Studying Driving Risk Factors using Multi-Source Mobile Computing Data

Xianbiao Hu[1*], Yi-Chang Chiu[2], Yu-Luen Ma[3], Lei Zhu[4]

[1*] Director of R&D, Metropia, Inc., Tucson, AZ, USA. xb.hu@metropia.com

[2]Associate Professor, Department of Civil Engineering and Engineering Mechanics, University of Arizona, Tucson, AZ, USA, chiu@email.arizona.edu

[3]Professor, Katie School of Insurance and Financial Services, Illinois State University, Normal, Il, USA, yma@ilstu.edu

[4]PhD student, Department of Civil Engineering and Engineering Mechanics, University of Arizona, Tucson, AZ, USA, leizhu@email.arizona.edu

ABSTRACT

Traffic congestion can largely be attributed to the issues related with driving behavior, which may cause vehicle crash, stop-and-go traffic due to frequent lane changing behaviors, etc., and makes the driving behavior research also of significance in the realm of traffic management and demand management. The emergence and subsequent rapid advances with new information and communication technologies (ICT) now offers the capability of collecting high-fidelity and high-resolution trajectory data in a cost-effective manner. In this research, we use a smartphone app to collect data for the purpose of studying driving risk factors. What's unique about the data in this research is its backend server also estimates traffic speed and volume for each link that the vehicle traverses. In order words, the data collected with build-in GPS modules in the smartphone include not only the vehicle spatial-temporal dimension location, which could be used to correlate the network geography attributes and/or real-time traffic condition, but also the detailed information about the vehicle dynamics including speed, acceleration, and deceleration, whereby a driver's control and maneuver of a vehicle can be analyzed in detail. Such type of dataset combining both user trajectory and link speed/volume information is rarely seen in prior research, permitting a unique opportunity to link critical traffic congestion factors leading to driving behavior and crash potential.

In this paper, the overall research framework used in this research is presented, which mainly includes data collection, data processing, calibration and analysis methodology. A preliminary case study — including data summary statistics and correlation analysis — is also presented. The results of our study will further existing knowledge about driving exposure factors that are closely linked to crash risk, and provide the foundation for advanced forms of Usage Based Insurance.

Keywords: Driving Risk Factors, Smartphone Trajectory Data, Information and Communication Technologies (ICT), Usage Based Insurance (UBI), Pay-As-You-Drive-And-You-Save (PAYDAYS)

1. INTRODUCTION

It is generally acknowledged that consistent driving behavior will smooth the traffic flow and as a result, have a beneficial impact on the traffic flow efficiency, while some other driving behaviors, such as frequent lane change, speeding, hard brakes may bring disruption of the traffic flow, and even cause vehicle crash which will significantly lower the network operational efficiency. In this regard, driving risks evaluation is not only of concern to the safety related study, but also to the traffic management and demand management overall. Traditional traffic safety related research is usually based on the past crash history and uses it to identify hazardous locations, as well as analyze the potential risk factors contributing to the crash events. This kind of research can only be performed at the aggregated level since it lacks the capability of tracking individual drivers continuously in order to collect the personalized driving data, which is a prerequisite of the microscopic safety analysis. The emergence and subsequent rapid advances with new information and communication technologies (ICT) such as GPS devices or smartphones now offer the capability of collecting high-fidelity and high-resolution travel data in a cost-effective manner. These technologies also permit continuous data collection so long as the vehicle/device is in operation. Given the ever growing cellular phone market (with 91% of adults in the United States own cellular phones and 56% own a smartphone) *(1)*, it has now become much easier to track and understand traveler activity, and travel patterns.

As a result, recent research started to examine the idea of applying ICT in the field experiments to collect the personal driving data and use it for safety analysis purpose. Examples of such experiments can be found in *(2-4)*. Such data usually include not only vehicle spatial-temporal dimension location, but more importantly also speed, acceleration, and deceleration, which are frequently used to evaluate a driver's driving behavior.

While all these efforts can be considered as already taking a major leap forward, one can argue that other important risk factors such as location information (e.g. an intersection of higher risk than a mid-link) and traffic flow are not traceable by the GPS devices and are absent in the previous researches. For example, if a driver exhibits stop-and-go or abrupt accelerate/decelerate behavior, without supplemental information on traffic condition, it is usually difficult to tell if this is simply due to the driver's behavior or because of heavy traffic conditions. Another intuitive idea is to compare the driver's instantaneous speed with the surrounding traffic, if a driver is observed to be consistently driving at a speed much higher than the other people around him/her, it's reasonable to assume most likely this driver is driving more aggressively than average drivers.

