Vehicle Occupants and Driver Behavior: A Novel Data Approach to Assessing Speeding

November 2019 | Final Report

SAFE-D
SAFETY THROUGH DISRUPTION

TRANSPORTATION INSTITUTE
VIRGINIA TECH.

Texas A&M Transportation Institute

San Diego State University
Leadership Starts Here

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. 02-009	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Vehicle Occupants and Driver Behavior: A Novel Data Approach to Assessing Speeding		5. Report Date November 2019 6. Performing Organization Code:
7. Author(s) Michael W. Martin Dr. Lisa L. Green Dr. Eva Shipp Byron Chigoy Rahul Mars		8. Performing Organization Report No. Report 02-009
9. Performing Organization Name and Address: Safe-D National UTC Texas A&M Transportation Institute The Texas A&M University System College Station, Texas 77843-3135 Virginia Tech Transportation Institute 3500 Transportation Research Plaza Blacksburg, VA 24061		10. Work Unit No. 11. Contract or Grant No. 69A3551747115/Project 02-009
12. Sponsoring Agency Name and Address Office of the Secretary of Transportation (OST) U.S. Department of Transportation (US DOT) State of Texas		13. Type of Report and Period Final Research Report 14. Sponsoring Agency Code

15. Supplementary Notes
This project was funded by the Safety through Disruption (Safe-D) National University Transportation Center, a grant from the U.S. Department of Transportation – Office of the Assistant Secretary for Research and Technology, University Transportation Centers Program, and, in part, with general revenue funds from the State of Texas.

16. Abstract
The question of whether driver behavior, and speeding in particular, differs based on passenger(s) presence requires the use of large amounts of data, some of which may be difficult to accurately obtain. Traditional methods of obtaining driver behavior information result in datasets that either lack passenger information altogether (i.e., insurance companies using telematics) or rely on rough estimates of passenger age and gender obtained from blurred photos (i.e., naturalistic driving studies like the Second Strategic Highway Research Program). This research project represents a novel, data-driven approach to assessing passenger impact on speeding. Household travel survey demographic information and GPS traces were linked to HERE network speed limit to study the impact of vehicle occupancy on speeding. Survey responses from 11 study areas were cleaned, merged, and ultimately used in developing binomial logistic regression models. Of particular interest were the following driver groups: teenagers, adults driving with child passenger(s), and older drivers. The models suggest that drivers speed less when there is a passenger in the vehicle, particularly adult drivers with a child passenger(s).

17. Key Words passengers, big data analytics, vehicle occupancy, speeding, binomial logistic regression models, household travel survey, GPS data	18. Distribution Statement No restrictions. This document is available to the public through the Safe-D National UTC website, as well as the following repositories: VTechWorks, The National Transportation Library, The Transportation Library, Volpe National Transportation Systems Center, Federal Highway Administration Research Library, and the National Technical Reports Library.		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 22	22. Price $0

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Abstract

The question of whether driver behavior, and speeding in particular, differs based on passenger(s) presence requires the use of large amounts of data, some of which may be difficult to accurately obtain. Traditional methods of obtaining driver behavior information result in datasets that either lack passenger information altogether (i.e., insurance companies using telematics) or rely on rough estimates of passenger age and gender obtained from blurred photos (i.e., naturalistic driving studies like the Second Strategic Highway Research Program). This research project represents a novel, data-driven approach to assessing passenger impact on speeding. Household travel survey demographic information and GPS traces were linked to HERE network speed limit to study the impact of vehicle occupancy on speeding. Survey responses from 11 study areas were cleaned, merged, and ultimately used in developing binomial logistic regression models. Of particular interest were the following driver groups: teenagers, adults driving with child passenger(s), and older drivers. The models suggest that drivers speed less when there is a passenger in the vehicle, particularly adult drivers with a child passenger(s).

Acknowledgements

This project was funded by the Safety through Disruption (Safe-D) National University Transportation Center, a grant from the U.S. Department of Transportation – Office of the Assistant Secretary for Research and Technology, University Transportation Centers Program, and, in part, with general revenue funds from the State of Texas.

We would like to thank the TxDOT Transportation Planning Program for allowing the use of the household travel survey data and acknowledge the work of Gargi Singh in developing the learning module example. We appreciate Robert Wunderlich taking the time to review this report and provide valuable feedback as the project's subject matter expert.

Table of Contents

List of Figures

List of Tables

Introduction

Obeying the speed limit is one of a driver's more important responsibilities, with close ties to safety. Speeding may also result in speeding tickets along with associated fines and penalties. And while speeding itself is not lethal, it may be causal factor in crash severity, with some crashes ultimately being fatal. In 2016, Texas had the highest number of speeding related fatalities (1,069) of any state in the U.S. [1]. This indicates the need to better understand speeding and the factors that influence it. Factors that can influence speeding include "public attitudes, personal behavior, vehicle performance, roadway characteristics, enforcement strategies and speed zoning" [1]. Another factor that may influence speeding is vehicle occupancy—both with respect to the number and characteristics of the occupant(s). This study explores whether and how the presence, number, and ages of vehicle passengers influence driver speed choice. Based on the models developed as part of this research, drivers speed less when traveling with a passenger in the vehicle, especially in the case of an adult driver traveling with a child passenger(s). These model results serve as a starting point for using the novel technique of combining travel survey data with GPS and HERE network data to gain insights into speeding behavior linked to passengers and various driver ages. Further modeling with the dataset derived as part of this research may lead to additional insights into the impact of passengers on speeding and help in the development of safety countermeasures targeted at those groups of drivers that may be especially prone to speeding with passengers.

