and low probability of crash being speeding-related. High probability was defined as the 30 highest outputs from each functional classification model. Low probability was defined as the 30 lowest outputs. Medium probability was defined by calculating the median probability in that model's high probability group. For example, if the median probability of the 30 crashes in the high probability group was 0.60, the median probability in the medium probability group would be 0.30. Nine hundred crash reports were initially sampled. Of the 900 reports sampled, only 604, or about two-thirds contained a valid crash narrative. Of note, all crash reports sampled from the interstate and freeway model contained a valid narrative. While there was a slight overrepresentation of interstate/freeway crashes, this overrepresentation was constant across the six groups. **Table 3** displays the sample of crash narratives sampled from the six groups. **Figure 4** displays a graphical example for Minor Arterials: Speeding-Related Designation of how crash narratives were sampled.

Table 4 displays the various crash types that were captured within the sampling. Single vehicle crashes were captured more frequently than other crash types due to the variables included in the logistic regression models.

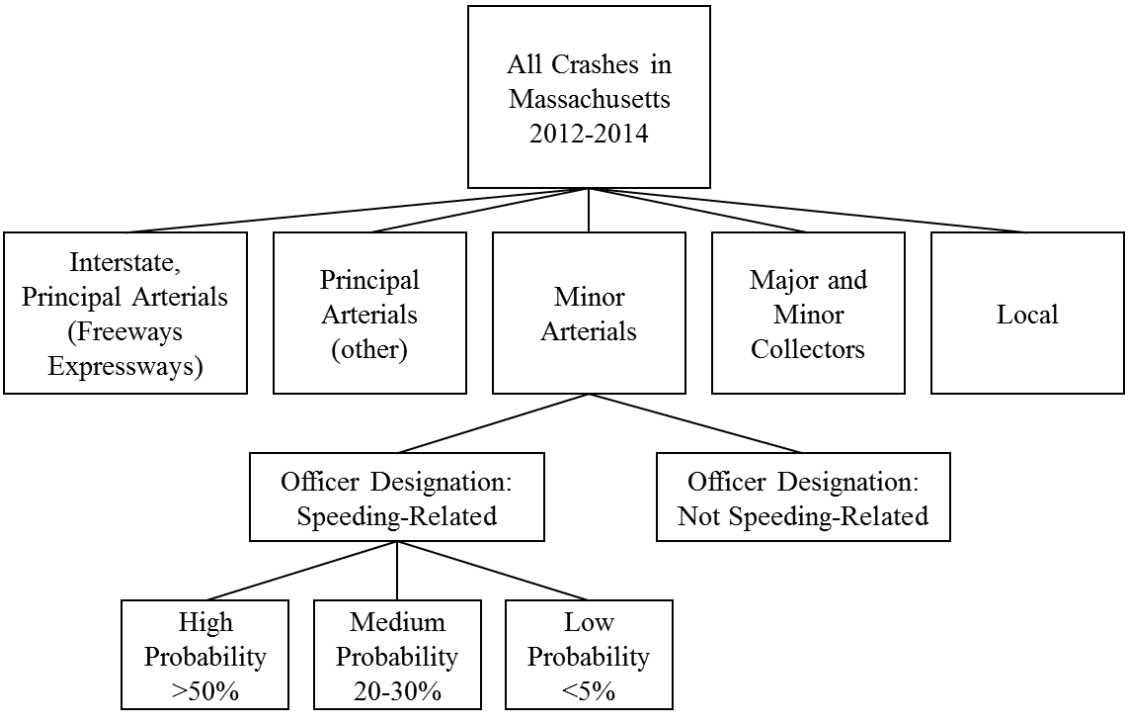


Figure 4. Data tree showing the sampling strategy from logistic regression models.

Table 3. Number of Sampled Crash Narratives from the Six Groups

Officer Designation	Model Output	Functional Classification Model					Total
		1 & 2	3	4	5 & 6	7	
Not Speeding-Related	High	30	19	12	20	12	**93**
	Med	30	18	17	20	18	**103**
	Low	30	16	20	17	22	**105**
Speeding-Related	High	30	20	19	17	15	**101**
	Med	30	24	12	19	17	**102**
	Low	30	17	21	13	19	**100**
Total		**180**	**114**	**101**	**106**	**103**	**604**

Note: Refer to **Table 1** for functional classification number definitions.

Table 4. Crash Types within the Sample Separated by Officer Speeding Designation

Manner of Collision	Not Speeding-Related	Speeding-Related	Total
Single Vehicle - Collision in roadway/median/roadside	135	166	301
Single Vehicle - Ran off road	36	42	78
Rear-end	35	46	81
Angle	54	28	82
Sideswipe	27	11	38
Head-on	7	6	13
Unknown	7	4	11

5.3.3 Double Blind Narrative Review

The sampled crash narratives were assigned to a team of six reviewers who were research assistants within the UMass Transportation Program. The reviewers were trained with example crash narratives where there was a clear and known answer. In addition, several narratives that were not included within the sample were reviewed by all reviewers on the review team to make sure that there was agreement. The narratives were distributed in such a manner so that each person reviewed an equal number of crashes from each of the six groups. Each crash was reviewed by two of the six reviewers, with each reviewer being blind to the group from which the crash belonged and blind to the identity of the other person reviewing that narrative. To eliminate the effect that reviewer bias or tendencies may have on the results, the narratives were assigned to reviewers in a manner that ensured that each reviewer reviewed an equal number of crash reports. In total, each reviewer read around 200 crash narratives. Each crash narrative review, which were compiled digitally into a spreadsheet, took one to three minutes to complete, depending on the length of the narrative.

Reviewers were also blind to the objectives and hypotheses of the study. They were instructed to decide whether the "Narrative indicates that the officer determined the crash

was at least partially caused by Motor Vehicle Operator 1 (MV1) speeding and/or driving too fast for conditions?"

The Cohen's kappa test was conducted in order to measure the agreement between reviewers. The test evaluates the level of agreement against the probability of the reviewers agreeing by chance *(48, 49)*. The test outputs a kappa value between 0 and 1 with 1 meaning perfect agreement and 0 meaning no agreement. The cases in which the reviewers did not agree were reviewed by a graduate researcher whose review counted as the tiebreaker. The results of the double-blind narrative review were then compared to six categories sampled from the logistic regression model.

5.4 Results and Discussion

A total of 604 crash narratives were reviewed by a team of six undergraduate students to answer the question of "does the crash narrative indicate that operator #1 was at fault due to exceeding the posted speed limit or traveling too fast for conditions?" The reviewers agreed on 542 of the 604 narratives (89.7%).

The kappa values between each of the six reviewers are shown in **Table 5**. The overall calculated kappa value of 0.77 suggests a good level of agreement not based on random chance. Thus, the results of these reviews were significant and can be utilized in further analyses. Prior to completion of further analyses, the 62 crashes in which the reviewers did not agree were reviewed by a graduate researcher whose review counted as the tiebreaker. Of the 62 disagreements, 26 were ultimately determined to have narratives indicating that the crash was speeding-related.

Table 5. Inter-Rater Agreement Test

Reviewer 1	Reviewer 2	Reviewer 1 Response/Reviewer 2 Response				κ
		Yes/Yes	Yes/No	No/Yes	No/No	
1	2	17	0	2	23	0.90
1	3	8	2	3	26	0.67
1	4	10	4	2	25	0.66
1	5	11	4	2	20	0.66
1	6	12	3	1	26	0.81
2	3	9	0	0	31	1.00
2	4	16	1	3	22	0.81
2	5	13	1	5	19	0.68
2	6	12	0	3	27	0.84
3	4	6	3	5	27	0.47
3	5	6	1	2	30	0.75
3	6	9	0	2	31	0.87
4	5	14	2	0	20	0.89
4	6	12	5	1	24	0.69
5	6	12	5	0	24	0.74
Total		**167**	**31**	**31**	**375**	**0.77**

Response to question: "Does narrative indicate the crash was speeding-related?"

5.4.1 Comparing review results to crash narrative length

It was earlier hypothesized that the logistic regression model could predict which crashes truly would or would not be speeding-related based on its crash narrative. As shown in **Table 6**, the model accurately identified crashes which were not originally designated as speeding-related but had narratives which indicated that speed was a causative factor. For crashes originally designated as speeding-related, the model was less accurate. This may have been due to the fact that only 164 of these 303 (54.1%) crashes designated by the officer as being speeding-related contained narratives which described speed as a reason for the crash. This low percentage may be partially explained by investigating the length of the crash narratives, **Table 7**. When examining these 303 crashes which had an officer designation of speeding-related, the 164 narratives which indicated speed as the crash causation had a mean narrative length of 174 words (St. Dev: 134). By contrast, the

139 crash narratives that did not indicate speed as a cause had a mean narrative length of 101 words (St. Dev: 101), a statistically significant difference (two sample t-test, p= <0.001). This relationship was also observed within the 301 crashes that were not officer-designated as speeding-related. The 29 narratives which contradicted the original designation to indicate speed as the causative factor had a mean length of 215 words (St. Dev: 171) as compared to a mean length of 131 words (St. Dev: 126) for the 272 narratives which did not indicate speeding-related, another statistically significant difference (p = 0.015).

Table 6. Crash Narrative Review Results Based on Sampling Category

Officer Designation	Model Prediction	# Speeding-Related Indicated by Narratives	Total Reviewed	% Narratives Indicating Speeding-Related
Not Speeding-Related	High	18	93	19.4
	Med	8	103	7.8
	Low	3	105	2.9
Speeding-Related	High	62	101	61.4
	Med	44	102	43.1
	Low	58	100	58.0

Table 7. Crash Narrative Review Results versus Narrative Length

Officer Designation	Narrative Indication	n	Mean Length (words)	Standard Deviation
Not Speeding-Related	Not Speeding-Related	272	131	126
	Speeding-Related	29	215	171
Speeding-Related	Not Speeding-Related	139	174	134
	Speeding-Related	164	116	101

Note: Both differences are statistically significant ($p < 0.05$)

5.4.2 Exceeding the Posted Speed Limit versus Driving Too Fast for Conditions

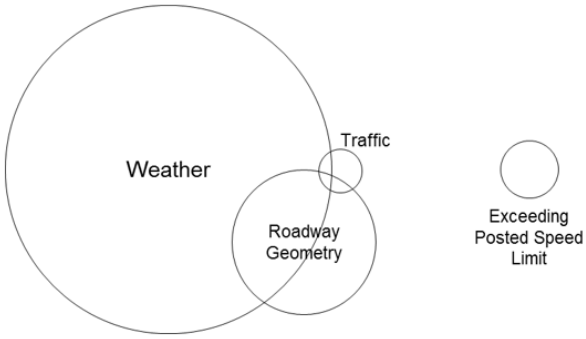
The low percentage (54.1%) of speeding-related crashes which contained narratives describing speed as a causative factor warranted additional investigation into the two Driver Contributing Codes (DCCs) which classify a crash as speeding-related. As shown

in **Table 8**, crashes with a DCC of "exceeding the posted speed limit" contain a narrative which mentions speed nearly 75% of time. However, crashes with "driving too fast for conditions" (DTFFC) as the DCC have narratives mentioning speed only 47.1% of the time, a statistically significant difference ($p = <0.001$, t-test). Interestingly, a small sample of crashes contained both driver contributing codes but only one contained a narrative which indicated speed as the crash causation.

Table 8. DCC-Based Narrative Results for Crashes Designated as Speeding-Related

Driver Contributing Code	Narrative Indicated Speeding-Related	Total Reviewed	Percent
Exceeding the Posted Speed Limit	64	87	73.5%
Driving Too Fast for Conditions	99	210	47.1%
Both DCCs	1	6	16.7%

Based upon this observation, the narratives of crashes with DTFFC as the DCC were investigated further. It was found that DTFFC was being utilized by officers in four situations: (1) too fast for weather conditions, (2) too fast for the roadway geometry (e.g. down a hill, around a curve), (3) too fast for congested traffic conditions, and (4) exceeding the posted speed limit (i.e. officer should have used "exceeding the posted speed limit" DCC). The narratives of crashes with DTFFC were further classified into these four situations, **Figure 5**.



Weather	49
Weather + Traffic	1
Weather + Roadway Geometry	17
Weather + Traffic + R. Geometry	1
Traffic	7
Traffic + Roadway Geometry	1
Roadway Geometry	12
Exceeding Posted Speed Limit	12

Figure 5. Conditions referred to in "driving too fast for conditions" narratives.

The majority of crashes coded as DTFFC were related to crashes which occurred in inclement weather. The second most common use for DTFFC was related to roadway geometry, which unsurprisingly overlapped often with weather conditions. Finally, the least common use for DTFFC was in congested traffic conditions. Often times in these crashes, drivers were cited for "Failure to Use Care While Stopping". Twelve out of the 100 crashes simply involved speeding and not DTFFC and should have been instead categorized as "exceeding the posted speed limit".