In this research, we use Metropia mobile app to collect data for research purpose *(5)*. When a user starts a trip with the app, the internal GPS module is activated and starts to record the second-by-second latitude/longitude data location and instantaneous moving speed. These data allow detailed position, velocity, acceleration and deceleration data to be stored and analyzed online or offline. *Further, what's unique about data in this research is its backend server also estimates traffic speed and volume for each link that the vehicle traverses*. In other words, the data collected with build-in GPS modules in the smartphone include not only the vehicle spatial-temporal dimension location, which could be used to correlate the network geography attributes and/or real-time traffic condition, but also the detailed information about the vehicle dynamics including speed, acceleration, and deceleration, whereby a driver's control and maneuver of a vehicle can be analyzed in detail. Such type of *multi-source* dataset including both user trajectory and link speed/volume information is rarely seen in prior research, permitting a unique opportunity to link critical traffic congestion factors leading to driving behavior and crash potential. With both data linked together, one can discern hazards caused by driving behavior and/or congestion levels.

In this research, we propose to answer the following major research question: How can we collect and use multi-source mobile computing data, including vehicle trajectories, associated geometric network information, link traffic volume and speed data, time of day information, to more accurately identify driving exposure factors, and quantify the relationship between hazard, driving behavior and congestion levels? In this paper, we present our overall research framework, which includes data collection, data processing, calibration and analysis methodology to answer this question. More specifically, model calibration has been conducted to calibrate a few key parameters of the system, risk measurements have been designed to characterize driving risks, and correlation analysis has been performed to reveal the relations between various risk measurements. A preliminary case study — including data summary statistics and correlation analysis — will also be presented.

The rest of this paper is organized as follows: Section 2 reviews the relevant literatures on the past research effort. Section 3 presents the overall methodology being developed for this research, including the general research workflow, data collection and processing procedure, calibration, and analysis methodology. Section 4 discusses the preliminary case study results. Section 5 concludes this research and presents future research directions.

2. LITERATURE REVIEW

Traditional traffic safety related research is usually based on the past crash history and uses it to identify hazardous locations, as well as analyze the potential risk factors contributing to the crash events. Examples of this kind of research can be found in *(6-10)*, they can only be performed at the aggregated level since it lacks the capability of tracking individual drivers continuously in order to collect the personalized driving data, which is a prerequisite of the microscopic safety analysis.

Researches of applying the latest ICT to collect data and evaluate driving behavior started to show up in the last decade. A recent effort in the individual driving behavior is the 100-car naturalistic driving study *(2)*. Drivers are monitored and recognized as unsafe, moderate safe and safe according to frequencies of crashes/near-crashes. The results indicate that hard braking, inattention, and tailgating are the top three at-risk behaviors among drivers. Unsafe drivers are more likely to engage in the at-risk behaviors and decelerate/swerve greater than the safe drivers. The results also imply that improper braking and inappropriate speeds are positively related to crash/near-crashes. Different traffic and weather conditions are also studied separately for the driving behavior and crash risk. The unsafe drivers drive more aggressively regardless of traffic conditions.

Another example of applying new ICT in the driver behavior evaluation can be found in *(4)*, where a framework for profiling drivers by at-risk behavior using driving pattern, spatial and temporal characteristics and driver characteristics was proposed. Second-by-second GPS data observations are collected from 106 drivers in Sydney over several weeks. Behavioral measures are summarized as maximum, average, minimum and standard deviation of speed, acceleration and deceleration, distance at 75% of speed limit or over speed, number of sharp celeration), etc.

Studies in driving pattern regarding fuel consumption and emission can also be found in the literatures. Both safe and green driving style drivers are observed to have less stop and hard braking, smooth acceleration and deceleration and moderate engine speed. *(11)* and *(12)* matched the driving pattern data to the transportation network and examined the variation of the driving patterns as a function of external conditions. Five cars were used in daily driving by 30 families for two weeks. The driving patterns are measured by aggregated speed, acceleration/deceleration, oscillation of speed and celeration, power use, engine speed and gear changing behavior for different street types. The parameters are defined as percent of time speed < 2km/h, frequency of local max/min values of speed curve, percent of time at acceleration over , speed celeration () distribution. A linear regression model is proposed to examine the relations between a certain driving pattern with street characteristics, traffic flow conditions, weather and drivers. The impact of these driving behaviors on emissions and fuel-use is further investigated *(13)*. This study suggests strategy for eco-driving to avoid heavy acceleration, large power demands and high engine speeds.

A number of studies have explored the relationship between traffic accidents and traffic volume or traffic flow. *(14)* discovers strong relationships between traffic flow conditions (mean volume, median speed, and temporal variations in volume and speed) and crash occurrence. Another research suggests the probability of an accident is high in congested conditions because of the high traffic density, short following distances, unexpected stops-and-starts of other vehicles, and emotional/aggressive driving behaviors *(15)*. However, these crashes tend to occur at slower speeds and cause fewer casualties and less external cost. On a road without congestion, the average traffic speed is relatively higher, resulting in more serious injuries or fatalities. A telematics system has been successfully demonstrated to be useful for improving motor carrier efficiency in *(16)*. In this particular field study, the research team demonstrated that telematics can be used to monitor and improve safe driving behavior as well as to monitor and improve fuel economy in trucks.