Analyzing the question of whether driver behavior, and speeding in particular, differs based on vehicle occupancy requires the use of large amounts of data, some of which may be difficult to accurately obtain. In determining passenger information, the Second Strategic Highway Research Program dataset reduction process relied on a blurred photo that only allowed for a rough *estimate* of passenger demographic information, such as age and gender [2]. Data collected by insurance companies that use telematic devices for vehicle tracking to assess discounts [3] capture details about drivers and their habits; however, their datasets lack information on passengers in the vehicle for a given trip. Lack of reliable passenger data limits the ability to confidently assess the impacts of passengers on speeding.

The proposed research addresses this gap using household travel survey data from the Texas Department of Transportation Travel Survey Program (TxDOT TSP) dataset, which has traditionally been used for transportation planning purposes, but was applied to the area of safety for this project. This type of novel approach in a safety setting adds value to this rich household travel survey dataset and makes greater use of data that can be expensive to collect. The objective of this research was to better understand the relationship between vehicle occupancy and driver speeding behavior. In particular, the focus was on the following driver groups of interest: teenagers, adults driving with child passenger(s), and older drivers. The dataset allowed for assessment of both passenger and driver characteristics in order to evaluate the impacts of passengers on speeding. This was accomplished by merging traditional household travel survey responses, GPS data linked to trips, and speed limit roadway network data. This approach removes the guesswork from linking passenger demographic information to trip speed information. The resulting dataset was used in the development of binomial logistic models, which were used in assessing the impact of vehicle occupancy on speeding.

Report Outline

This report is presented as follows. First, the Introduction presents the objective of this research and gives an overview of the report structure. This is followed by the Background, which provides information on research germane to vehicle occupancy and speeding. Specific focus is placed on the following driver groups: teenagers, adult drivers traveling with a passenger(s), and older drivers. The Methodology section includes a description of the data used in the analysis and the cleaning process for making the data usable in addressing the research objective. A description of how models were developed to determine if vehicle occupancy is related to speeding tendencies of driver groups is also included. The Results section summarizes the model results and explains how they can be interpreted. This is followed by a Discussion of some of the statistical modeling and results and limitations associated with this research project. Lastly, the Conclusions and Recommendations section includes a discussion of further research that could build on the results of this research project and might benefit from utilization of the dataset that resulted from extensive cleaning and coding.

Background

Effects of Passengers

There are a wide array of findings within the literature concerning the effect of passenger presence on driver behavior. Some research reports that driving with passengers has a positive effect on minimizing speeding and/or increasing safety. Fleiter et al. [4] note that drivers consider safety and comfort when driving with passengers and subsequently speed less. This may also be a reflection of drivers' desires to appear concerned about passenger welfare [4]. Lee and Abdel-Aty found that drivers are generally more likely to practice safe driving habits, like wearing a seatbelt and not driving while impaired by alcohol, when traveling with passengers [5]. Vollrath et al. found that passengers decreased accident risk most for older drivers (age 50+) and that this positive effect of passenger presence decreased with driver age [6]. Preusser et al. found that passenger presence decreased the proportion of at-fault crashes for drivers aged 30-years and older [7].

While several researchers found passenger presence to have a positive effect on drivers, others found the opposite. Preusser et al. found that in fatal crashes, in contrast to their previously cited finding regarding reduced passenger-related at-fault crash risks for drivers 30-years-old and older, passenger presence increased the proportion of at-fault fatal crashes in drivers aged 24-years and younger [7]. Similarly, Doherty et al. found that, for drivers aged 16–19, the rate of accident involvement was roughly twice as high when traveling with passengers than when traveling alone [8]. Aldridge et al. found that young drivers (aged 16–20) driving with peer passengers resulted in an increased propensity to cause accidents. However, when young drivers drove with adults and/or children their propensity to cause accidents was lowered [9].

Several authors cite the *number* of passengers being an important factor in safety. Fu and Wilmot found that fatal crash risk increased for young drivers as the number of passengers increased, particularly for drivers aged 16–17 [10]. Similarly, Lam et al. found that for younger drivers, traveling with multiple passengers increased car crash injury risk, though the same was not true for older drivers [11]. Rosenbloom and Perlman found that more passengers resulted in fewer traffic violations [12]. Lee and Abdel-Aty also found this to be true, positing that traveling with more passengers reduced driver crash potential as drivers felt a greater sense of responsibility when traveling with more passengers [5]. Keall et al. found that after controlling for blood alcohol content and age, traveling with one passenger was ideal in reducing the risk of a fatal injury to the driver during night-time travel. In fact, relative to driving alone or with two or more passengers, driving with just one passenger actually halved the risk of a fatal injury to the driver traveling at night [13]. Doherty et al. found that accident rates were higher with two or more passengers than with only one passenger [8]. Chen et al. found that the death risk of 16–17-year-old drivers increased with number of passengers, and that death rates in crashes for this young group of drivers was especially high when carrying male passengers or passengers younger than 30 [14]. As this review of the literature reveals, the findings related to the impact of the number of passengers is somewhat split regarding whether more passengers are a positive or negative influence from a safety perspective. Additionally, in some cases, the findings are not linear with respect to number of passengers. Likewise, the age make-up and gender of the driver and/or passengers may also influence how the presence of passengers impacts safety.

Driver Groups of Interest

The focus of this research project is on the following driver groups of interest: teenagers, adults driving with child passenger(s) (originally phrased as "parents;" further explanation for the change in terminology is provided later in the report), and older drivers. This section contains background information specific to each of these groups that supports the further study of how vehicle occupants affect the speeding behavior of each group of drivers.