These findings suggest that, while weather is the most common use for DTFFC, the DCC is also being used in situations referring to roadway geometry or congested traffic. If engineers and researchers were not aware of this fact, they may simply assume that DTFFC implies driving too fast for weather conditions. Perhaps separating DTFFC into three separate DCCs, relating to weather, roadway geometry, and traffic, would help officers select the most appropriate DCC and would improve safety analyses performed by engineers and researchers.

5.4.3 Identifying speeding-related crashes originally misclassified

The sampling method using the logistic regression model identified 29 crashes with narratives indicating speeding as a factor which were not originally designated as speeding-related. Of these crashes, 82% (24 of 29) contained a DCC of "Failure to keep in proper lane or running off road" and "Operating vehicle in erratic, reckless, careless, negligent or aggressive manner". These DCCs could be incorporated into the speeding-related crash typology and utilized when performing future analyses such as the one conducted in this paper.

Interestingly, 24 out of 29 (82%) of the misclassified crashes only contained one DCC when officers have the ability to enter two. While the use of only one DCC is a potential reason these crashes were misidentified, crashes originally identified as speeding-related contained only one DCC in 225 out of the 303 reviewed (74%), a statistically insignificant difference ($p = 0.25$). Officers should be encouraged to use more than one DCC when completing a crash report and should be educated as to how this additional information is useful to engineers and safety practitioners. If officers were to indicate a second DCC more often, it is less likely that speeding-related crashes, in addition to all other crashes, would be incorrectly classified.

5.4.4 Recommendations

Based upon the investigation of crashes involving the DCC of "driving too fast for conditions" (DTFFC), it is recommended that this DCC be separated into three DCCs: "driving too fast for weather conditions", "driving too fast for traffic conditions", and "driving too fast for roadway geometry". While these specific details can be obtained from the crash narratives, such a change would benefit engineers and researchers when conducting safety analyses. Crash and speeding mitigation strategies greatly depend on the

types of crashes occurring on a specific roadway, and more specific DCCs would increase accuracy. The “driving too fast for conditions” DCC is currently an all-encompassing option of officers, and officers would be better equipped with more distinct and intuitive DCCs. In the past, paper crash reports necessitated concise crash report forms. However, the recent digitization of crash reporting allows for more crash report fields. This change should be considered when states are updating their crash report forms.

When conducting analyses using existing speeding-related crash data, engineers should attempt to obtain full crash narratives whenever possible. The crash narrative, in combination with the standard crash information, can provide additional insight into why speeds are of concern at a particular location.

5.4.5 Future Work and Limitations

The crash narrative review was conducted manually as the speeding-related designation is very subjective and it was important to accurately review data. In the future, an automated crash narrative review process as demonstrated by *(50)* could be developed for speeding-related crashes. An automated narrative review process could be used in conjunction with the standard crash information to most accurately identify crashes which were related to speeding. Additionally, the length of the crash narrative may be used in future models to gauge the confidence in the model’s prediction. For example, a two paragraph crash narrative is likely to encompass all of the crash details, including the causation, whereas a one or two sentence narrative is not likely to provide sufficient detail.

The crash narratives which were reviewed only encompassed crashes occurring in Massachusetts. A more robust study could sample crash narratives from multiple states to see if the findings match the conclusions from this research. Another interesting analysis would be to stratify the model results by agency type to determine the extent to which

reporting practices may be consistent within an agency. Crash reports filled out on highways and interstates always had a crash narrative, whereas only two-thirds of all crashes originally sampled contained a valid narrative. This may be due to the fact that these crashes fall within State Police jurisdiction, which may be indicative of consistent reporting practices and increased training within a single agency. Future studies could investigate the accuracy of the speeding-related crash designation as it relates to the responding officer's jurisdiction.

5.5 Conclusions

In order to investigate the speeding-related crash designation, logistic regression models were developed based upon established speeding-related crash typologies. These models were used to sample 604 crash reports for a double-blind crash narrative review conducted by a team of six reviewers to determine if the officer deemed MV1 at fault for speeding and/or driving too fast for conditions. The resulting reviews were in agreement 89.7% of the time and disputed narratives were further analyzed by a member of the research team.

Hypothesis 1, related to the level of correlation between the developed logistic regression models and the crash causations. This hypothesis is partially accepted. The logistic regression model accurately identified crashes which were not originally designated as speeding-related but had crash narratives that suggested speeding as a causative factor. However, little agreement was seen between the model and crashes originally designated as speeding-related, which may have been due to the fact that only 53.4% of these narratives described speeding as a causative factor.

Hypothesis 2, was related to the ability to identify commonalities between crashes that are misclassified with respect to their speed related crash causation level. This

hypothesis was accepted. Further investigation of misclassified crashes revealed that the DCC of “driving too fast for conditions” was being used in three separate situations. Additionally, this DCC was also incorrectly used when “exceeding the posted speed limit” would have been more appropriate. Finally, it was determined that the responding officer only utilized one DCC in 82% of crashes not designated as speeding-related but contained a narrative indicating speed as a factor.

In summary, the use of logistic regression models based upon speeding-related crash typologies offers a promising method by which all possible speeding-related crashes could be identified. The review of crash narratives associated with speeding-related crashes revealed three distinct ways in which the DCC of “driving too fast for conditions” was being used. Whenever feasible, crash narratives should be reviewed when selecting safety countermeasures at a high crash location.

CHAPTER 6
THE APPLICATION OF CONTINUOUS SPEED DATA FOR SETTING RATIONAL SPEED LIMITS AND IMPROVING ROADWAY SAFETY

6.1 Summary

Research on rational speed limits suggests that simply lowering speed limits does not necessarily result in safer roadways; thus, there is a need to revisit the process by which speed limits, which are the front lines of any speed management program, are established. Traditionally, speed studies are conducted by taking spot speed observations at varying intervals along a roadway, however it would be ideal to have speed values continuously along a roadway. The specific objective of this research effort was to compare a continuous data collection method with existing methods and develop a methodology for integrating them to improve roadway safety. In this study, a group of drivers were equipped with a smartphone application which continuously captured video, vehicle speeds, and location data. The continuous speeds were then compared to speeds captured at eight fixed points. The results identified similarities in the 85th percentile speeds observed using the various data collection methods and a case study was conducted using FHWA's expert system, USLimits2. The results provide evidence for a successful proof of concept for mapping continuous speed data to traditional speed data collection points that may help in the speed limit setting process as well as the establishment of appropriate advisory speed zones. This research endeavor outlined a methodology which may be utilized to improve the process by which engineers determine speed limits and advisory speed zones.

6.2 Project Motivation

Traffic engineers typically employ conventional processes for the task of setting speed limits using operating speed data collected at fixed points. However, these data may

be misleading as they do not capture the entire speed profile, and the selection of the data collection location may ultimately bias the resulting speed data and established speed limit. As an example, the Massachusetts Department of Transportation (MassDOT) acknowledges in their 'Procedures for Speed Zoning on State and Municipal Roadways' that "it would be ideal to have speed checks continuously along the roadway so that the 85th percentile speed could be computed at all points." However, as of the last edition of the guidelines in 2012, MassDOT concedes that this type of data collection would not be practical. Given recent advances in smartphone technology, it is prudent to revisit the data collection process for setting speed limits. The advent and proliferation of mobile phone devices with GPS capabilities allows data aggregators such as Inrix ®, Google ®, and TomTom ® to report real-time traffic conditions. These crowd sourced data sets use speeds that are calculated nearly instantaneously and continuously updated. These anonymous data sets could likely be used to sample the traveling public and utilized as a basis for speed limit determination.

Traditionally, speed limits on new construction are based on the design speed of the roadway segment. Many speed limits remain vestiges of the highway building boom era of the 20th century and remain inappropriate for the current conditions. Present-day speed limit modifications are prompted by several means: town or city officials may have received complaints, the roadway may be under a rehabilitation, or crash history may warrant a speed limit change.

Crowd-sourced data would provide agencies with an active approach to speed management. Instead of waiting for crashes, road redesign or complaints, agencies could utilize these robust data sets to improve road safety. In addition, police departments could

use the data to determine uniform and consistent speed enforcement thresholds, and town engineers could use the data to identify differences between speed limits and active free-flow speeds.

The specific objective of this research is to explore the feasibility of linking and integrating continuous speed data collection with traditional speed limit setting practices. In this study a methodology was developed which utilizes Ubipix, a traffic and video data collection smartphone application (app), in order to generate continuous speed profiles for use in setting speed limits and determining speed advisory zones.

6.3 Methods

Speed data was captured on a 1.75 mile stretch of rural road in Amherst, Massachusetts, **Figure 6**. South East Street was selected for its varying speed limit and popularity among commuters. Additionally, the roadway has a frequent history of speeding violations and is under review by the Town of Amherst to explore possible speed management strategies. Despite the high prevalence of speeding, this location has an average crash frequency, meaning operating speeds are more influential in the speed limit setting process. Speed data was captured using three different methods: equipped volunteer drivers with the Ubipix app, eight installations of pneumatic tubes with automated traffic recorders (ATRs), and LiDAR spot speed collections at the same eight locations of the ATRs.

Figure 6. Study location, South East St in Amherst, Massachusetts.

6.3.1 Pneumatic Tubes with Automated Traffic Recorders

Eight ATRs were installed along the road at 1,000 to 1,500 foot intervals. To verify the results of the smartphone app, Ubipix, the ATRs were installed during the same time period that the trial drives occurred.

6.3.2 LiDAR Spot Speed Collection

One hundred free-flow vehicle speeds were captured from each direction of travel during daylight hours. Data were collected at the same eight locations along the route where ATRs were deployed. Ideally, LiDAR data collection would have been conducted concurrently with the probe drives and ATRs. However, in order to not influence the trial

drivers' behavior, LiDAR data collection occurred after these drives occurred when the ATRs were not installed. LiDAR data were collected during the same time of day, under similar weather conditions, and during the same time of year as the probe drives and ATRs.

6.3.3 Smartphone App

Twenty sample drives were collected by five subject drivers. Drivers were asked to download and install Ubipix on their smartphone, and were then provided with a mount so that their phone would be positioned to capture video as they drove. Each of the five drivers drove the 1.75-mile route twice in each direction. The volunteers consisted of two males in their mid-20's, two females in their mid-20's and a 60+ year old female. Admittedly, the sample size and range of ages was limited, however, this research was intended to be a proof of concept. In the future, a more diverse sample would be desired, and this would be possible when using data obtained from one of the large traffic data providers. The drivers were asked to simply drive as they normally do and they were informed that the app would not be capturing audio.

6.3.4 Smartphone Data Output

After the four trial drives for each driver, the app data was uploaded. The standard Ubipix web interface and data platform is presented in **Figure 7**.

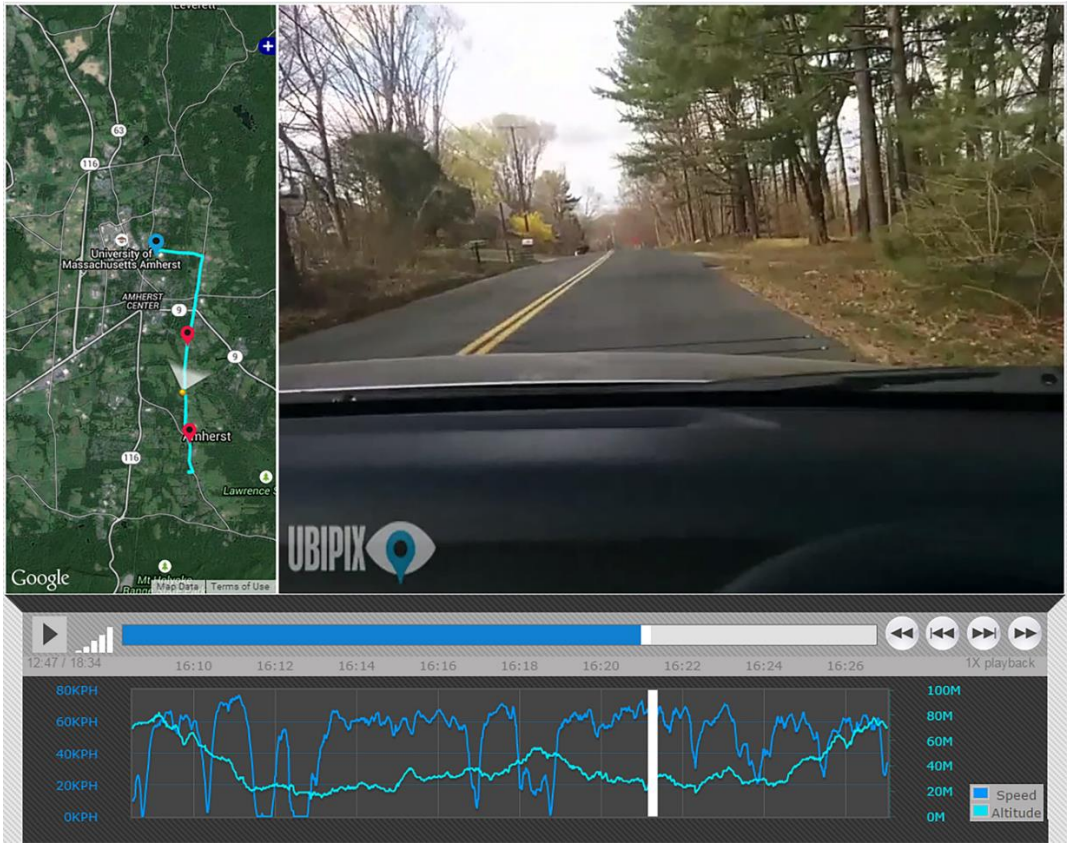


Figure 7. Standard output from the Ubipix app.