A spatial analysis of the relationship between congestion and pedestrian safety in London is conducted in *(17)*. No conclusive impacts of congestion on safety can be confirmed. The congestion level in this study is measured using location and employment variables. *(18)* employs a new approach for congestion index according to travel time and free flow travel time, and explore the effects of traffic congestion on road accidents using a similar spatial approach on a London orbital motorway. Traffic congestion is found to have little or no impact on the frequency of injured and fatal crashes.

In the last decade the Data Acquisition System (DAS) or In-Vehicle Data Recorders (IVDR) have also been introduced to collect detailed driving behavior data, but the usage was limited due to the high hardware cost, and the research sample size are usually insufficient *(19, 20)*. Furthermore, with those devices, drivers may not act exactly in the same way as they normally do during the experiment period; in other words, these methods may introduce certain biases to the experiment.

3. METHODOLOGY

3.1 Overall research framework

In this section, a comprehensive research framework and approach to analyze vehicle use and/or driver behavior data to advance knowledge about driving exposure factors that are closely linked to crash risk is presented. The main modules in this research include the following:

- Data collection: specifying the raw data we are collecting
- Data processing: describing how raw data will be processed to extract useful information
- Calibration: specifying how the threshold of hard brake will be calibrated
- Analysis Methodology: explaining how the data will be used in this research

Figure 1 presents the overall research work flow. After the GPS trajectory is collected and sent back to the cloud server, those data will be stored together with network dataset and time-dependent traffic condition and processed for research purposes. More details will be explained in the following sections.

3.2 Data collection

In this research, in order to answer the question of what are the factors that are attributable to auto accidents, and how can we use various types of data to more accurately quantify the relationship between driving, traffic and auto accidents, we propose to collect various risk factor data through Metropia Mobile, a recently developed smartphone-based app designed to improve mobility management. The primary data are categorized into:

1) Vehicle trajectory data
2) Roadway geometry
3) Time-varying traffic dynamic

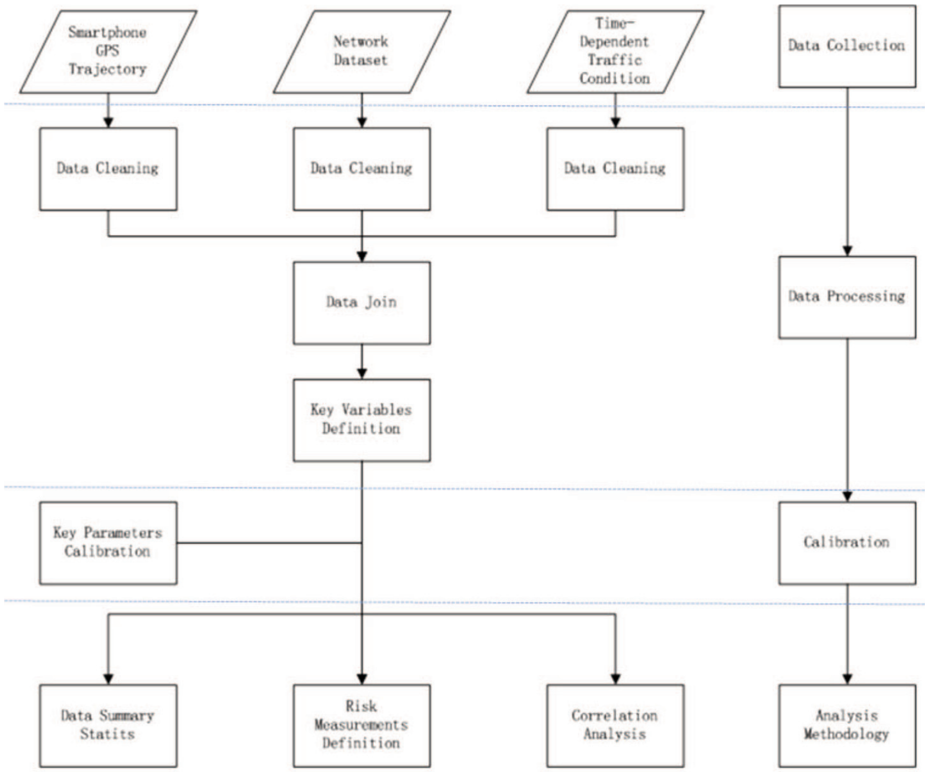

Figure 1. Overall research framework

3.2.1 vehicle trajectory data

Detailed trajectory data and driving behavior data for each trip are collected during the trip validation process. When a user starts a trip with the app, the internal GPS module built in the smartphone is activated and starts to record the second-by-second data. These data, including detailed position such as latitude, longitude velocity and acceleration will be collected at fine time interval and sent back to the cloud server, where they will be stored and used for further analysis. Following is a detailed list of attributes of the trajectory data.