Teenagers

Teenagers were identified as a driver group of interest for this study given their relative newness to driving and the associated safety concerns. Several of the articles referenced in the literature review specifically focused on teenagers driving with passengers [7-11, 14]. The *California Driver Handbook* states that, "If you are under 18 years old, your risk of a fatal collision is about 2½ times that of the 'average' driver. Your risk of an injury collision is 3 times higher than the average driver's risk " [15, p. 14]. Given teenage drivers' statistically increased risk of both fatal and injury collisions, special consideration of passenger effects on these drivers is warranted.

Graduated Driver Licensing (GDL) programs are one measure to help address safety concerns of teenagers driving with passengers. GDL programs are designed to help safely transition teenagers to driving, and while the specifics of GDL programs may vary by country or state, one of the main components of GDL programs includes limiting the allowable number and/or type of passengers [16]. Based on a meta-anlaysis of GDL programs, Shope found that GDL programs reduce crash risk in young drivers by 20–40% [17].

Parents

Parents have a responsibility to keep the children under their supervision safe, and thus were identified as a driver group of interest for this research. Much of the literature on parent drivers traveling with children revolves around parental impacts on teenage drivers. However, a handful of articles mention the effects of parents driving with children in their vehicle. Koppel et al. found that only 12% of driver distractions related to child occupants. They also found that males/fathers were more likely to be distracted by child occupants both in terms of frequency of distractions and duration of distractions [18]. Dingus et al. found that despite children being a distraction, interacting with children in the backseat actually has a "protective effect." This may be a reflection of parents traveling with children driving slower or with greater headway to other vehicles [19]. Given these mixed findings that children can be a source of distraction but may also provide extra motivation for safe driving, parents (i.e., adults driving with child passenger(s)) were a driver group of interest within this research project.

Older Drivers

Older drivers were identified as a driver group of interest given their susceptibility to injuries in crashes. Based on a review of the literature, only a few studies have specifically examined the effect of passengers on older driver behavior and how the presence of passengers can influence safety. Hing et al. found that older drivers (age 75+) were less likely to be at fault in a crash when traveling with two or more passengers during night travel [20]. Braver and Trempel found that passengers of older drivers (age 75+) who were involved in a fatal crash had a higher death rate than passengers traveling with drivers aged 30–59 who were involved in a fatal accident. However, this may partially be a reflection of older drivers' passengers tendency to be older and more susceptible to injury [21]. Most of the reviewed literature focuses on the effect of passengers on young drivers. Our study strives to reduce the gap in older driver/passenger risk information, as the household survey data used to develop the models encompass a large range of driver age groups, including older drivers.

Methodology

TxDOT Household Travel Survey Data

Travel surveys are typically conducted to support the development of a regional travel demand model. Generalized travel behavior characteristics are produced at the household and person levels to estimate travel within a specific region. Since 2000, the TxDOT TSP has conducted household surveys around the state for each of its 25 Metropolitan Planning Organizations, as shown in Figure 1.

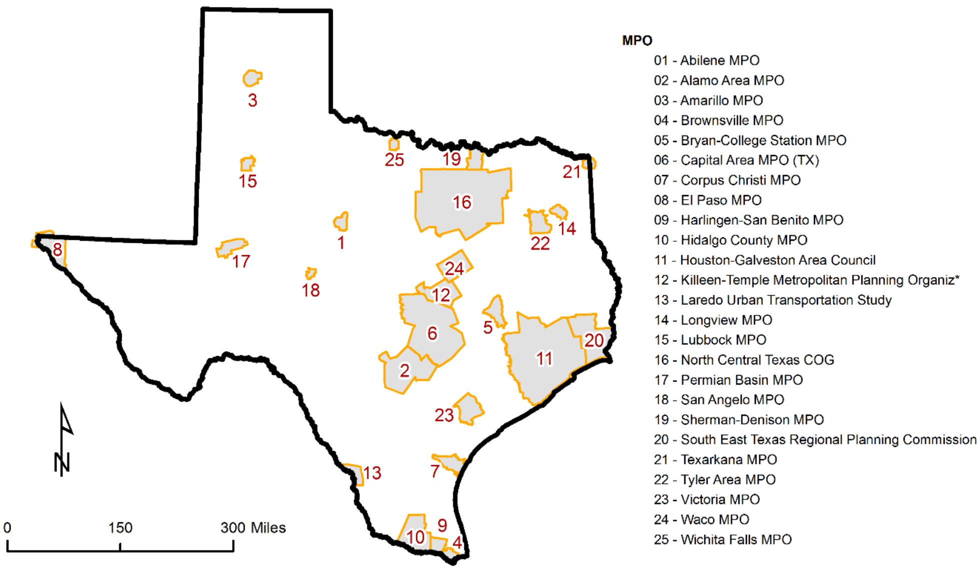

Figure 1. Texas metropolitan planning organizations.

These surveys collect information about a household, the persons that make up each household, vehicle details, and trips made by members of the household. Figure 2 illustrates that each of these data elements can be related using a unique identifier for households, persons, vehicles, and trips.