The smartphone app's graphical interface consists of three primary sections, the map view, camera view and speed/altitude graph. On the map to the left, the blue pin marker represents the starting point of the drive and the red pin markers are tags that show the 1.75 mile segment of road being studied. The drive began approximately 2.5 miles in advance of the test segment to engage drivers in the regular driving task prior to the experimental segment. It was our hope that this 2.5 mile warmup period would be sufficiently long for the drivers to forget that their drive was being recorded and their behavior would not be altered for the trial segment in any meaningful way. The yellow dot represents the position of the vehicle at the corresponding video point with the sight triangle indicating the direction of travel. In the example shown in **Figure 7**, one of the ATR setups is visible just above the dashboard.

6.3.5 Data Manipulation

As seen in the graphical user interface, the speeds are only presented graphically. It is challenging to do more than a visual inspection of the data based upon this graph. However, the raw data of the drive can be extracted from the app's platform. The data is exported in an unformatted '.json' file which, when formatted, can then be converted into a spreadsheet file. The raw data recorded at a 1 hz frequency included vector data of latitude, longitude, bearing, and speed. From the given coordinates the distance between the last data point was calculated using the formula for distance between Latitude and Longitude points on a WGS-84 coordinate system as shown in Equation 1.

Equation 1: Vincenty's Ellipsoidal Formula *(51)*.

$$Distance\ between\ two\ points = \cos^{-1}(\sin(Latitude_1)) * \sin(Latitude_2) + \cos(Latitude_1) * \cos(Latitude_2) * \cos(Longitude_2 - Longitude_1) * \alpha$$
$$where\ \alpha = 3958.756\ Mile\ (The\ Ellipsoid\ Radius\ of\ the\ earth\ in\ WGS-84\ Coordinates)$$

Next, by calculating the cumulative distance traveled, a speed versus position relationship was developed similar to the graph seen in the graphical interface.

6.3.6 Overlaying ATR Locations with Ubipix Data

Using the app's built in geotagging system, video from one of the drives was utilized to tag the coordinates of the eight ATRs to determine their exact location. Using these coordinates in conjunction with the coordinates from the Ubipix drives, the location of the ATRs were geotagged on the speed versus location graph.

6.4 Results

Using the continuously collected speeds from the smartphone data, a comparison was made between the other two data collection methods. These speed data points were

calculated every 50 feet for the Ubipix drives and at the eight fixed locations for the ATRs and the LiDAR spot speed observations.

6.4.1 Verification of Smartphone Data Accuracy

ATRs and LiDAR guns have been shown to be accurate to 0.5 m/s (~1 mph) *(19)*, and GPS devices which capture ground speed, have been shown to have similar accuracy *(52)*. However, it is prudent to verify Ubipix specifically. Since the Ubipix drives occurred when the ATRs were active, the ranges of speeds observed in the drives can be compared to the ranges of speeds collected by the pneumatic tubes. While ATRs are unable to automatically link an observation to a specific driver, visual inspection of the graphs in **Figure 8** and **Figure 9** show that the ranges of speeds observed in the ten drives match (i.e. in each direction of travel) closely to the ranges of speeds measured by the ATRs.

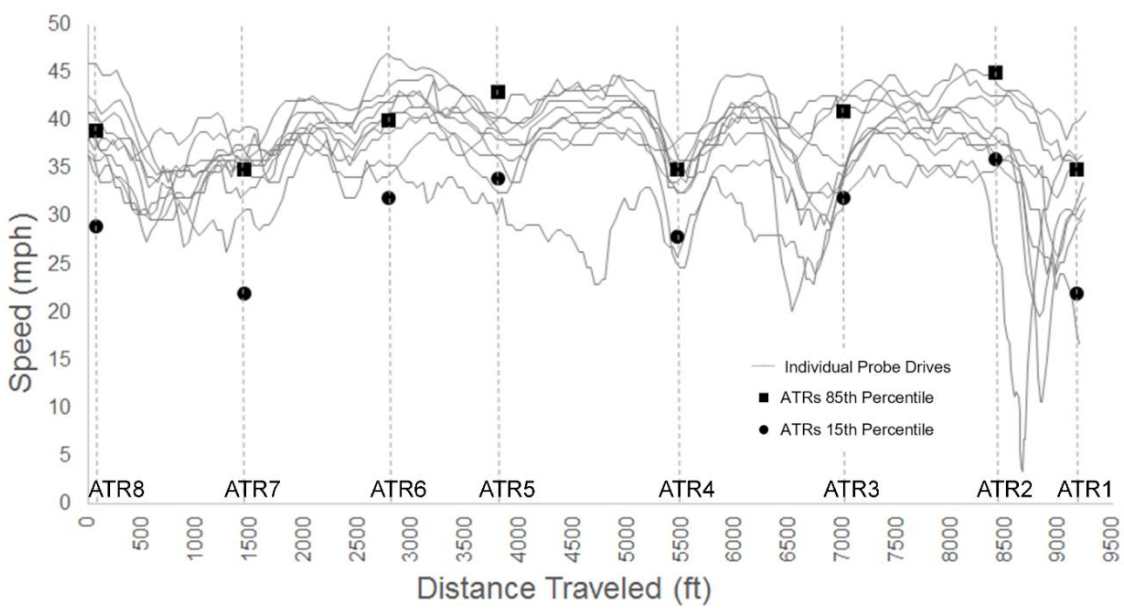


Figure 8. Northbound Ubipix drives versus eight ATR locations.

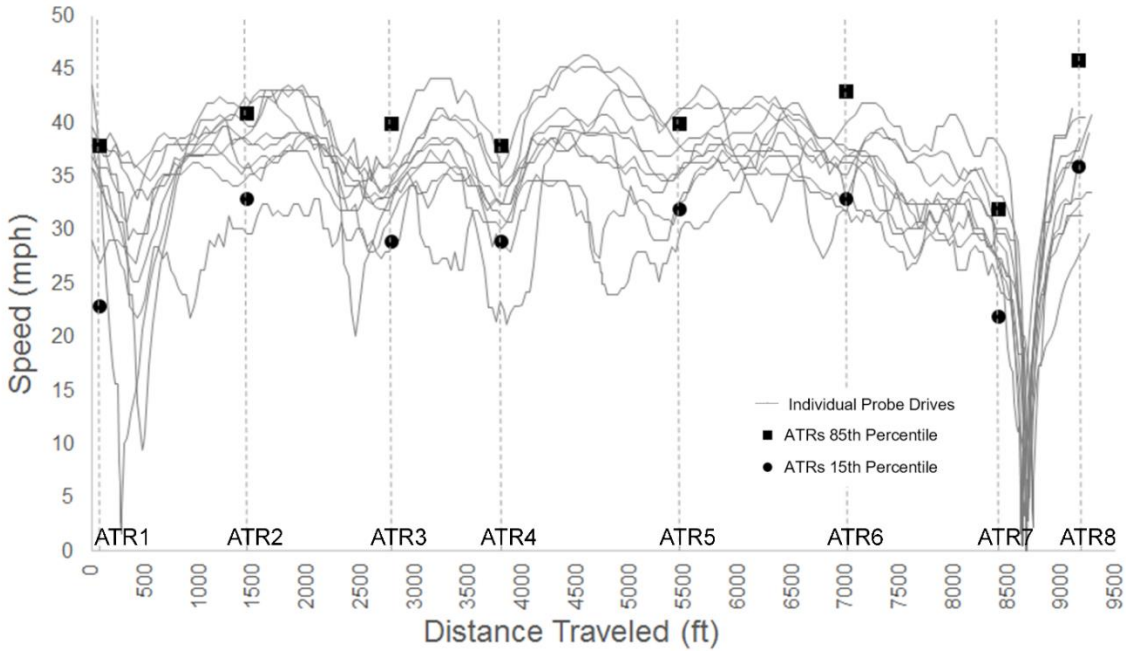


Figure 9. Southbound Ubipix drives versus eight ATR locations.

6.4.2 Comparison of Data Collection Techniques

Spot speed data was collected via a LiDAR gun. Unlike the ATR data, all spot speed measurements consisted of vehicles in free-flow conditions. As such, most 15th and 85th percentile speeds collected via LiDAR were higher than the speeds collected by the ATRs. These comparisons are demonstrated for the northbound and southbound drives in **Figure 10** and **Figure 11**, respectively. The trial drives with the smartphone app provide speed data between the eight fixed points. The size of each bar represents the difference between the 15th percentile and 85th percentile speed. This allows for an inspection of the speed variability along the 1.75 mile route. For both the northbound and southbound drives, the trial drive variability falls within the range of the ATR and LiDAR data at the eight fixed points. This suggests that while ten drives in each direction may not be enough to generate a representative sample, a continuous data collection method may necessitate smaller sample sizes than other spot speed methods.

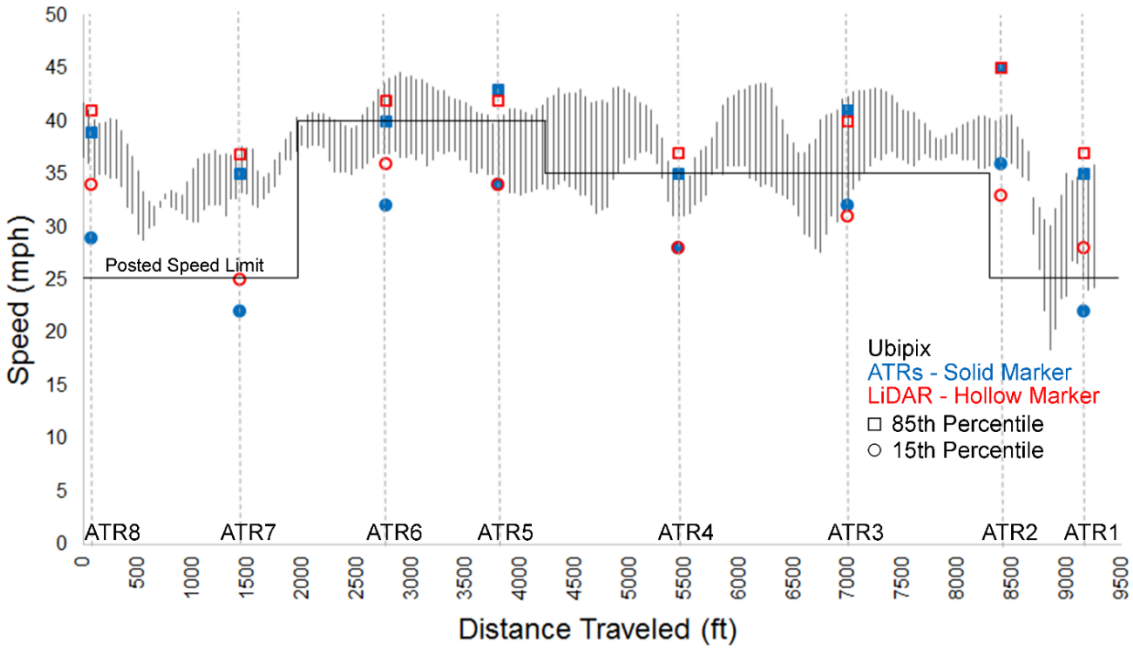


Figure 10. Speeds and variance across data collection methods, NB direction.