- UserID: The unique ID of the user
- TripID: The unique ID of the particular trip
- Latitude: The latitude of this particular GPS point
- Longitude: The longitude of this GPS point
- Altitude: The altitude of this GPS point
- Heading: The heading of vehicle at this moment
- Speed: The instantaneous travel speed of the vehicle

- Acceleration: The instantaneous acceleration value of the vehicle
- Timestamp: The timestamp of this GPS point
- LinkID: The ID of the road segment that user is currently driving on

3.2.2. Roadway geometry

The geographic network dataset is stored on cloud server and includes a set of links and nodes, i.e. $G = (N, A)$ where N is the set of nodes and the set of links. The key attributes relevant with this research include location of the intersection, type of a road segment, speed limit and so on. The main attributes of network data can be found below.

- LinkID: The ID of a road segment
- Speed Limit: The speed limit of a road segment
- Facility Type: The type of links, such as freeway, arterial, local street, ramp
- Number of lanes: Number of lanes of a road segment
- Link length: The length of a road segment
- Start nodeID: ID of the start node for a particular road segment
- Start latitude: The latitude of the start node
- Start Longitude: The longitude of the start node
- End nodeID: The ID of the end node for a particular road segment
- End Latitude: The latitude of the end node
- End Longitude: The longitude of the end node

3.2.3. Time-varying traffic dynamic

On the cloud server, the time-varying traffic dynamic, i.e. the traffic dataset including time-dependent traffic speed, traffic flow, etc. for each traversed link is hosted and keeps updating every 5 minutes. Below is a list of attributes.

- LinkID: The unique ID of a road segment
- Date: The date of the traffic data
- Timestamp: The timestamp of the traffic data
- Link average speed: The instantaneous average speed of this road segment
- Link volume: The instantaneous traffic volume of this road segment

3.3. Data processing

3.3.1 Data cleaning

Data cleaning process will be applied prior to the data being used in the research, which is very important to ensure the quality of system input and accurate analysis result. Some typical invalid data scenarios include but are not limited to:

- GPS being inaccurate under certain scenarios, such as around tall buildings or indoors
- GPS being inaccurate due to network connection issues
- GPS point “jumps” when the speed of vehicle moving is low
- Issues with the data transmission

- Geometric network quality issue

3.3.2. Data joins and sample data

For the analysis purpose, if we perform data joins between the multi-source data in Section 3.2, in the end for each GPS point, we will have the following attributes:

- UserID
- TripID
- Latitude
- Longitude
- Altitude
- Speed
- Acceleration
- Timestamp
- LinkID
- Distance to Downstream Intersection: Distance to the next intersection that drivers need to turn
- Next Movement: The next turning movement at the intersection, such as turn right, turn left, U turn
- Facility Type
- Number of lanes
- Link average speed
- Link volume
- Speed limit

Most attributes listed above are straightforward and could be obtained by simple data joins. One thing to mention is some special work is needed for the "Next Movement" and "Distance to downstream intersection", which requires the algorithm to scan through the user's whole trip trajectory, extract the intersections where drivers need to make turns, and compute these 2 values for each GPS point.

Following is the sample data after the data joins.

3.3.3. Key variables description

We envision the key variables of measuring driving hazards to include the variables that can be observed from the user GPS trajectory or can be derived after data joins. For example, since driving speed and acceleration/deceleration can be obtained from the second-by-second GPS trajectory, by associating with the road segment speed limit information, it can be determined whether the driver drove at a speed higher than the speed limit. In addition, it can also be inferred how many times the driver has made harsh brake during the trip by looking at the deceleration values.

For each trip, the following key variables can be defined and calculated:

- Speeding: if driver is driving at a speed higher than speed limit of the road segment by a certain threshold

GPS Point ID	User information									Network information					Traffic information		
	UserID	TripID	Latidude	Longtitude	Altitude	Heading	Speed (mph)	Acceleration (ft/s^2)	Time stamp	Metropia LinkID	Downstream Intersection (miles)	Next Movement	Faculity Type	# of lanes	Link Avg Speed (mph)	Link Volume (vehicle/hour)	Speed limit (mph)
1	499	trip_2014(	32.25037003	-110.9282074	2351.0	272.1	34.7	-0.5	4/9/2014 7:58:11 AM GMT-7	9171	1.4	Turn Left	Arterial	2	36.6	940	40
2	499	trip_2014(	32.2503624	-110.9283981	2350.4	269.9	36.4	2.5	4/9/2014 7:58:12 AM GMT-7	9171	1.4	Turn Left	Arterial	2	36.6	940	40
3	499	trip_2014(	32.2503624	-110.9285736	2349.4	269.8	36.9	0.8	4/9/2014 7:58:13 AM GMT-7	9171	1.4	Turn Left	Arterial	2	36.6	940	40
4	499	trip_2014(	32.2503624	-110.9287643	2348.1	269.7	38.0	1.6	4/9/2014 7:58:14 AM GMT-7	9171	1.4	Turn Left	Arterial	2	36.6	940	40
5	499	trip_2014(	32.2503624	-110.9289474	2346.1	269.9	38.6	0.8	4/9/2014 7:58:15 AM GMT-7	9171	1.4	Turn Left	Arterial	2	36.6	940	40
6	499	trip_2014(	32.25037003	-110.9291458	2345.5	269.9	38.6	0.0	4/9/2014 7:58:16 AM GMT-7	4584	1.4	Turn Left	Arterial	2	37.1	890	40