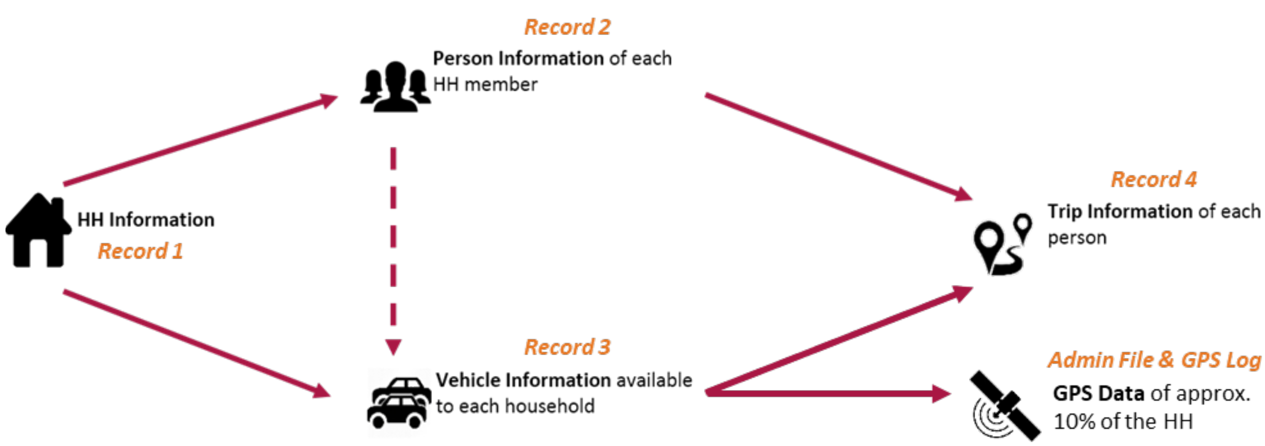

Figure 2. Travel survey data elements.

Trip information is captured via a travel diary that respondents use to self-report details on each trip that they make during a specified travel day. The trip records are dependent on respondents accurately recording their details in daily trip diary, and/or by how well they respond to queries by trained surveyors collecting this information. GPS data for household surveys were collected for one to three vehicles in a household for a 10% sub-sample of households surveyed. The data from the vehicles with a GPS unit were tied back to the data about the vehicle's passengers using the trip records from the survey. The GPS units were installed in respondents' vehicles and used logged vehicle travel information in 1-second intervals while the vehicle was running and in motion.

GPS data for travel surveys is collected as a quality control and error correction measure to supplement the surveys. A well-known travel data collection problem is that respondents often do not enter information about their travel day correctly, do so carelessly, or fail to record small trips. The intent of the GPS data is to identify these gaps.

Typically, great effort has to be taken to tie a vehicle's entire travel day GPS traces back to the person(s) who used the vehicle. To do this, practitioners have developed a number of data cleaning algorithms that correct for information that is not found in the trip records but is found in the GPS records. One example of a typical error is that a respondent records starting a trip at a certain time but the GPS data suggests the trip started earlier. Based on the GPS data, the respondent's start time can be corrected.

Another common problem is that two respondents may record driving the same vehicle at the same time. For this research, data cleaning focused on identifying only those trips where researchers had a high confidence that the trip and GPS records were in close alignment. As such, this focus shifts analysis of household GPS data from understanding travel for a day and localizes it to individual trips.

Traditional uses of the TxDOT TSP household GPS data may disregard errors in the GPS data as impertinent to the research. These errors include positional shifting that occurs when the vehicle is stopped or slow moving, gaps in the GPS records when the vehicle passes under bridges and other overhead obstacles, and the canyon effect, which represents a large positional shift between GPS records due to buildings and other reflective objects along the vehicle's path. However, for this research project, the team conducted a thorough review of the GPS records to identify these problems and address them using corrective methods.

GPS Data Processing and Travel Survey Linking

Linking GPS trace data to household travel surveys is a multi-step procedure. Household surveys include four survey files:

- **Household File:** Information about a household, such as income and size.
- **Person File**: Information about the persons in the household, such as age and gender.
- **Vehicle File**: Information about the household vehicles, such as make and model.
- **Trip File:** Information about the travel of each person in the household for a day, including their mode of travel and geographic data for each travel location.

GPS data for household travel surveys include two primary datasets: (1) an administration file, hereafter referred to as the "admin file," and (2) the GPS logger file, hereafter referred to as the "log" file. The log file contains vehicles' GPS trace data. The admin file includes crosswalk information that links the log file to its respective surveyed household and household vehicle. Both the admin file and log file are singular and contain all GPS data for all households whose surveys include a GPS component. These files have multiple entries stemming from the same GPS logger, though for different households, vehicles, and days.

Both the admin file and the log file represent semi-raw data. That is, the data collector provides only the minimum amount of information necessary to link the two files. It is incumbent on analysts to process these files into a suitable form for analysis. Generally, this includes two processes. The first is data linking, which establishes key variables that link the admin file and log file to the appropriate household using data from the household survey and "standardizing" variables in the survey file to match those in the admin file and log file. The second process is GPS processing to attribute and segment GPS trace data into trips. Trips are defined here as the one-way travel between a definite origin and destination. For example, one trip is travel from home to work with no stop in between. If the trip is taken in a vehicle, this would be present in the GPS trace data, as shown in Figure 3.