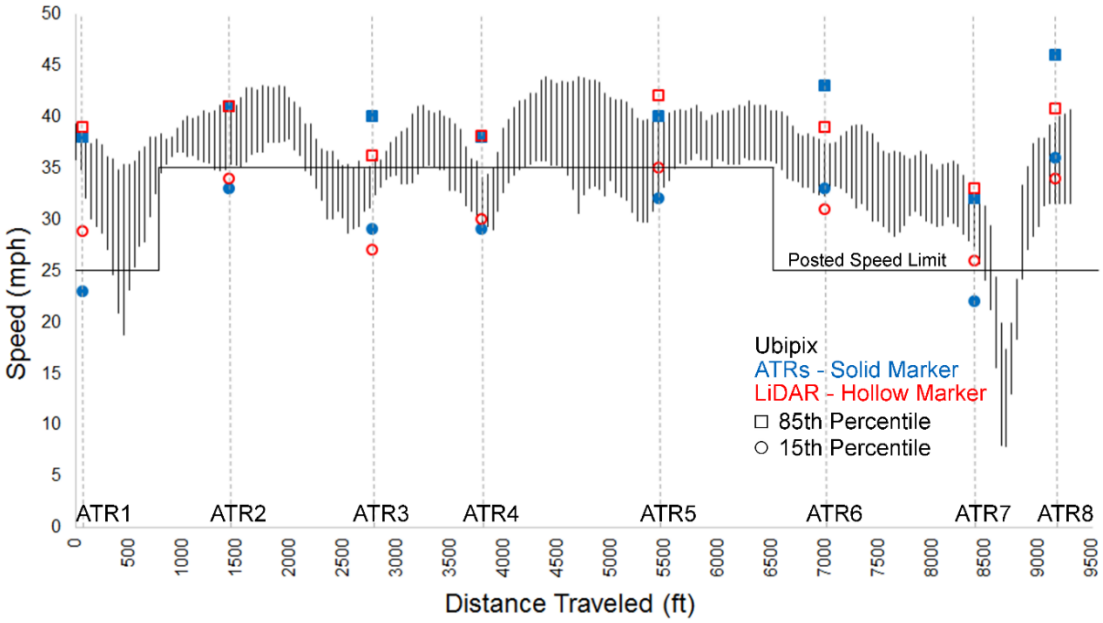


Figure 11. Speeds and variance across data collection methods, SB direction.

6.4.3 USLimits2 Case Study

In order to compare the three speed data collection techniques employed in this study, the 35 mph posted speed limit segment in the northbound direction was evaluated using USLimits2. As explained previously, USLimits2 is an expert system developed by FHWA to assist practitioners in determining a safe and sensible speed limit on a given roadway segment. The system considers operating speeds, roadway geometry, surrounding land use, and crash history in order to recommend the most appropriate speed limit. The segment, which can be seen on **Figure 10** from approximately 4300 feet to approximately 8400 feet, was selected due to its relatively high variance in speed. The variable inputs for the 0.8 mile segment are shown in **Table 9**. Crash history was obtained via MassDOT's Crash Portal. For the segment in question, nine years of crash history were available with 17 total crashes including four non-fatal injury crashes. The crash rates and injury rates were compared against the rates included in USLimits2 which were obtained from the Highway Safety Information System database *(53)*.

Table 9. Variable Inputs for USLimits2 Case Study

Category	Selected Option
Route Type	Road Section in Developed Area
Section Length (miles)	0.8
Average Annual Daily Traffic	2000
Adverse Alignment	No
Statutory Speed Limit for this Type of Road	40 mph
Is this a one-way street?	No
Divided/Undivided	Undivided
Number of Through Lanes	2
Area Type	Residential-Collector
Total number of driveways and unsignalized access points in the section	28
Total number of signals in the section	0
On Street Parking and Usage	Not High
Pedestrian/Bicyclist Activity	Not High

The inputs for 50th and 85th percentile speed were varied depending on the three data collection methods and the location at which data were collected. Determining which speeds to input relies heavily upon "engineering judgment". The USLimits2 User Guide does provide the following guidance, "The 85th percentile speed used in the analysis for a general speed limit should not be taken from data collected in the adversely aligned section." *(37)* As shown in **Table 10**, eight data points were selected for input into the system, six points from the two fixed locations where ATR and LiDAR data were collected, and two points between those fixed points. The results show that the recommended speed limit varies at the two locations with all three methods of speed data collection. This demonstrates how a practitioner may utilize continuous speed data. If a change to the speed

limit was desired, the continuous data could determine the exact location where that change should occur.

Table 10. USLimits2 Speed Limit Recommendations for Each Method

Collection Point (feet)	Method of Data Collection	50th Percentile Speed (mph)	85th Percentile Speed (mph)	USLimits2 Recommended Speed Limit (mph)
5500*	ATR	32	35	35
	Ubipix	33	35	35
	LiDAR	33	37	35
6600	Ubipix	34	37	35
7000	LiDAR	35	40	40
	ATR	36	41	40
	Ubipix	37	42	40
4700	Ubipix	38	43	40

Note: The collection point refers to the distance traveled scale on **Error! Reference source not found.**. The p osted speed limit for this segment is currently 35 mph.
(*) indicates that data was collected at "adversely aligned section"

Currently, practitioners must decide which location and which method of data collection to use. Smartphone applications have the ability to provide continuous data, which when presented graphically, can provide extra context when engineering judgment is needed. For example, speed peaks can be identified and explained by viewing the video, as demonstrated on the southbound drive shown in **Figure 12**.

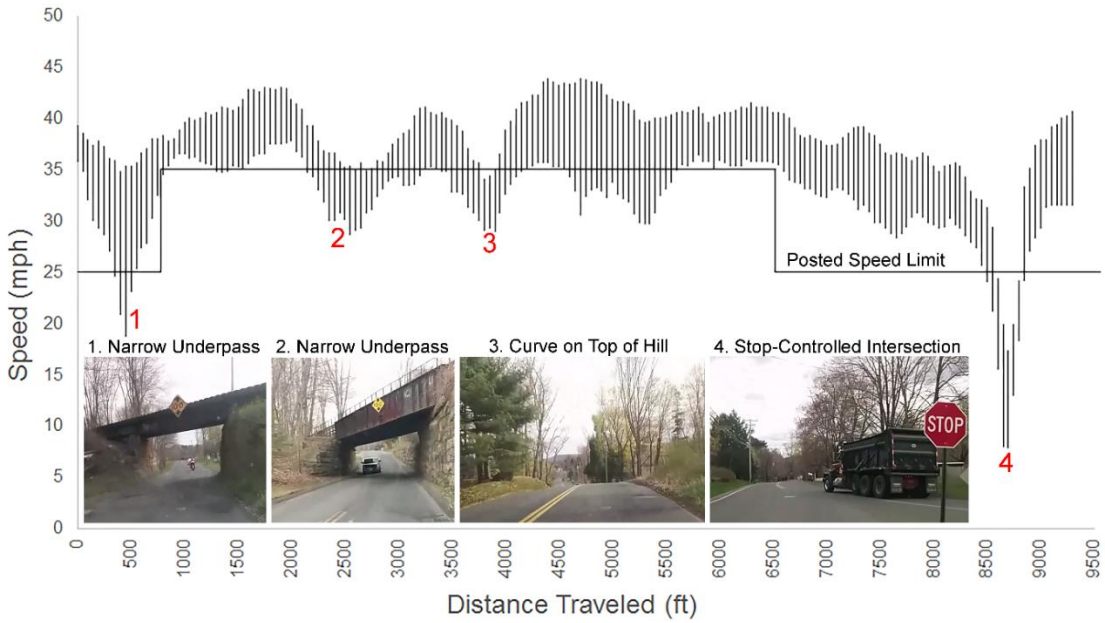


Figure 12. Southbound Ubipix drives with 15th to 85th percentile speed variance, showing four screenshots from the graphical interface.

6.5 Discussion

Based on the relationships explored in **Figure 10** and **Figure 11**, Ubipix data satisfies the need for continuous speed data along roadway segments and may, in certain situations, be a substitute for ATR or LiDAR data. Continuous speed provides numerous benefits over traditional data collection techniques such as: inexpensive collection, no need for specialized equipment beyond a smartphone, and short data turnaround time. The availability of continuous speed data has significant positive implications for both engineers and law enforcement, alike.

With the adoption of expert speed setting systems, such as USLIMITS2, 85th percentile speeds are no longer the sole input when determining a posted speed limit. While operating speeds are still the most valuable input into the system, spot speeds are no longer the sole determinant in the speed limit decision. With the push to set more appropriate speed limits, continuous speed data collection techniques, such as the one outlined in this

paper, may be a substitute for, or complement to, traditional spot speed methods. While these traditional methods may have more accuracy at a specific location due to larger sample sizes, these methods do not provide data along the entire speed zone being studied. For example, in the USLimits2 case study, the recommended speed limit was the same for all three methods when looking at the fixed points of data collection. However, the system recommended a higher speed limit at one point meaning a decision must be made as to the location of the speed zone change. In Massachusetts and Ohio for example, a speed zone must be at least 0.5 miles long and be rounded to 0.1 mile increments *(32)*. Data collection via a smartphone app could help practitioners decide the optimal speed zone length or provide information as to where additional data collection is needed.

This research endeavor outlined a methodology which may be used by engineers when setting speed limits. Ideally, continuous speed data obtained from trial runs, would be collected in coordination with ATR or spot speed data. As was presented in **Figure 8** and **Figure 9**, the speed data captured using ATRs can be used to verify the accuracy of the continuous speed data and simultaneously provide a larger sample of valid data points. This would allow continuous speed data to be used in analysis between the fixed points of ATR data collection. By plotting the continuous data versus the posted speed limit, roadway segments which may be good candidates for additional advisory speed signage, could be easily identified. Admittedly, there remains several next steps related to partnerships that would need to be established between large data suppliers and the agencies wishing to make use of the data; however the frequency of these partnerships is increasing on a routine basis.

Companies which monitor and provide real time traffic conditions mine data through continuous collection of their users' operating speeds. Access to this data would allow an agency to analyze vehicle speeds on any roadway. While these data would not include video and may not be as granular as the data in this study, it would allow engineers to alter speed limits or implement traffic calming designs. The main challenge associated with these partnerships would be the privacy issues of using these data. Perhaps it may require users to opt-in which may skew the pool of users or prohibitively reduce the sample size. Such partnerships would require unknown up-front costs and encounter possible privacy concerns. A future study should attempt to establish a partnership in order to quantify these costs and establish a methodology to mitigate privacy issues.

6.6 Conclusions

Continuous speed data was collected via a smartphone app, Ubipix, on a 1.75 mile rural road. These data were compared to spot speed data collected at eight locations along the route via a LiDAR gun and pneumatic tubes with ATRs. Despite the limited sample size of the continuous data, this method of collection still exceeded the three trial drives that are used to supplement the spot speed data as specified in the MassDOT guidelines. Ranges of continuous speeds observed were consistent with speeds collected via LiDAR and ATRs. A USLimits2 case study demonstrated the importance of data collection location to the outputted recommendation and suggested that continuous speed collection may provide valuable context to the practitioner conducting the speed study. In the future, partnerships may be developed with large data suppliers such as Inrix, Google, and TomTom to obtain this data without the need for trial runs. Future research should target refining the procedure outlined in this paper, establishing pilot partnerships with large data suppliers, and developing methods to automatically pull and analyze these data obtained

from such partnerships. These partnerships and methods would allow agencies to cost-effectively monitor operating speeds on their roadways. These data could enable a preventative approach to enforcement and speed management rather than waiting for serious or fatal crashes to occur.

CHAPTER 7
PERCEPTION OF TIME'S INFLUENCE ON DRIVER BEHAVIOR

7.1 Summary

Speeding greatly attributes to traffic safety with approximately a third of fatal crashes in the United States being speeding-related. Previous research has identified being late as a primary cause of speeding. In this driving simulator study, a virtual drive was constructed to evaluate how time pressures, or hurried driving, affected driver speed choice and driver behavior. In particular, acceleration profiles, gap acceptance, willingness to pass, and dilemma zone behavior were used, in addition to speed, as measures to evaluate whether being late increased risky and aggressive driving behaviors. Thirty-six drivers were recruited with an equal male/female split and a broad distribution of ages. Financial incentives and completion time goals calibrated from a control group were used to generate a *Hurried* and *Very Hurried* experimental group. As compared to the control group, *Very Hurried* drivers selected higher speeds, accelerated faster after red lights, accepted smaller gaps on left turns, were more likely to pass a slow vehicle, and were more likely to run a yellow light in a dilemma zone situation. These trends were statistically significant and were also evident with the *Hurried* group but a larger sample would be needed to show statistical significance. The findings from this study provide evidence that hurried drivers select higher speeds and exhibit riskier driving behaviors. These conclusive results have possible implications in areas such as transportation funding and autonomous vehicle design.