Figure 2. Sample data after data joins

- Relative Speed: if the driver's driving speed deviates from other drivers on the same road at the same time for more than a certain threshold
- Braking: if deceleration is lower than given threshold
- Time in traffic: the time when traveling in traffic and driving slowly
- Acceleration: if acceleration is higher than given threshold
- Peak Time: if the trip happens during peak hours
- Late time: if the trip happens at midnight or early morning
- Left Turn: the number of times driver makes a left turn during the trip
- Right Turn: the number of times driver makes a right turn during the trip
- Mileage: the total distance of the trip
- Travel time: the total length of time traveled during this trip

3.4 Calibration

The calibration module to better calibrate the threshold for driving risk measurements will directly affect the data analysis result later on. In the past, we've relied on literature to define this type of threshold value, which usually works fine for most risk measurements.

However for the purpose of detecting hard brake, such a value may not be applicable for our purpose because of the differing measurement instruments used in literature. A special field experiment was conducted earlier this year specially for this research. We collected a set of field data from a controlled experiment that involved actual hard break behavior by different drivers using various types of vehicles. We performed statistical analysis on all collected data to identify the threshold, which is used to detect hard brake for the research purpose later on.

More details about the hard experiment can be found at *(21)*. A video showing an example of the procedure for this test can be viewed at http://youtu.be/uwr8YaV0kXc

3.5. Analysis methodology

Figure 1 described the systematic framework of how the GPS trajectory data will be collected and processed for data analysis purpose. In this section, the analysis methodology used in this research will be presented. The analysis methodology will be performed at three different levels:

1) Data summary statistics
2) Definition of the key risk measurements and computation
3) Correlation analysis between risk factors

3.5.1 Data summary statistics

The first step of the analysis methodology is the data summary statistics, with the goal of revealing the characteristics of our data samples. This includes mean values, standard deviations, minimums, maximums and the distribution of each variable at the macro level. This allows us to develop a baseline comparison of different risk factors and answer some basic questions such as the percentage of people drive over speed limit, or whether "hard braking" a common habit, etc.

The univariate distribution of each variable will be used to determine the appropriate bucket for each variable. The potential skewness of the variables will also be informative, as it reveals whether the data is in need of transformation in the consequent statistical models. For the data summary statistics, the team will perform the statistical analysis for key variables mentioned in 3.3.2 and 3.3.3 Specially, the focus will be put on the following variables.

- Speed
- Acceleration
- Deceleration
- Hard brakes
- Harsh start

3.5.2. Key risk measurements definition

Based on the key variables defined in Section 3.3.3, various risk measurements can be defined for individual driver at the trip-based level. In addition to the total distance and time traveled, to ensure comparability of data, trajectory variables will be standardized by time, for example number of hard brakes per unit of time, or percentage of time when driver is speeding.

Table 1 summarizes the key risk measurements considered in this research. For the hard brake threshold being used in this research, as discussed in Section 3.4, the team performed a hard brake field experiment to better calibrate this critical parameter. For the other measurements, we're relying on literature in transportation engineering or insurance industry to define threshold values.

Table 1. Key risk measurements definition

User ID	Unique ID of the user
Trip ID	Unique ID of this trip
Speeding 1	Percentage of time when driving speed is higher than speed limit for more than 5mph
Speeding 2	Percentage of time when driving speed is higher than speed limit for more than 10mph
Relative Speed 1	Percentage of time when driving speed deviates from other drivers on the same road for more than 5mph
Relative Speed 2	Percentage of time when driving speed deviates from other drivers on the same road for more than 10mph
Braking 1	# of times deceleration is lower than given threshold (-10ft/s2) per unit time
Braking 2	# of times deceleration is lower than given threshold (-15ft/s2) per unit time
Braking 3	# of times deceleration is lower than given threshold (-20ft/s2) per unit time
Time in Traffic 1	Percent of time when traveling at a speed lower than 85% of the link speed limit
Time in Traffic 2	Percent of time when traveling at a speed lower than 80% of the link speed limit
Time in Traffic 3	Percent of time when traveling at a speed lower than 75% of the link speed limit
Acceleration 1	Based on # of times acceleration is higher than given threshold (10ft/s2) per unit of time
Acceleration 2	Based on # of times acceleration is higher than given threshold (15ft/s2) per unit of time
Acceleration 3	Based on # of times acceleration is higher than given threshold (20ft/s2) per unit of time
Peak Time	Percentage of time when driving during peak hours
Late Time	Percentage of time when driving between 12:00am to 4:00am
Left Turn	# of times user makes a left turn during the trip per unit of time
Right Turn	# of times user makes a right turn during the trip per unit of time
Mileage	Total mileage driven
Travel time	Total length of time traveled during this trip

3.5.3. Correlation analysis

After defining the risk measurement, performing correlation analysis of risk factors is the next step. The goal of this analysis is to identify the risk factors that are contributing to the accidents, and quantify the correlations between those risk factors.