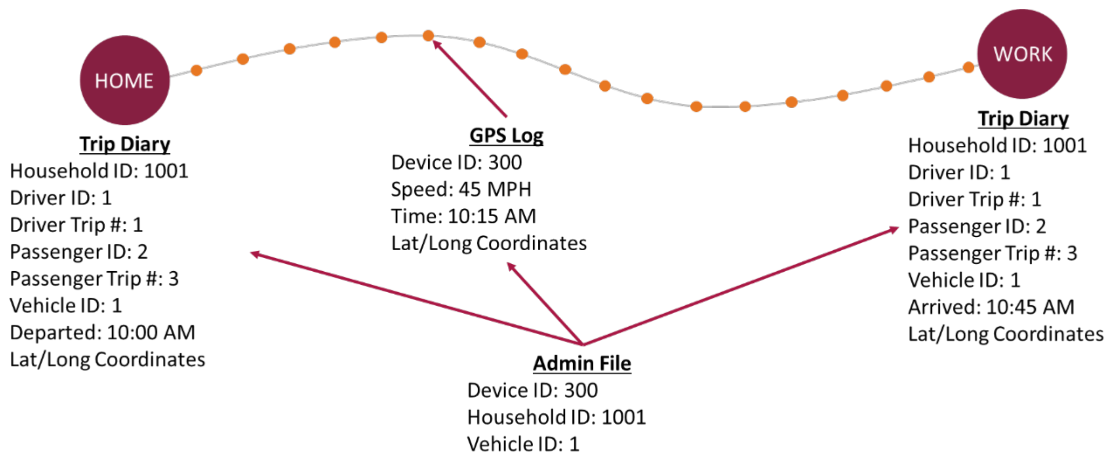

Figure 3. Trip diary and GPS log linkage example.

Data File Linking

Key Variables

The admin file and the survey file link together via unique household and vehicle identifiers. The log file initially only links to the admin file via a GPS logger unit ID. However, GPS technicians may utilize the same GPS unit in several vehicles over the survey period. As a result, linking the GPS admin file and log file to the appropriate household and vehicle requires a match on unit ID *and* the survey date. Establishing these keys provides for linking the GPS admin and log files to the household survey household and vehicle files.

Survey/GPS Standardization

Each household survey has varying data definitions. Accordingly, standardization procedures apply common variables to all surveys representing income, travel mode, trip purpose, and time stamps. Time stamps in both the admin file and log file were also standardized into consistent formats representing local rather than Greenwich Mean Time (GMT).

GPS Data Processing

GPS Attribution

GPS data in the log file have sparse attributes, which include unit, date, GMT, coordinate data (latitude/longitude), velocity, heading, and elevation. For this study, additional attributes were computed from the GPS data and are included in Table 1.

Table 1. Supplemental GPS Attributes

Name	Definition
Log Order	Integer order of GPS points by unit ID, household, and vehicle
Filter	Filter variable to disregard points not on study day or for excluded households
Point to Point Time	Time in seconds between each GPS point
Point to Point Distance	Distance in feet between each GPS point
Point to next 5th Point Heading	Heading from each point to the next 5th point
Point to Point Speed	Speed in miles per hour between each point
Cumulative Time	Total time elapsed for survey day at vehicle start up to log point
Cumulative Distance	Total distance traveled for survey day at vehicle start up to log point

GPS Trip Segmentation

GPS trip segmentation utilizes the attributed data and a heuristic algorithm presented in Figure 4.

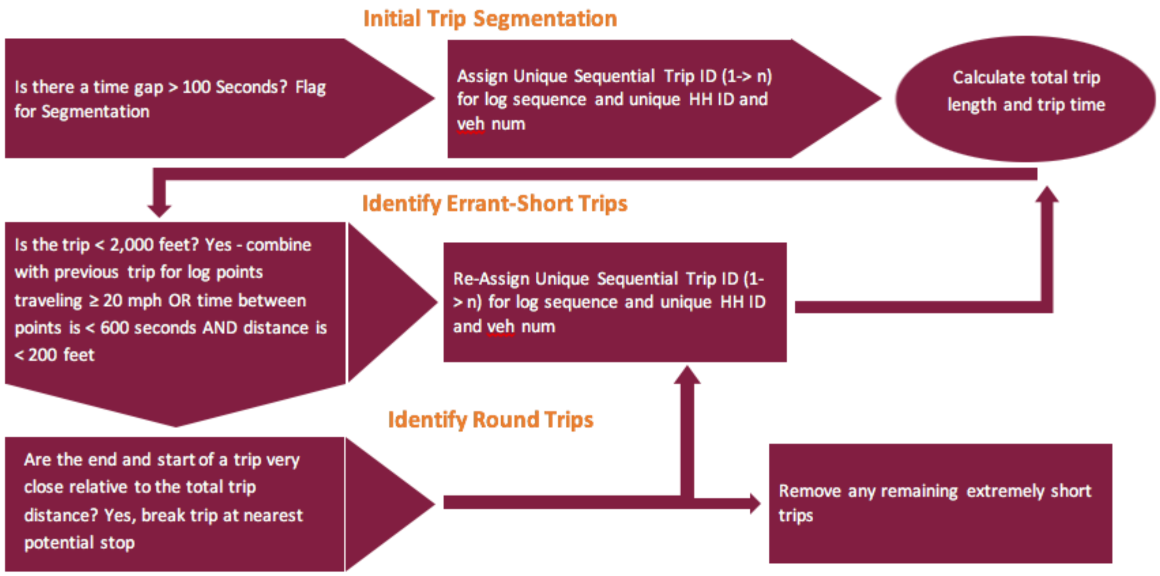

Figure 4. GPS trip segmentation algorithm.

The GPS loggers used for most surveys are designed to shut down when the vehicle is turned off. Ideally, this would create a clear break point indicating separate trips. However, GPS log data are subject to a number of effects which deviate from the ideal. These include user interference, long stops, gaps in log data from structural interference (i.e., no satellite reception beneath bridges) and other anomalies. Initial trip breaks occur when the time between points is greater than 100 seconds.