7.2 Project Motivation

The objective of this current study was to determine how time pressures, or drivers' perception of time, influenced speed choice and driver aggressiveness. This

objective was addressed by manipulating participants' perception of time and investigating the different outcomes. There were three overarching hypotheses:

1. Drivers in the control group, who are given no incentives, will make appropriate speed choices based on roadway conditions and posted speed limits and will not exhibit overly risky behaviors.
2. Drivers who are given an incentivized completion time goal, based on the 85th percentile completion time from the control group, will choose higher speeds and exhibit more aggressive driving behaviors than the control group as a whole.
3. Drivers who are given an incentivized completion time goal equal, based on the 15th percentile completion time from the control group, will similarly choose higher speeds and make riskier maneuvers than the control group but not the aggressive group.

7.3 Methods

A between subject experimental design was developed based upon existing literature to examine the effect that peoples' "perception of time" influences their driving behavior. The following section outlines the research tasks that were employed to address the objectives of this study.

7.3.1 Apparatus

A Realtime Technologies Inc. (RTI) driving simulator, depicted in **Figure 14** used in the current study is a full-cab, fixed-base, setup that includes a fully equipped 1996 Saturn sedan, with three screens subtending 135 degrees horizontally. At a resolution of 1024 x 768 pixels and at a frequency of 60Hz, the virtual environment is projected on each screen through a network of four advanced Realtime Technologies (RTI) simulator servers

equipped with high-end, multimedia chips. The participant sits in the driver's seat and operates the controls, just as he or she would in a normal car. A Dolby surround system consisting of side speakers and two sub woofers located under the hood of the car provides realistic wind, road and other vehicle noises with appropriate direction, intensity and Doppler Shift.

Figure 13. Driving simulator at Arbella Insurance Human Performance Lab, University of Massachusetts Amherst.

7.3.2 Measures and Associated Hypotheses

The independent variables were elements within the virtual drive listed in **Table 11**, which were the same for both the control and experimental groups. These various elements were used to evaluate drivers' aggressiveness and included: unprotected left turns with oncoming vehicles, red lights, a slow lead vehicle within a passing zone, progress updates throughout the drive, and dilemma zones. The dependent variables were the participants' reaction to these situations. **Table 11** contains the independent and dependent variables along with the hypothesized results from these variables.

Table 11. Variables and Associated Hypotheses

Independent Variable	Dependent Variable	Hypothesized Result
Speed checkpoints	Speed	Drivers in the experimental groups would select higher speeds.
Red lights	Acceleration profile after light turned green	Drivers in the experimental groups would accelerate faster after a red light.
Unprotected left turn with oncoming vehicle gaps: 3s, 3s, 1.5s, 2s, 2.5s, 3s, 3.5s, 4s, 4.5s, 6s, 10s	Size of gap accepted	Drivers in the experimental groups would accept smaller gaps than the control group.
Slow lead vehicle in passing zone	Willingness to pass	A larger percentage of drivers in the experimental groups would be willing to pass.
Dilemma zones	Willingness to abruptly stop for yellow light	A higher percentage of drivers in the experimental groups would be willing to run a yellow light in a dilemma zone situation.

7.3.3 Participants & Procedure

Before recruiting participants, three years of crash data (2012-2014) from the state of Massachusetts were analyzed to determine a logical distribution of participant ages. Since speed and driver aggression were a large focus of this study, the proportion of speeding-related crashes as a function of age was examined. While the proportion of crashes caused by speeding declined with age for both males and females, there seemed to be an inflection point around 30 years old when the decline became less pronounced. This inflection point is visualized in **Figure 14** by linear best fit lines for before and after 30 years old. Based on this data, participants were recruited to achieve an equal split of participants under 30 years old and over 30 years old in addition to an equal male/female split.

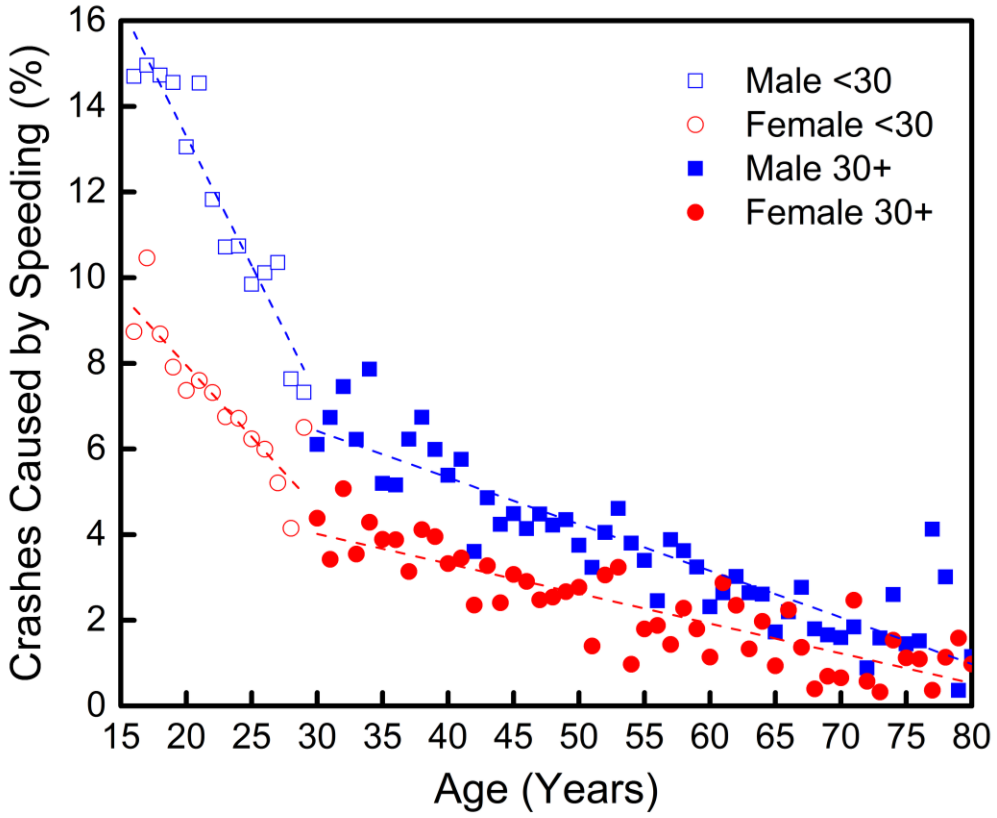


Figure 14. Proportion of crashes caused by speeding by age for males and females.

A total of 36 licensed drivers (18 years and older; 18 males and 18 females) from the greater Amherst, Massachusetts area were recruited as simulator participants. During recruitment, it was advertised that participants would be paid $15-30 compensation for their time. Participants were provided five minutes to drive in a practice training scenario to become familiar with the performance capabilities of the driving simulator prior to their experimental drive.

The experiment consisted of three groups, all of which drove the same virtual scenario, **Figure 15**. The first 12 subjects were placed in the control group. The overall travel times from the control group were then utilized to determine the incentive times used in the experimental groups. Each group consisted of three males and three females under

30 years old and three males and three females over 30 years old. One participant in the *Very Hurried* experimental group withdrew due to simulator sickness resulting in a partial dataset. A full comparison of participant demographics by group is shown in **Table 12.**

Table 12. Participant Demographics

Group	Male Participants	Female Participants	Driver Age (yr) Mean ± Std. Dev.	Driving Experience (yr) Mean ± Std. Dev.
Control	6	6	31.1 ± 12.0	14.9 ± 12.1
Experimental (*Hurried*)	6	6	32.9 ± 13.9	15.9 ± 14.3
Experimental (*Very Hurried*)	6	6	30.6 ± 9.2	13.3 ± 9.5

Before the virtual drive, participants completed a questionnaire which evaluated the frequency at which they exhibited aggressive driving tendencies. Participants were asked to rate each question either "Never", "Rarely", "Sometimes", or "Often". The questionnaire included 13 actions such as "Tailgate others to force move" and "Deliberately prevent other from passing". By assigning a value of 1-4 for Never to Often, a mean aggressiveness score could be computed for each participant and thus each group. The mean scores, with a lower value meaning less aggressive, were 1.74 (control), 1.56 (*Very Hurried*), and 1.72 (*Hurried*). None of these differences were statistically significant. When coupled with the balancing of age and sex the lack of statistical differences in the aggressiveness scores suggests that each of the groups were identical.

Control group: Participants in the control group were instructed to drive as they normally would. They were informed that the compensation range was simply used for recruiting purposes and that they would receive the full $30. Drivers in the control group

saw pop-up notifications throughout the drive at 25%, 50% and 75% drive progress. These notifications only displayed the percentage of drive complete and made no mention of time elapsed. The 85th percentile drive time was approximately 14 minutes and the 15th percentile time was approximately 16 minutes. These values were used as the incentive times for the aggressive and passive experimental groups, respectively.

Experimental groups: Participants were informed that they would receive $30 if they i) avoided getting in any crashes, ii) avoided any "tickets" and iii) finished the drive in under 14 minutes (*Very Hurried*) or under 16 minutes (*Hurried*). Otherwise, they would receive the baseline $15 as compensation. In order to conform to IRB requirements, all participants in the experimental group had to receive the full $30 compensation regardless of driving performance. However, this information was withheld from participants until after the drive to ensure that the incentive remained. Participants in these groups were also informed that they would see progress markers pop-up on the simulator screen at 25%, 50% and 75% drive progress. In addition to the drive progress, these pop-ups displayed the percentage of the 14/16 minutes that had elapsed and allowed participants to quickly evaluate whether they needed to speed up to meet the 14 or 16-minute deadline. These pop-ups would be analogous to drivers comparing their time remaining from GPS navigation versus their clock.

All procedures including informed consent, payment, and participant recruitment followed Protocol ID#: 2016-3343 as approved by the Institutional Review Board (IRB) of the University of Massachusetts.

7.3.4 Experimental Design

The entire drive consisted of a rural two-lane roadway with a 40-mph posted speed limit and contained nine signalized intersections, **Figure 15**. At two of the intersections,

drivers were instructed to turn left, and oncoming vehicles were scripted to test participants' gap acceptance while making a left turn. Four intersections were scripted to remain red until drivers reached the stop bar, and then would turn green. Sensors were built in so that participants' acceleration profile could be easily measured after each intersection. Two intersections near the end of the virtual drive were scripted so that the light turned from green to yellow when drivers were four seconds away, putting the driver in a dilemma zone situation.

There were five left horizontal curves and three right horizontal curves. Each curve had a length of 157 m and radius of 100 m. Lanes were 3.66 m wide (12 ft) with a 0.30 m shoulder (1 ft). There were no significant roadside objects or hazards.

Near the halfway point of the drive, a truck pulled out in front of the participants and traveled at 35 mph along a straightaway. A "Pass with Care" sign reminded participants that passing was allowed at that segment. After about a quarter mile, the slow-moving truck turned right at an un-signalized intersection which allowed participants who chose not to pass the truck to resume traveling at a free-flow speed. Ambient traffic throughout the drive was individually scripted so that oncoming traffic was consistent for all participants.

Progress updates were placed at the 25, 50, and 75% points of the drive and speed collection points were placed in a manner to capture speeds before and after each of these three updates. The total drive lasted 14-16 minutes. A full layout of the virtual drive is depicted in **Figure 15**.