The correlation of the array of risk factors allows us to identify which risk factor is most likely to be an attribute for a crash, or might otherwise be a contributing factor to a severe incident. A comparison of correlation by other factors (e.g. time of day) can be useful information as to the likelihood of a varying marginal effect of different combination of risk factors. Given that no crash data is currently available, hard brake will be used in this research to represent the probability of drivers running into accident, i.e. the more hard brake happens, the probability of a driver run into accident will be higher.

For the correlation analysis, Statistical Analysis System (SAS) software will be used to reveal the correlations between the risk measurements defined in Section 3.5.2s.

4. CASE STUDY

In this section, we present a preliminary case study to demonstrate the feasibility and preliminary result of this research using existing trajectory data. The next two sections present the summary statistics and correlation analysis on the data we collected.

For a proof of concept purpose, we used partial data from the Tucson network. After confirming the validity of the trips, we conducted our analysis on a total of 171 trips.

4.1. Data summary statistics

4.1.1. Speed, Deceleration and Acceleration Distributions

The distribution of speed, deceleration and acceleration is analyzed in this section.

The average speed of all trips is 25 mph, with the minimum speed equals to 5.1 mph and maximum speed equals to 59.75 mph. Given that all trips are taken in Tucson, the range of speed is expected.

Acceleration and deceleration measures reflect the changes in speed in terms of feet per square second (ft/s^2). Both deceleration and acceleration show skewness in the distributions. The average deceleration is 4.6 ft/s^2. The minimum and maximum deceleration are 2.22 ft/s^2 and 19.99 ft/s^2, respectively. The histogram shows that most drivers hit the brake at the rate of 4 to 5 ft/s^2 and very few had deceleration greater than 10 ft/s^2.

Acceleration distribution shows a similar pattern. The average acceleration is 4.4 ft/s^2. The minimum and maximum acceleration are 2.46 ft/s^2 and 27.77 ft/s^2, respectively. Most acceleration ranges from 3 to 5 ft/s^2, with a few acceleration values greater than 11 ft/s^2. Compared to deceleration, the acceleration distribution shows a slightly greater variation.

The number of trips and the corresponding percentage of observations in each speed range are summarized below.

- Average speed: out of 171 trips, 10 have average speed below 10 mph, which counts for 5.8% of our sample. The most comment speed range is 25 to 29.99

mph. There are 35 trips in this speed range, which counts for 20.5 percent of the sample.

- Average deceleration: 152 out of 171 trips have deceleration in the range of 3 to 5.99 ft/s^2, which counts for 88.9% of the trips. There was only one trip where the deceleration is greater than 12 ft/s^2 and the deceleration is in the range of 18 to 21 ft/s^2.
- Average acceleration: average acceleration is less than 6 ft/s^2. Out of 171 trips, 161 trips have acceleration in the range of 3 to 5.99 ft/s^2, which counts for 94.2% of the total trips. Only three trips have acceleration greater than 9 ft/s^2. Among those three trips, one is in the range of 21 to 23.99 ft/s^2 and one in the 27 to 30 ft/s^2 range.

4.12. Hard Brake and Hard Acceleration

To identify the tendency of having hard brakes or hard acceleration, for each trip we count number of deceleration exceeding certain thresholds and standardize the number by the length of the trip. When hard brake is measured by a greater than 10 ft/s^2 deceleration, the average number of hard brake per trip is 0.17 per minute. The maximum number of observed hard brake is 1.33 per minute. When we use a more restrictive threshold, the number of detected hard brake drops. For example, when we define hard brake as greater than 15 ft/s^2 deceleration, the average number of hard brake becomes 0.06 per minute. When we use 20 ft/s^2 as the threshold which is the median value from our hard brake experiment, the average number of hard brake further drops to 0.03 per minute.

For hard acceleration, the average numbers of hard acceleration per minute are 0.20, 0.07 and 0.03 for a threshold of 10 ft/s^2, 15 ft/s^2 and 20 ft/s^2, respectively. The maximum number of hard acceleration is 1.5 per minute when we use the threshold 10 ft/s^2. The count drops to 0.83 per minute when a 20 ft/s^2 threshold is adopted.

Figure 3 show the number and the percentage of trips in each frequency range. Under this definition, most trips have less than 0.1 hard brake per minute. 169 out of 171 trips have less than 0.3 hard brake per minute, which counts for 98.8% of the observations. There are two trips with more than 0.5 hard brakes per minute, which counts for 1.2% of the observations. We can also observe that the number of qualifying hard brakes gets lower as the threshold increases, as expected. For example, when 10 ft/s^2 is used to define a hard brake, 16 trips have greater than 0.6 hard brake per minute. When 15 ft/s^2 is used as a threshold, only 2 trips have more than 0.6 hard brake per minute, which is significantly lower than when 10 ft/s^2 was used as a threshold.