After initial trip segmentation, each trip is tested to identify very short trips. Most often, these trips occur because the GPS logger turned off at a stop light or the vehicle was in heavy congestion. These trips are recombined with the preceding trip and the test is run again. Additionally, the segmentation procedure identifies round trips, which are trips with start and end points at or near each other. These often occur where the trip is a drop-off and the minimum stopping time is unmet.

GPS Trips to Travel Diary Trips Matching

There are no explicit or direct links connecting household survey trips to GPS log trips. Therefore, matching GPS trips to survey trips relies on indirect methods, which include geospatial and temporal matches. Matching trips using this method does not always ensure that all trips from either the GPS data or trip diary have a match. This is largely caused by households not reporting all trips in their travel diaries and is the primary reason why household travel surveys collect GPS data to offset the frequent underreporting of trips. Underreporting occurs when the person surveyed forgets to report a trip or does not appreciate that a trip they have made needs to be recorded. This phenomenon is usually observed for trips not directly connected to the home, such as travel from one store to another, passenger drop-off trips, and short stops at a gas station. In addition to underreporting, survey diaries rely on a person’s recall of trip information, which leads to misreporting errors such as wrong times, wrong location address, etc. Finally, trip data may have data entry errors attributable to data collection.

Matching trips between the travel diary and GPS loggers means identifying, in the log file, the person from the household making the trip along with their associated trip number. Matching is also to associate the GPS trip ID for a vehicle in the travel diary.

The first step of the matching procedure is to convert the GPS log file to a line file representative of the sequential order of the GPS points. The procedure utilizes geographic data of the trip’s beginning- and endpoints from the trip diary. These data are used to create a boundary plane of the diary trip. Figure 5 illustrates both geographic data components.

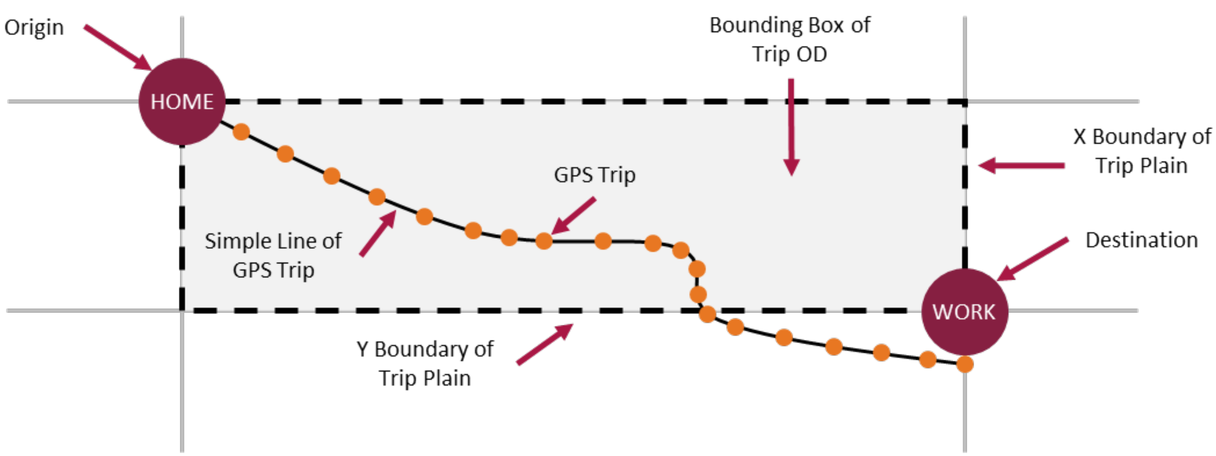

Figure 5. Trip matching geometries.

In addition to the geographic component, the process utilizes the start and end time of the trip as recorded in the travel diary and the start and end time of the trip based on the GPS data. The algorithm matches GPS log trips to travel diary trips sharing the same household and vehicle IDs. Matches are further constrained based on whether the GPS trip line crossed the trip's Y and/or X boundaries, and whether the trips are related in time. The boundary match is preferable over other geographic matches, such as an intersection with a bounding box or GPS point proximity to the origin and destination points. Using a bounding box does not ensure that the trip's path was within the probable path between the origin and destination. The same is true for a proximity test. Each type of match is assigned a hierarchical grade from 1–6, with 1 being a probable best match. Table 2 illustrates a simple set of matched and unmatched trips.

Table 2. Trip Matching Example

Trip File	Trip ID	HH ID	Veh Num	Trip Begin	Trip End	Line Crosses Origin Y	Line Crosses Dest Y	Line Crosses Origin X	Line Crosses Dest X	Line Inter-sects Boun-ding Box	Match Grade
Diary Trip	4	1	2	10:50	11:35	-	-	-	-	-	-
GPS	3	1	2	10:45	11:40	Yes	Yes	Yes	Yes	Yes	1
GPS	3	1	2	10:55	11:30	Yes	Yes	No	No	Yes	1
GPS	3	1	2	10:55	11:40	Yes	Yes	No	No	Yes	2
GPS	4	1	2	10:30+ 10 Min	11:25+ 10 Min	Yes	Yes	No	No	Yes	4
GPS	3	1	2	10:45	11:40	No	No	No	No	Yes	5
GPS	6	1	2	12:45	13:40	Yes	Yes	Yes	Yes	Yes	No Match
GPS	6	1	2	10:45	11:40	No	No	No	No	No	No Match

*Match Grade is a hierarchical scale where 1 represents the highest-grade match and 5 the lowest.