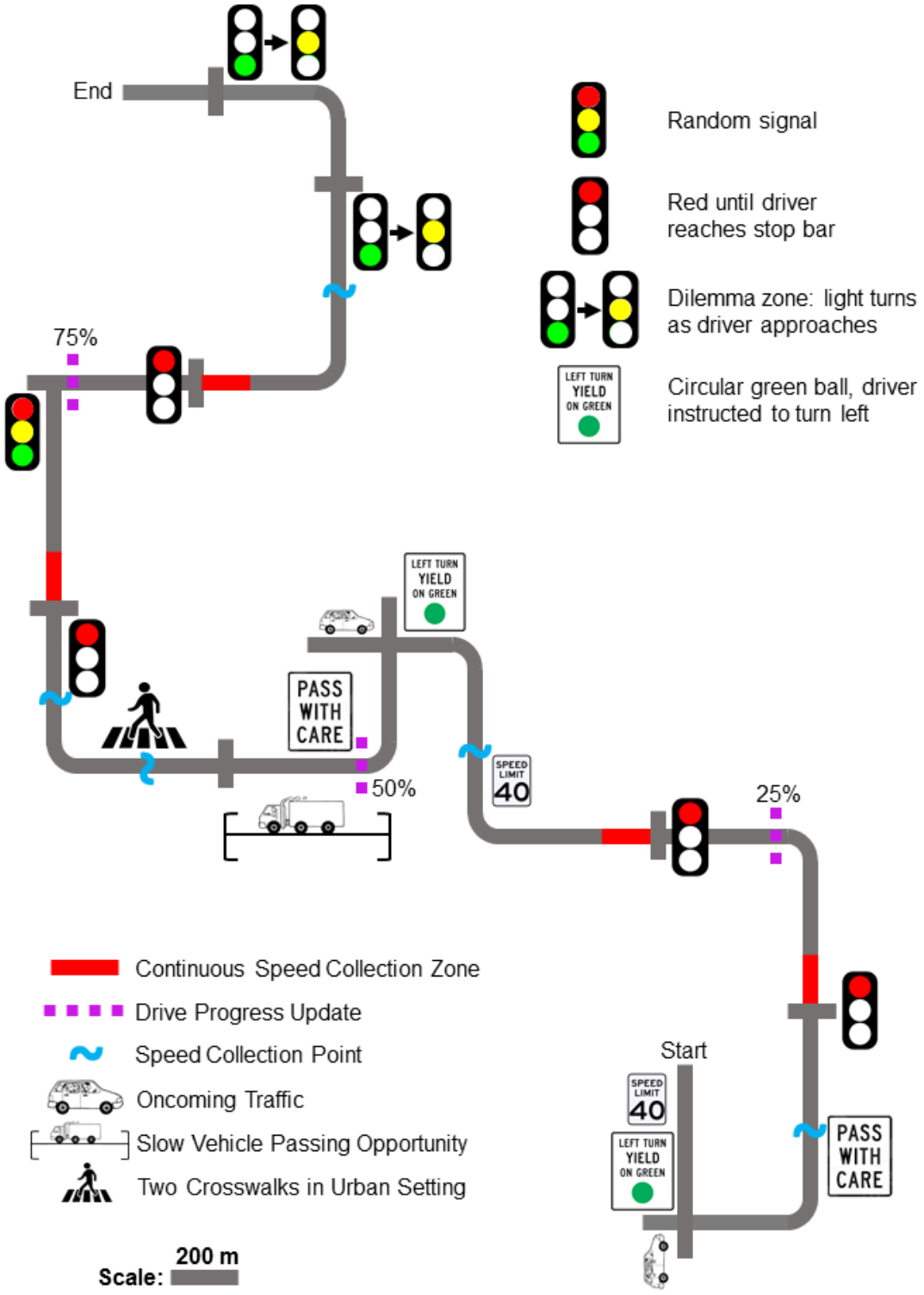


Figure 15. Schematic of virtual drive depicting elements participants encountered.

7.4 Results and discussion

The current driving simulator study examines how time pressures, or a driver's perception of time, impact driving performance. A between-subjects experimental design was utilized where each participant was placed in either the control, the *Hurried* experimental group, or the *Very Hurried* experimental group. The controlled laboratory settings allowed for the consistent manipulation of critical variables as well as the direct measurement of dependent variables.

All statistical tests conducted were unpaired two-sample Student's t-tests using the software package Minitab. All error bars represent 95% confidence intervals and a statistically significant difference ($p < 0.05$) from the control group is denoted by (*). Statistical significance ($p < 0.05$) between checkpoints within a group is denoted by a black bar.

7.4.1 Speed and Acceleration

The mean speed collected at five separate checkpoints is displayed in **Figure 16**. The drivers in the control chose a consistent speed throughout the duration of the drive, only statistically increasing their speed after the urban crosswalk section of the drive; checkpoint three ($M = 39.8$, $SD = 5.1$), checkpoint four ($M = 45.3$, $SD = 4.7$); $t(21) = -2.74$, $p = 0.012$. Participants within the *Hurried* group chose similar speeds as the control group and also only statistically increased their speed after the urban crosswalk setting, checkpoint three ($M = 41.1$, $SD = 5.0$), checkpoint four ($M = 46.5$, $SD = 5.0$); $t(21) = -2.69$, $p = 0.017$. This indicates that the time pressure placed on *Hurried* drivers was not enough to significantly alter their speed choice. Similar to the control and *Hurried* groups, *Very Hurried* drivers also statistically increased their speed after the urban crosswalk

setting, checkpoint three (M = 46.6, SD = 5.7), checkpoint four (M = 50.8, SD = 3.2); t(15) = -2.13, p = 0.050.

In the control and *Hurried* experimental groups, participants reduced their speed in the urban area with two crosswalks (speed checkpoint #3) as compared to their initial speed choice (speed checkpoint #1). By contrast, participants in the *Very Hurried* experimental group still selected a higher speed in the urban crosswalk setting as compared to their initial speed choice. While these differences were not statistically significant, this observation supports the overall hypothesis that *Very Hurried* drivers would be more likely to engage in riskier behavior.

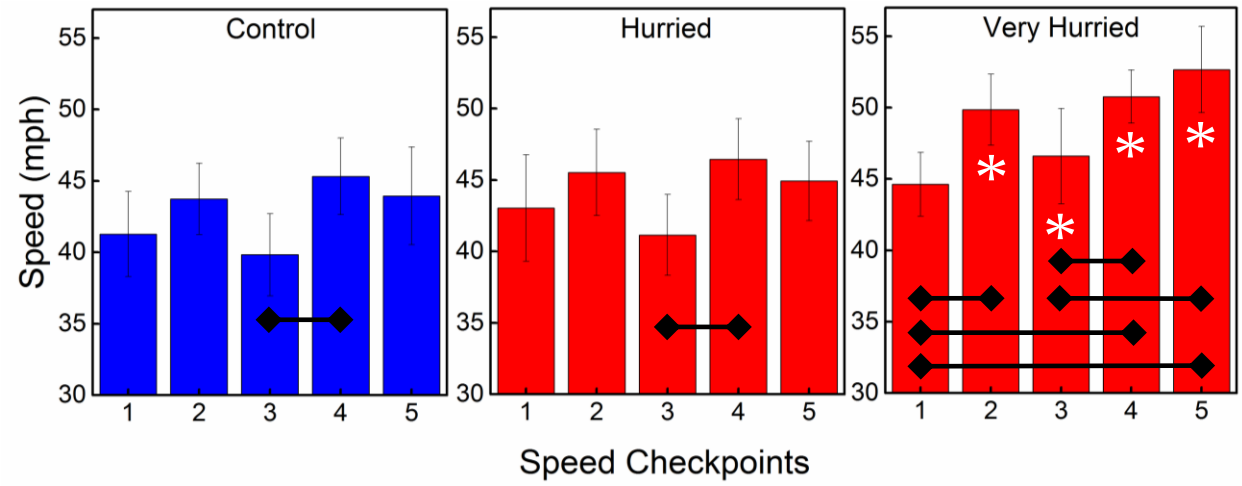


Figure 16. Mean speeds for each group at the five speed checkpoints.

The *Very Hurried* drivers initially selected a speed similar to both the control and *Hurried* drivers, indicating that all drivers initially had the same perception of time. After the first progress update, the *Very Hurried* participants drove at statistically higher speeds than the control group for the rest of the drive, (checkpoints 2-5, **Table 13**).

Table 13. Statistical Speed Checkpoint Comparison Versus Control Group

Checkpoint	*Hurried*			*Very Hurried*		
	df	**T**	**P-Value**	**df**	**T**	**P-Value**
1	20	-0.72	0.478	20	-1.76	0.093
2	21	-0.90	0.381	21	-3.40	0.003*
3	21	-0.64	0.528	20	-3.01	0.007*
4	21	-0.57	0.573	19	-3.27	0.004*
5	21	-0.44	0.666	20	-3.74	0.001*

(*) indicates statistical significance at 95% confidence.

With the exception of within the urban crosswalk setting, *Very Hurried* drivers statistically increased their speed after the first speed checkpoint, **Table 14**. The increased speed selection can be attributed to drivers gaining a better perception of time from the first progress update.

Table 14. Statistical Speed Checkpoint Comparisons Within *Very Hurried* Group

	1 vs.				**2 vs.**			**3 vs.**		**4 vs.**
	2	**3**	**4**	**5**	**3**	**4**	**5**	**4**	**5**	**5**
df	21	17	20	18	18	19	19	15	19	16
T	-3.06	-0.96	-4.13	-4.19	1.53	-0.57	-1.40	-2.13	-2.64	-1.05
P-Value	0.006*	0.349	0.001*	0.001*	0.143	0.575	0.176	0.050*	0.016*	0.310

(*) indicates statistical significance at 95% confidence.

Vehicle speeds were continuously collected 600 feet downstream of the four red lights, enabling average acceleration to be calculated over that segment. Participants in the control group, accelerated slower after red lights (M = 1.579, SD = 0.34) than participants in the *Very Hurried* experimental group (M = 1.963, SD = 0.32); t(91) = -5.63, p = 0.000. Participants in the *Hurried* experimental group accelerated faster than the control group but not as fast as the *Very Hurried* group, however these differences were not statistically significant, **Table 15**.

Table 15. Mean Accelerations after Red Lights and Inferential Statistics

Group	Sample Size	Mean Acceleration after Red Light (ft/sec^2)	Statistical Comparisons to Control		
			df	T	P-Value
Control	48	1.579	-	-	-
Hurried	48	1.658	93	-1.10	0.273
Very Hurried	46	1.963	91	-5.63	0.000*

(*) indicates statistical significance at 95% confidence.

7.4.2 Gap Acceptance

Drivers in all three groups encountered two unprotected left turns with oncoming vehicles with fixed gap sizes which became progressively larger. The critical gap, defined as the gap size at which 50% of drivers will accept and 50% will reject, was found by plotting the cumulative acceptance rate of the nine gap sizes presented to participants in the virtual drive, **Figure 17**. Similar to speed and acceleration results, drivers in the *Very Hurried* group were most aggressive and had the smallest critical gap (4.8 sec). *Hurried* drivers had a critical gap (6.0 sec), which was higher than the *Very Hurried* group but lower than the control (6.4 sec).

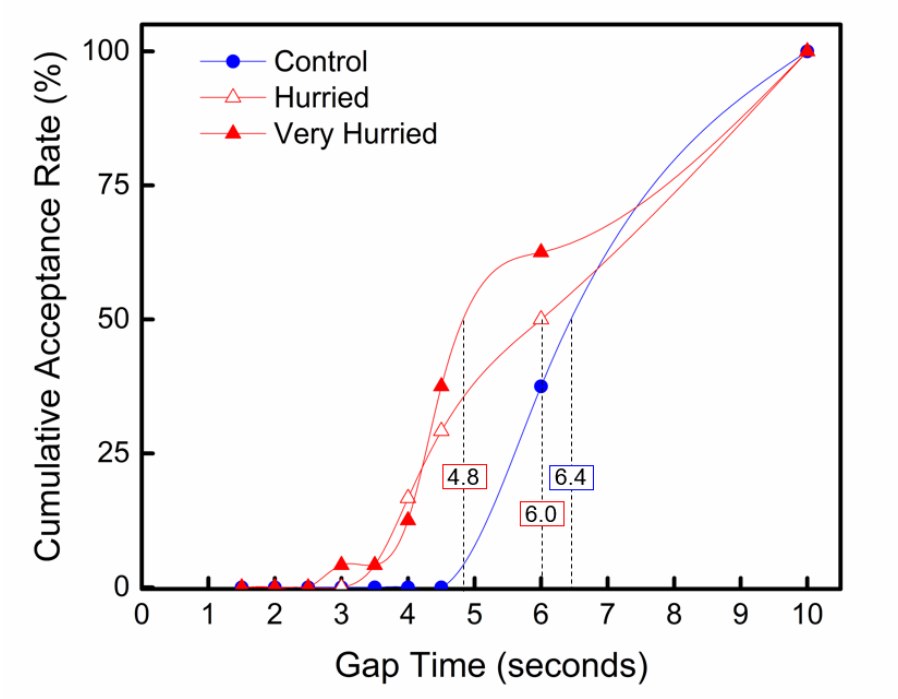


Figure 17. Cumulative gap acceptance rate for left turns with critical gaps.

Since a statistical test of the critical gap is not possible, a further examination of the differences was conducted by calculating the mean accepted gaps for the three groups, **Table 16**. For all three groups, the mean accepted gap was higher than the critical gap, which was likely due to the scripting of oncoming vehicles. If 7, 8, or 9 second gaps had been scripted, the mean for all three groups would likely have been lower. Nevertheless, the mean accepted gaps followed the same trends as the critical gap with the *Very Hurried* group selecting the most aggressive gap which was statistically different from the control.

Table 16. Mean Accepted Gaps for Unprotected Left Turns

Group	Mean Accepted Gap (sec)	Statistical Comparisons to Control		
		df	T	P-Value
Control	8.5	-	-	-
Hurried	7.5	42	1.53	0.134
Very Hurried	6.7	41	2.68	0.011*

(*) indicates statistical significance at 95% confidence.

7.4.3 Other Aggressiveness Measures

In addition to speed, acceleration, and gap acceptance as dependent measures, participants had the opportunity to pass a slow-moving vehicle and were subjected to dilemma zones, **Figure 18**. These scenarios further tested how time pressures affect driver performance. Around the halfway point of the virtual drive, a truck turned out in front of participants and drove at 35 mph, below the posted speed limit of 40 mph, and the participants did not know that the truck was going to turn off the road in a quarter mile. In the control group, only one driver chose to pass the slow-moving truck (8.3%). In the experimental groups, 4 of 12 *Hurried* drivers and 5 of 11 *Very Hurried* drivers passed the truck before it turned off the roadway, with the latter representing a statistically significant difference from the control (p = 0.029).