The same pattern is also observed for hard acceleration. Using 20 ft/s^2 as an example, almost 90 percent of the trips have less than 0.1 harsh acceleration per minute. It's also found that the number and percentage of trips in each frequency range, 168 trips have less than 0.3 hard acceleration per minute and only three trips have more than 0.4 hard acceleration per minute.

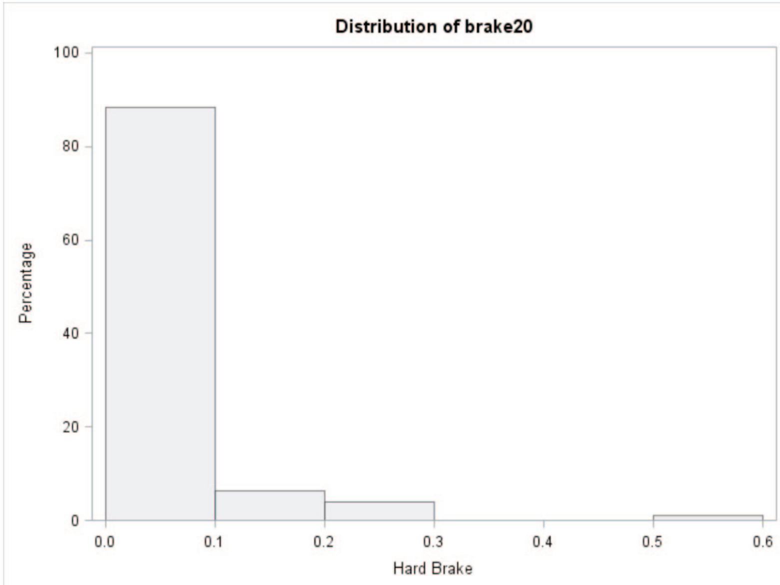

Figure 3. Histogram of Number of Hard Brake (20 ft/s^2) per Minute

4.1.3. Hard Brake and Speed

We divide our sample into two groups based on number of hard brakes: those in the upper 50% (trips with more hard brakes) and those in the lower 50% (trips with fewer hard brakes) so we have two groups with close to equal sizes. Table 2Table 2 shows the comparison of the acceleration pattern, average speed and relative speed to other cars between these two groups. The group with more hard brakes tend to drive faster, have more hard acceleration and driving speed is more likely to deviate from other drivers. Specifically, comparing to those with no or few hard brakes, those with more hard brakes drive 4.3 mph faster, have 0.122 more hard acceleration per minute and spend 0.79% more time with driving speed deviating from other drivers on the same road for more than 5mph.

Table 2. Comparison of Various Speed Related Variables by Hard Brake

	More Hard Brakes	Fewer Hard Brakes
Number of Trips	86	85
Number of times when acceleration is greater than 10ft/s^2 per minute	0.256	0.134
Average speed measured by mph	27.152	22.813
Percentage of time when driving speed deviates from other drivers on the same road for more than 5mph	0.609	0.531

4.2. Correlation analysis

Table 3 shows the correlation between hard brake and other driving attributes, including hard acceleration, relative speed to other drivers on the same road, speeding, number of left turn and number of right turn. Pearson correlation coefficients are reported for each pair of variables. Given the small sample we have in the trial sample, we use 10 ft/s^2 as the threshold for defining hard brake and acceleration in this section of the research. We find that hard brake is positively correlated with hard acceleration, speed variation with other drivers and number of left turns, even though the correlation between hard brake and number of left turn is only marginally significant. The results imply that divers with hard brakes more frequently are also more likely to have hard acceleration, travel at a speed that is different from other cars on the same road and possibly make more left turns.

Table 3. Correlation between Hard Brake and Other Driving Attributes

	Definition	Correlation
Hard Acceleration	Number of times when acceleration is greater than 10ft/s^2 per minute	0.459***
Relative Speed	Percentage of time when driving speed deviates from other drivers on the same road for more than 5mph	0.228***
Speeding	Percentage of time when driving speed is higher than speed limit for more than 10mph	-0.061
Left Turn	# of times user makes left turn during the trip per minute	0.139*
Right Turn	# of times user makes right turn during the trip per minute	0.081

*** statistically significant at 1%
** statistically significant at 5%
* statistically significant at 10%

5. CONCLUSIONS AND FUTURE RESEARCH

In this research, we use a smartphone app to collect the data for the purpose of studying driving risk factors. The uniqueness of the data in this research is its backend server also estimates traffic speed and volume for each link that the vehicle traverses. The overall research framework used in this research is presented, which mainly includes data collection, data processing, calibration and analysis methodology. A preliminary case study including data summary statistics and correlation analysis is also presented. The case study reveals that the group with more hard brakes tends to drive faster, have more hard acceleration and driving speed is more likely to deviate from other drivers, all of which indicate aggressive driving behavior and high driving risks.