A single diary trip can have multiple matches, as shown in Table 2. To identify the best match, each GPS trip is compared to its GPS match to calculate the departure time and distance differences between the trips. The trip matching algorithm uses the departure time difference, distance difference, and match grade to weight each match. As shown in Table 3, to calculate the weight, each metric is first scaled relative to the highest difference.

Table 3. Trip Match Scaling

Survey Trip ID	HH ID	Veh Num	GPS Trip ID	Match Grade	Dep Time Diff	Trip Distance Diff	Match Grade Scale*	Dep Time Diff Scale*	Trip Dist Diff Scale*	Match Grade Scaled#	Dep Tme Diff Scaled	Trip Dist Diff Scaled
2	1	2	3	1	20	5	75	1.5	1	75	30	5
2	1	2	5	2	10	50	75	1.5	1	150	15	50
2	1	2	4	4	200	300	75	1.5	1	300	300	300

*Based on the maximum difference for the trip ID; in the example it is 300 (gray shaded cell).

*Scale = Max Difference (300)/Each Difference).

 # scaled value = scale difference value

From these values, a composite ranking weight is calculated, as shown in Table 4. This composite weight utilizes factors giving highest value to differences in trip distance (60%), then match grade (30%), then time (10%). Determination of these weighting values was based on iterative testing of trips to see which combination produced the best results based on an empirical review of the outcomes.

Table 4. Trip Match Weighting

Survey Trip ID	HH ID	Veh Num	GPS Trip ID	Match Grade Weight	Dep Time Difference Weight	Trip Distance Difference Weight	Rank Weight	Rank
2	1	2	3	22.5	9.0	1.5	33.0	1
2	1	2	5	45.0	0.5	0.3	45.8	2
2	1	2	4	90.0	90.0	90.0	270.0	3

The outcome of these ranking is used to assign the best matched GPS trip to the best matched diary trip successively, starting at the first GPS trip. As each trip in the diary and GPS log is matched, it is marked as "matched" in both datasets so that it is not used for further matching.

Posted Speed Limit Data

High quality posted speed limit (PSL) data on a highly detailed road network was required to make a comparison with drivers' traveling speeds to determine whether they were speeding. Many free or open-source spatial datasets for road networks do not accurately represent the as-built environment due to a lack of detail (e.g., TxDOT Road Inventory Network) and have incomplete or imprecise PSL delineations (e.g., OpenStreetMap). However, it is possible to purchase preparatory data that is both detailed and includes reliable PSL data. This project acquired the navigational-quality HERE Streets spatial data set with speed limit data (https://www.here.com/products/mapping/map-data). The HERE Streets speed limit data have two sources: posted and derived. Posted speeds are based on a posted speed limit sign or speed limit information painted on the road, which the HERE field crews gather when they drive the roadways collecting street data. Posted speeds also include data obtained from official sources. Derived speeds are based on administrative regulations, such as state-specific speed limits on highways. Table 5 provides a breakdown of the HERE speed limit data source by roadway functional class.

Table 5. HERE Speed Limit Data Source by Roadway Functional Class

Functional Class	1 – Posted Sign	2 – Derived	3 – No Speed Limit
1	99%	1%	0%
2	98%	2%	0%
3	94%	6%	0%
4	91%	9%	0%
5	4%	1%	95%
Total	23%	2%	75%

GPS Log and Posted Speed Limit Linking

To discern whether or not a driver was speeding, the posted speed limits were joined to each GPS trace point based on its location within a Geographic Information Systems (GIS) in a process known as map matching. Each GPS trace point was plotted and spatially joined with the three nearest HERE network segments. The distance between the point and each segment was retained for later use. Next, a map matching algorithm was developed to correct for errors created during the spatial joining process. This was necessary because the nearest segment is not always assigned to the correct segment in terms of network connectivity and travel path direction. For example, GPS trace points that cross a grade separated intersection (i.e., overpass) can be incorrectly joined to the opposing roadway purely because it was the closest segment. The map matching algorithm takes into account the segment ID and roadway name assigned to the point before and after to assess with which segment the current point should actually be joined. The map matching algorithm is done on a trip-by-trip, point-by-point basis to check roadway segment IDs and name, as illustrated in Figure 6.

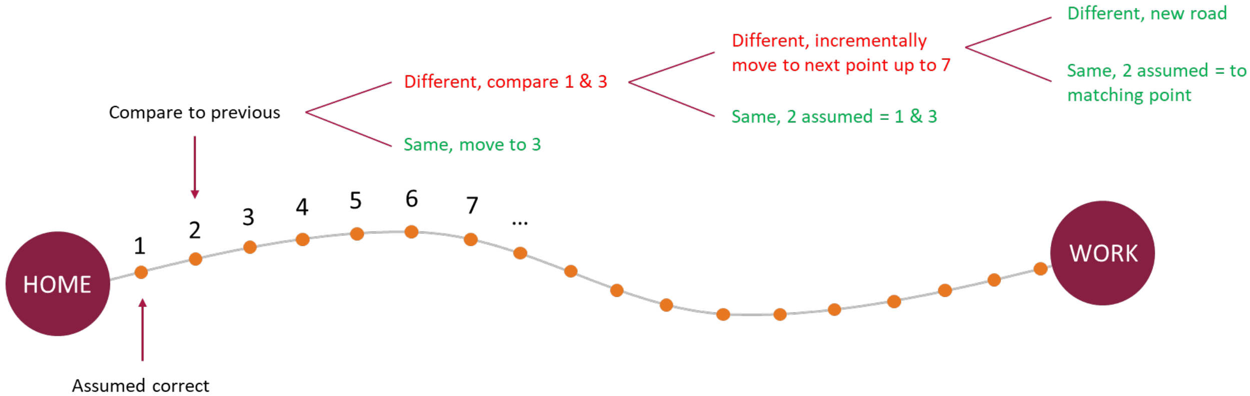

Figure 6. Map matching algorithm.