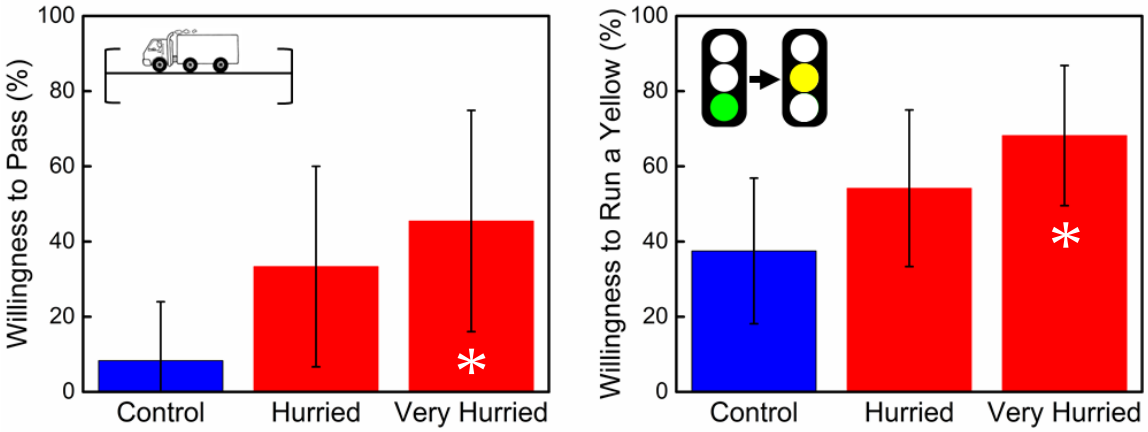


Figure 18. Willingness to (left) pass a slow vehicle and (right) run a yellow right.

When drivers were nearing the end of the drive, two signalized intersection put drivers in a dilemma zone situation forcing a stop or go decision. Specifically, these final two intersections were coded to be green as the drivers approached, but turn yellow when the driver was four seconds from the intersection. Participants in the control group ran the yellow light 9 of 24 times (38%), *Hurried* drivers ran the yellow 13 of 24 times (54%), and

Very Hurried drivers ran the yellow 15 of 22 times (68%). While *Hurried* drivers displayed riskier tendencies than the control group, this difference was not statistically significant. However, the difference between *Very Hurried* drivers and the control group was statistically significant ($p = 0.029$).

7.5 Conclusions

Thirty-six drivers participated in a driving simulator study which evaluated how time pressures, or a drivers' perception of time, impacted driving behavior. Travel times from a control group were used to determine the incentive thresholds for the experimental groups. The *Hurried* group had a goal time based on the passive drivers in the control and the *Very Hurried* group had a goal time based on the aggressive drivers in the control. Speeds, accelerations, gap acceptance, dilemma zones, and a passing zone all tested participants' aggressiveness and risk tolerance.

7.5.1 Evaluation of Hypotheses

The overarching hypothesis of this research project was that drivers would choose higher speeds and make riskier decisions when subjected to greater time pressures. Five specific hypotheses related to elements within the virtual drive were used to investigate the overarching hypothesis.

The first hypothesis predicted that drivers in the experimental groups would select higher speeds. After receiving the first progress update, *Very Hurried* drivers selected statistically higher speeds than the control. While *Hurried* drivers selected higher speeds than the control at all five speed checkpoints, these differences were not statistically significant.

The second hypothesis predicted that drivers in the experimental groups would accelerate faster after a red light. In the 600-foot segment following a red light, *Very*

Hurried drivers accelerated statistically faster than control drivers. While *Hurried* drivers also accelerated faster than the control, this difference was not statistically significant.

The third hypothesis predicted that drivers in the experimental groups would accept smaller gaps than the control group. *Very Hurried* and *Hurried* drivers had lower critical gaps and lower mean accepted gaps than the control group for the two unprotected left turns in the virtual drive. While a statistical comparison is not possible for critical gaps, the mean accepted gap for *Very Hurried* drivers was found to be statistically lower than the control group.

The fourth hypothesis predicted that a larger percentage of drivers in the experimental groups would be willing to pass. More *Hurried* drivers passed the slow-moving truck than control drivers, however due to the small sample this difference was not statistically significant. However, the *Very Hurried* drivers passed even more often than the control, resulting in a statistically significant difference.

The fifth, and final, hypothesis predicted that a higher percentage of drivers in the experimental groups would be willing to run a yellow light in a dilemma zone situation. Both *Hurried* and *Very Hurried* drivers were more likely than control drivers to run a yellow light in the two dilemma zone situations. However, only the *Very Hurried* group displayed a statistically significant difference from the control group.

The hypotheses examined in this study all showed that time pressures placed on drivers resulted in more aggressive, riskier behavior. The most notable statistical differences came from a comparison of the control and the *Very Hurried* group. The *Hurried* group, who had less of a time pressure, also displayed the same qualities of the *Very Hurried* drivers such as increased speeds and accelerations, a smaller critical gap, and

increased willingness to pass and run a yellow light as compared to the control. These findings between the *Hurried* and control would most likely become statistically significant with a larger sample size.

7.5.2 Limitations & Future Work

In this study, for the experimental groups, the consequences of getting in a crash or getting a ticket were equal to the consequence of not making the goal time, a loss of $15. In real life, the consequence of a crash would be demonstrably higher than being a few minutes late and the consequences of a ticket would also be higher. This alignment of risks and benefits likely exaggerated some of the results of this study. While it is not possible in a driving simulator to simulate physical harm from a crash or financial hardship from a speeding ticket, future studies could use different incentive and penalty structures in an attempt to validate the findings from this study.

An instrumented driver study could further investigate the impacts of time pressures by pairing the naturalistic driving data with a journal or log of the participant's daily schedule. Such a log would enable the linkage of the participant's time pressure on a specific day with their recorded driving performance.

7.5.3 Practical Implications

The practical implications from this research are abstract but nonetheless significant. The findings from this research indicate that drivers who are in a hurry select higher speeds and make riskier driving decisions. With the proliferation of GPS, drivers can monitor their projected arrival time in real-time and reroute themselves through local or neighborhood roads to avoid congestion. Projects primarily focused on relieving congestion, may also yield safety benefits on surrounding roads in the network, as drivers may make more aggressive decisions based on the difference between their remaining

projected travel time and their desired remaining travel time. This finding is important as funding agencies often have one pot of money for congestion projects, and another pot of money for safety projects. The results of this study indicate that a project aimed to reduce congestion may also legitimately claim safety benefits as well.

Finally, the findings from this study may influence autonomous driving design. As autonomous vehicles begin to join the market, manual driving will still be possible. In order to achieve maximum safety benefits, the vehicle should seek to have its driver remain in autonomous mode as often as possible. If the car can learn the driver's schedule and sense when they may be more hurried than usual, the autonomous mode may drive a little faster and be a little more aggressive than usual in order to meet the operator's preference to minimize their travel time. While this aggressive autonomous mode would be sub-optimal compared to factory settings, this mode would likely still be safer than a human driver.

CHAPTER 8
CONCLUSIONS AND FUTURE WORK

8.1 Conclusions

In the United States, nearly a third of fatal crashes are due to speeding *(8)*. There are many specific methods aimed to reduce speeding, such as crash data analysis, outreach campaigns, targeted enforcement, and understanding speed selection. Within these existing methods to reduce the safety impacts of speeding, there is a need for innovative approaches to data quality, speed data collection, and its utilization. In this dissertation, a multi-faceted approach was taken to improve roadway safety by examining the speeding-related crash designation, improving speed limit setting practices, and understanding the causes of speeding. An in-depth analysis of speeding-related crashes was constructed to build a logistic regression model to predict the probability of specific crashes being designated as speeding-related. A crash narrative review of the crashes identified by the model revealed that the Driver Contributing Code (DCC) in crash report could be improved, which would result in better crash data. A method was developed to capture continuous speed profiles from drivers instrumented with only a smart phone. These continuous profiles showed promise as a more accurate methodology to set speed limits. To understand driver speed choice and behavior, a driver simulator study was conducted to place drivers under different time pressures. When participants were hurried and very hurried, they chose higher speeds and engaged in riskier behavior, highlighting that projects to minimize congestion should also be able to use safety funding due to the safety benefits on the surrounding network. Overall, these projects targeted three specific areas of speeding research in an effort to improve traffic safety.

8.2 Practical Implications to Specific Stakeholders

The findings and recommendations that resulted from this dissertation work have practical implications to various groups such as the Massachusetts Department of Transportation, transportation engineers, police departments, and vehicle manufacturers.

8.2.1 MassDOT

The Massachusetts Department of Transportation may be one of the stakeholders most interested in this work specifically the projects discussed in chapters four, five, and six, which have direct implications to the agency. Currently, MassDOT is responsible for distributing crash data to consultants on roadway improvement projects. This process can consume considerable manpower and can often result in delays to the design of the project. MassDOT may be interested in working with a group, such as UMass Safe, who could alleviate the work load on MassDOT by distributing these data in a timely fashion and adding additional value by addressing know data quality concerns.

Data quality is always a concern to MassDOT, the recommendations from the work regarding speeding-related crashes would be of interest to the department as they play a large role in shaping the specifics of the crash report. Additionally, the logistic regression method, if refined, could be applied to other fields within the crash report and could increase the accuracy and completeness of crash data, overall.

The continuous speed collection method developed in this dissertation may be of interest to MassDOT as a replacement for the “Trial Runs” required in the state guidelines for speed studies. These trial runs require three people in the vehicle, one person driving, another monitoring the odometer, and the third monitoring the speed. A continuous speed method may require the same number of people, but the data would be easier to collect and provide more insight into the speed conditions of the roadway than trial runs.

8.2.2 Transportation Engineers

Transportation engineers could be interested in a service which provides crash reports in a timely manner and constructs the collision diagrams necessary for safety improvement projects. Engineers would also be invested in working with the most accurate data possible on their projects. While engineers would be unable to dedicate the time to perform their own analyses to improve the data, if the findings from the speeding-related crash work were applied then transportation engineers would surely benefit. For example, if the driver contributing code of "Driving too fast for conditions" was split into three separate codes, as recommended, engineers would be able to choose the most appropriate safety countermeasures Finally, the continuous speed method could be of great use to engineers working on traffic calming projects as such a method would allow specific locations along a roadway to be targeted for speed countermeasures.

8.2.3 Law Enforcement

Law enforcement officers are one of the primary stakeholders relating to the crash report form. During the first project, one of the common complaints we heard was that the various fields on the crash report were ambiguous. The speeding-related crash data project addressed one of these ambiguous fields, "Driving too Fast for Conditions", and recommended that the field be changed to be more intuitive. Having a more intuitive crash report would allow officers to be accurate on their reports and fill them out faster. The third project, developing a continuous speed data collection technique, could be of benefit to law enforcement. This procedure could allow for targeted speed enforcement. By having a continuous speed profile, officers could look for locations where the 85th percentile speeds are near the posted speed limit. Then, at this location the threshold for pulling a vehicle and issuing a speeding citation would be lower than what they may usually use. By

targeting enforcement in such a manner, law enforcement could increase safety by addressing the specific locations which may be unsafe. It may be the case that 5 mph over the speed limit in one location is more unsafe than 15 mph over the speed limit in another location, continuous speed profiles would enable officers to know when this is the case.

8.2.4 Researchers

While all the work contained in this dissertation would be of interest to researchers, project four, understanding how perception of time influences driver behavior, would be of particular interest to researchers working on connected and autonomous vehicles. Trust will be a main hurdle to overcome when introducing autonomous vehicles, by understanding the user's perception of time, the autonomous vehicle can make decisions that gain the user's trust. If the autonomous vehicle drives too slowly or too passively for the user's perception of time, the user may disable autonomous mode for manual mode, lessening the safety impacts of the technology. Smartphones can already learn your schedule and provide traffic notifications when you are on your way to or from work, vehicles will be able to do the same thing. If the vehicle knows that the user left the house too late, or if it reads the user's calendar and sees an appointment, the vehicle can drive a little more aggressively to appease the user's desire to minimize travel time. Ultimately, autonomous vehicles seek to remove the human element completely from the driving task, but until manual driving becomes illegal, humans will have the ability to override their autonomous vehicle if they are unhappy with its driving behavior.

8.3 Future Work

In this dissertation, recommendations to the crash report were identified, a new method to collect speeds continuously was developed, and the effect of time pressures on speed selection and driver behavior was better understood. Future work should include

furthering model prediction to improve crash data quality, integrating advanced technologies into transportation research and practices, and better understanding driver behavior through driving simulation.