Future research directions include collecting crash data and claim amount through partnership extension with insurance companies, and build actuarial analysis model to quantify the relationships between hazards and accident claims. We believe that as time goes, more personalized driving data will become available and used in the statistical model for analysis, the research team can further move this research forward, so that in the end, the results of our study will further existing knowledge about driving exposure factors that are closely linked to crash risk. The manner in which these factors affect loss frequency and claim severity will also help insurers in the design and implementation of Usage Based Insurance program and provide the actuarial foundation for advanced forms of Pay-As-You-Drive-And-You-Save (PAYDAYS) insurance pricing.

6. ACKNOWLEDGEMENT

This research is supported by Federal Highway Administration Broad Agency Announcement "Pay-As-You-Drive-And-You-Save (PAYDAYS) Insurance Actuarial Study" project. Contract award number DTFH61-13-C-00033.

7. REFERENCES

1. RAINIE, L. *Cell phone ownership hits 91% of adults*. 2014; Available from: http://www.pewresearch.org/fact-tank/2013/06/06/cell-phone-ownership-hits-91-of-adults/.
2. Klauer, S.G., T.A. Dingus, V.L. Neale, et al., *Comparing Real-World Behaviors of Drivers With High versus Low Rates of Crashes and Near-Crashes*, 2009, Virginia Tech Transportation Institute.
3. Af Wåhlberg, A.E., *Driver celeration behaviour and accidents – an analysis*. Theoretical Issues in Ergonomics Science, 2008. **9**(5): p. 383-403.
4. Ellison, A.B., S.P. Greaves, and R. Daniels, *Profiling drivers' risky behaviour towards all road users*, in *A safe system: expanding the reach: Australasian College of Road Safety national conference*2012: Sydney.
5. Metropia. *Metropia website (http://www.metropia.com)*. 2014.
6. Ulleberg, P., *Personality subtypes of young drivers. Relationship to risk-taking preferences, accident involvement, and response to a traffic safety campaign*. Transportation Research Part F: Traffic Psychology and Behaviour, 2001. **4**(4): p. 279-297.
7. Preusser, D.F., A.F. Williams, S.A. Ferguson, et al., *Fatal crash risk for older drivers at intersections*. Accident Analysis & Prevention, 1998. **30**(2): p. 151-159.
8. Evans, L., *The effectiveness of safety belts in preventing fatalities*. Accident Analysis & Prevention, 1986. **18**(3): p. 229-241.
9. Mannering, F.L. and L.L. Grodsky, *Statistical analysis of motorcyclists' perceived accident risk*. Accident Analysis & Prevention, 1995. **27**(1): p. 21-31.
10. Kim, K., L. Nitz, J. Richardson, and L. Li, *Personal and behavioral predictors of automobile crash and injury severity*. Accident Analysis & Prevention, 1995. **27**(4): p. 469-481.
11. Young, M.S., S.A. Birrell, and N.A. Stanton, *Safe driving in a green world: A review of driver performance benchmarks and technologies to support 'smart' driving*. Applied Ergonomics, 2011. **42**(4): p. 533-539.
12. Ericsson, E., *Driving pattern in urban areas-descriptive analysis and initial prediction model*, 2000, Lund Institute of Technology, Department of Technology and Society, Traffic Planning: Lund.
13. Ericsson, E., *Independent driving pattern factors and their influence on fuel-use and exhaust emission factors*. Transportation Research Part D: Transport and Environment, 2001. **6**(5): p. 325-345.
14. Golob, T.F., W.W. Recker, and V.M. Alvarez, *Freeway safety as a function of traffic flow*. Accident Analysis & Prevention, 2004. **36**(6): p. 933-946.
15. Shefer, D. and P. Rietveld, *Congestion and Safety on Highways: Towards an Analytical Model*. Urban Studies, 1997. **34**(4): p. 679-692.
16. USDOT, *Study of the Impact of a Telematics System on Safe and Fuel-efficient Driving in Trucks*, 2014.

17. Noland, R.B. and M.A. Quddus, *A spatially disaggregate analysis of road casualties in England*. Accident Analysis & Prevention, 2004. **36**(6): p. 973-984.
18. Wang, C., M.A. Quddus, and S.G. Ison, *Impact of traffic congestion on road accidents: A spatial analysis of the M25 motorway in England*. Accident Analysis & Prevention, 2009. **41**(4): p. 798-808.
19. Toledo, T., O. Musicant, and T. Lotan, *In-vehicle data recorders for monitoring and feedback on drivers*. Transportation Research Part C: Emerging Technologies 2008. **16**(3): p. 320-331.
20. Toledo, T. and T. Lotan, *In-Vehicle Data Recorder for Evaluation of Driving Behavior and Safety*. Transportation Research Record: Journal of the Transportation Research Board 2006. **1953**: p. 112-119.
21. Chiu, Y.-C., Y.-L. Ma, and X. Hu. *Identifying Driving Risk Factors to Support Usage-Based Insurance using Smartphone Personalized Driving Data*. in *Workshops on Big Data and Urban Informatics*. 2014. Chicago, IL.