8.2.1 Speeding-Related Crashes

The crash narrative review was conducted manually as the speeding-related designation is very subjective and it was important to accurately review data. In the future, an automated crash narrative review process as demonstrated by *(50)* could be developed for speeding-related crashes. An automated narrative review process could be used in conjunction with the standard crash information to most accurately identify crashes which were related to speeding. Additionally, the length of the crash narrative may be used in future models to gauge the confidence in the model's prediction. For example, a two paragraph crash narrative is likely to encompass all of the crash details, including the causation, whereas a one or two sentence narrative is not likely to provide sufficient detail.

The crash narratives which were reviewed only encompassed crashes occurring in Massachusetts. A more robust study could sample crash narratives from multiple states to see if the findings match the conclusions from this research. Another interesting analysis would be to stratify the model results by agency type to determine the extent to which reporting practices may be consistent within an agency. Crash reports filled out on highways and interstates always had a crash narrative, whereas only two-thirds of all crashes originally sampled contained a valid narrative. This may be due to the fact that these crashes fall within State Police jurisdiction, which may be indicative of consistent reporting practices and increased training within a single agency. Future studies could investigate the accuracy of the speeding-related crash designation as it relates to the responding officer's jurisdiction.

Finally, a similar method of prediction could be applied to every field within the crash report. Such a method could predicts the most likely input for each field within the crash report along with the confidence associated with that prediction. Validation of the model would include conducting a manual review of the fields, narrative and crash diagrams of 200 crash reports. The outcome of this potential research would be a statistical model which could be broadly applied to crash datasets to quickly identify inaccuracies.

8.2.2 Continuous Speed Data

Companies which monitor and provide real time traffic conditions mine data through continuous collection of their users' operating speeds. Access to this data would allow an agency to analyze vehicle speeds on any roadway. While these data would not include video and may not be as granular as the data in this study, it would allow engineers to alter speed limits or implement traffic calming designs. The main challenge associated with these partnerships would be the privacy issues of using these data. Perhaps it may require users to opt-in which may skew the pool of users or prohibitively reduce the sample size. Such partnerships would require unknown up-front costs and encounter possible privacy concerns. A future study should attempt to establish a partnership in order to quantify these costs and establish a methodology to mitigate privacy issues.

The rise of unmanned aerial vehicles (UAVs), or drones, provides an opportunity to revolutionize traffic data collection techniques. Previously, aerial studies were infeasible due to the high costs of helicopters and studies conducted at ground level could only capture a specific location. A future study should compare the use of UAVs to traditional speed data collection instruments in order to evaluate the feasibility of UAVs as a traffic data collection tool.

8.2.3 Perception of Time

Future studies could use different incentive and penalty structures in an attempt to validate the findings from this dissertation. By testing other incentive/penalty structures, the cost/benefit balance of speeding can be manipulated. For example, in this research, a crash was equal in consequence to getting a speeding ticket, and both were equal consequence to being late to the destination. Future studies should make the cost of a crash higher than a speeding ticket, and make the cost of a speeding ticket higher than the cost of being late.

An instrumented driver study could further investigate the impacts of time pressures by pairing the naturalistic driving data with a journal or log of the participant's daily schedule. Such a log would enable the linkage of the participant's time pressure on a specific day with their recorded driving performance.

REFERENCES

1. National Highway Traffic Administration. Traffic Safety Facts 2013. July. 1–238. http://www-nrd.nhtsa.dot.gov/Pubs/812124.pdf. Accessed Nov. 4, 2015.

2. Center for Disease Control. *Injury Prevention & Control: Data & Statistics (WISQARSTM)*. 2013.

3. World Health Organization. *Global status report on road safety*. 2013.

4. Singh, S. *Critical reasons for crashes investigated in the National Motor Vehicle Crash Causation Survey*. 2015.

5. Aarts, L., and I. van Schagen. Driving speed and the risk of road crashes: a review. *Accident Analysis & Prevention*, Vol. 38, No. 2, 2006, pp. 215–24.

6. Rosén, E., H. Stigson, and U. Sander. Literature review of pedestrian fatality risk as a function of car impact speed. *Accident; analysis and prevention*, Vol. 43, No. 1, 2011, pp. 25–33.

7. Mao, Y., J. Zhang, G. Robbins, K. Clarke, M. Lam, and W. Pickett. Factors affecting the severity of motor vehicle traffic crashes involving young drivers in Ontario. *Injury prevention : journal of the International Society for Child and Adolescent Injury Prevention*, Vol. 3, No. 3, 1997, pp. 183–9.

8. Liu, C., and C.-L. Chen. *An Analysis of Speeding-Related Crashes: Definitions and the Effects of Road Environments*. 2009.

9. MassDOT. *Procedures for Speed Zoning on State and Municipal Roadways*. 2012.

10. Neuman, T. R., K. L. Slack, K. K. Hardy, V. L. Bond, I. Potts, and N. Lerner. *Guidance for Implementation of the AASHTO Strategic Highway Safety Plan: Volume 23 A Guide for Reducing Speeding-Related Crashes*. 2009.

11. Monsere, C. M., R. L. Bertini, P. G. Bosa, D. Chi, C. Nolan, and T. A. El-Seoud. *Comparison of Identification and Ranking Methodologies for Speed-Related Crash Locations*. 2006.

12. Council, F. M., M. Reurings, R. Srinivasan, S. Masten, and D. Carter. *Development of a Speeding-Related Crash Typology*. 2010.

13. *Speed Management Program Plan*. Washington DC, 2014.

14. McKnight, A. J., and A. S. McKnight. Young novice drivers: careless or clueless? *Accident Analysis & Prevention*, Vol. 35, No. 6, 2003, pp. 921–925.

15. Pollack, K. M., N. Yee, M. Canham-Chervak, L. Rossen, K. E. Bachynski, and S. P. Baker. Narrative text analysis to identify technologies to prevent motor vehicle crashes: examples from military vehicles. *Journal of safety research*, Vol. 44, 2013, pp. 45–9.

16. Swansen, E., I. A. McKinnon, and M. A. Knodler. Integration of Crash Report Narratives for Identification of Work Zone-Related Crash Classification. 2013.

17. Graves, J. M., J. M. Whitehill, B. E. Hagel, and F. P. Rivara. Making the most of injury surveillance data: using narrative text to identify exposure information in case-control studies. *Injury*, Vol. 46, No. 5, 2015, pp. 891–7.

18. Dube, C., C. D. Fitzpatrick, and M. A. Knodler. Improved Identification of Distraction-Related and the Impact of Distraction-Free Driving Laws. 2016.

19. Gates, T. J., S. D. Schrock, and J. A. Bonneson. Comparison of Portable Speed Measurement Devices. *Transportation Research Record*, No. 1870, 2004.

20. Jasrotia, R. *Effect on Speed Distribution due to Intrusive and Non-Intrusive Portable Speed Measurement Devices*. University of Kansas, 2011.

21. Bagdade, J., D. Nabors, H. McGee, R. Miller, and R. Retting. *FHWA: Speed Management: A Manual for Local Rural Road Owners*. 2012.

22. Middleton, D., and R. Parker. *Vehicle Detector Evaluation*. Austin, TX, 2002.

23. Lu, X.-Y., P. Varaiya, R. Horowitz, Z. Guo, and J. Palen. Estimating Traffic Speed with Single Inductive Loop Event Data. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2308, 2012, pp. 157–166.

24. Marti, M., R. Kuehl, and S. Petersen. *Traffic Data Collection Improvements*. St. Paul, MN, 2014.

25. Dingus, T. a., S. G. Klauer, V. L. Neale, A. Petersen, S. E. Lee, J. Sudweeks, M. a. Perez, J. Hankey, D. Ramsey, S. Gupta, C. Bucher, Z. R. Doerzaph, J. Jermeland, and R. R. Knipling. *The 100-Car naturalistic driving study phase II – Results of the 100-Car field experiment*. Washington DC, 2006.

26. Intelligence to Drive. *i2D Fact Sheet*. 2015.

27. Ltd., iGeotech T. Ubipix - Technology. http://www.ubipix.com/the-technology/. Accessed Jul. 1, 2015.

28. Parker, M. R. *Effects of Raising and Lowering Speed Limits on Selected Roadway Sections*. 1997.

29. AASHTO. *Strategic Highway Safety Plan: A Comprehensive Plan to Substantially Reduce Vehicle-Related Fatalities and Injuries on the Nation's Highways*. 1998.

30. AASHTO. *A Policy on Geometric Design of Highways and Streets, 6th Edition*. 2011.

31. Donnell, E. T., S. C. Himes, K. M. Mahoney, and R. J. Porter. Understanding Speed Concepts: Key Definitions and Case Study Examples. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2120, 2009.

32. Forbes, G. J., T. Gardner, H. McGee, and R. Srinivasan. *Methods and Practices for Setting Speed Limits: An Informational Report*. Washington DC, 2012.

33. Freedman, M., D. De Leonardis, A. Polson, S. Levi, and E. Burkhardt. *Field Test of the Impact of Setting and Enforcing Rational Speed Limits in Gulfport , Mississippi*. 2007.

34. Knodler, M. A., D. Hurwitz, and H. A. Rothenberg. Evaluation of Rationally Implemented Speed Limits on Collector Roadways. 2008.

35. Son, H., M. D. Fontaine, and B. Park. Long-Term Speed Compliance and Safety Impacts of Rational Speed Limits. *Journal of Transportation Engineering*, Vol. 135, No. 8, 2009.

36. Srinivasan, R., D. L. Harkey, D. Tharpe, R. Sumner, and M. R. Parker. Development of a Web-Based Expert System for Setting Speed Limits in Speed Zones. 2008.

37. Srinivasan, R., M. Parker, D. Harkey, D. Tharpe, and R. Sumner. *Expert System for Recommending Speed Limits in Speed Zones*. 2006.

38. Minnesota Department of Transportation. *Methods for Setting Posted Speed Limits*. 2012.

39. Schroeder, P., L. Kostyniuk, and M. Mack. *2011 National Survey of Speeding Attributes and Behaviors*. 2011.

40. Beck, K. H., M. Q. Wang, and A. F. Yan. Hurried Driver Dispositions: Their Relationship to Risky Traffic Behaviors. *American Journal of Health Behavior*, Vol. 36, No. 1, 2012, pp. 86–95.

41. Beck, K. H., S. B. Daughters, and B. Ali. Hurried driving: Relationship to distress tolerance, driver anger, aggressive and risky driving in college students. *Accident Analysis & Prevention*, Vol. 51, 2013, pp. 51–55.

42. Zhang, W., K. Wang, L. Wang, Z. Feng, and Y. Du. Exploring factors affecting pedestrians' red-light running behaviors at intersections in China. *Accident Analysis & Prevention*, Vol. 96, 2016, pp. 71–78.

43. Charron, C., A. Festoc, and N. Guéguen. Do child pedestrians deliberately take risks when they are in a hurry? An experimental study on a simulator. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 15, No. 6, 2012, pp. 635–643.

44. Bertola, M. A., S. A. Balk, and J. Shurbutt. Evaluating Driver Performance on Rural Two-Lane Horizontal Curved Roadways Using a Driving Simulator. *USDOT: FHWA*, No. FHWA-HRT-12-073, 2012, pp. 1–37.

45. NSF Innovation Corps. http://www.nsf.gov/news/special_reports/i-corps/. Accessed Feb. 1, 2016.

46. Stoltzfus, J. C. Logistic regression: a brief primer. *Academic emergency medicine : official journal of the Society for Academic Emergency Medicine*, Vol. 18, No. 10, 2011, pp. 1099–104.

47. Hosmer Jr., D. W., S. Lemeshow, and R. X. Sturdivant. *Applied Logistic Regression*. John Wiley & Sons, Inc., Hoboken, NJ, USA, 2013.

48. Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, Vol. 20, No. 1, 1960, pp. 37–46.

49. Smeeton, N. C. Early history of the kappa statistic. *Biometrics*, Vol. 41, 1984.

50. Bertke, S. J., A. R. Meyers, S. J. Wurzelbacher, A. Measure, M. P. Lampl, and D. Robins. Comparison of methods for auto-coding causation of injury narratives. *Accident; analysis and prevention*, Vol. 88, 2016, pp. 117–23.

51. Vincenty, T. T. Vincenty (1975): Geodetic inverse solution between antipodal points. 1975.

52. Witte, T. H., and A. M. Wilson. Accuracy of non-differential GPS for the determination of speed over ground. *Journal of biomechanics*, Vol. 37, No. 12, 2004, pp. 1891–8.

53. FHWA. HSIS - Highway Safety Information System. http://www.hsisinfo.org/index.cfm. Accessed Jan. 4, 2016.