Trip Summarization

The final analysis dataset consists of a summary for each trip in terms of distance and time traveled, driver and passenger demographics, number of passengers, and percentage of free-flow time that the driver was speeding relative to the posted speed limit. Free-flow speed was estimated to be any traveling speed of 5 mph below the posted speed limit or greater at any given GPS trip point. This prevented percentage of free-flow speeding from being biased by things like stops at signalized intersections.

Any parts of GPS trips that occurred on private road segments, parking lots, alleys, or that were unassigned to a segment were removed from the dataset before trip summarization, along with the first and last 30 seconds of the trip. To anonymize the data, no personally identifiable information was carried over to the trip summarization file.

Statistical Analysis and Modeling

The overarching goal of the statistical analysis was to demonstrate how household travel data could be used to address research questions relating to traffic safety. The test case for this project was to identify factors associated with speeding behavior among driver groups of interest. Three user groups of interest were identified:

- younger drivers (16–24 years),
- all drivers 16–55 years with a child passenger(s) (0–15 years)[1], and
- older drivers (65+ years).

Three separate statistical models were developed, one for each vulnerable user group. Coding difficulties precluded the strict development of a "parent" driver group, which is discussed in greater detail in the Limitations section.

Model Variable Definitions

For the test case analysis, the dependent variable was dichotomous and was coded as (1) if the driver was speeding for at least 20% of the free flow duration of each trip and (0) otherwise. Generally, the main independent variable of interest also was dichotomous and coded as (1) if a passenger with a specific characteristic was present in the vehicle and (0) otherwise. For the older drivers, the adult passenger variable was coded as (1) if a passenger 25 years or older was present and (0) otherwise. For the younger drivers, the passenger variables included both the adult passenger variable previously described and the child passenger variable, which was coded as (1) if a passenger 0–15 years was present and (0) otherwise. For all drivers 16–55 years, the same child passenger variable was included. Seven additional variables also were considered as independent variables, as shown in Table 6.

Table 6. Independent Variables

Variable	Codes
Driver's Gender	0 = female, 1 = male
Driver's Age	Continuous
Driver's Ethnicity	0 = White/Caucasian, 1 = Black/African American, 2 = Hispanic/Mexican American, 3 = Other
Driver's Employment	0 = unemployed, 1 = employed
Study Area	See Table 7
Trip Occupancy Type	0 = 0 passengers, 1 = 1 passengers, and 2 = 2 or more persons
Child Occupancy Type	0 = 0 children, 1 = 1 child, and 2 = 2 or more children
Driver with Child	0 = Drivers without children 0–15 years as passengers, 1 = Drivers with children 0–15 years as passengers
Driver with Adult	Drivers with adults 25 years and up as passengers

[1] Note: While one of the most critical driver groups of interest was parents, there was not a straight-forward method of coding the parent-child relationship. This category was thus recast as all drivers 16–55 years with a child passenger(s) (0–15 years).

Statistical Analysis

Initially, all variables were described using counts and proportions, or means, medians, and ranges, depending on the form (i.e., categorical or continuous) of the variable. This included cross tabulations of each independent variable by the dependent variable. Next, statistical models were constructed to quantify the magnitude of the association between passenger presence and speeding behavior while accounting for the effect of potential confounders and identifying other important independent variables. Separate models were constructed for each driver group of interest. Given that the outcome variable was dichotomous, binomial logistic regression was selected for modeling. The level of significance was set a priori at $\alpha = 0.05$. Adjusted odds ratios and associated 95% confidence intervals were computed. Since the data was at the trip level with multiple trips per driver, robust standard errors were computed to address the potential violation of the model assumption of independence of observations. In addition, terms for study areas were forced into all models to account for similarities in speeding behavior by study area. Initially, univariate models were constructed with the outcome and each independent variable. Variables with a $p < 0.25$ were included in an intermediate model. Using a backward selection approach, variables with an adjusted odds ratios closest to the null value (OR = 1.0) and with the highest p-values were removed one at a time until all variables in the final model, beyond the main independent variable of interest and study area, were statistically significant. This approach is only one way of analyzing these data. Other approaches include multi-level analysis and analyzing speeding in a continuous form. We selected this approach for the test case demonstration since it could be implemented relatively easily and produces results that are fairly intuitive in their interpretation.

Descriptive Statistics

The following tables describe the reduced household travel survey data for each driver group of interest in terms of persons and trips by study area. Table 7 illustrates the distribution of households and persons per study area. Houston was the highest contributor at 20% of both categories and Corpus Christi the lowest at 5% for both categories

Table 7. Distribution of Households, Persons, and Vulnerable Users

Study Area	Households (N = 2,058)	Persons (N = 2,939)	Older Drivers (N = 744)	Younger Drivers (N = 240)	Drivers 16–55 with a child passenger(s) (N = 1,605)
	N (%)	N (%)	N (%)	N (%)	N (%)
Abilene	146 (7)	186 (6)	37 (5)	17 (7)	111 (7)
Bryan-College Station	166 (8)	213 (7)	51 (7)	11 (5)	104 (6)
Corpus Christi	98 (5)	135 (5)	27 (4)	7 (3)	71 (4)
El Paso	231 (11)	350 (12)	75 (10)	37 (15)	216 (13)
Houston	410 (20)	589 (20)	114 (15)	52 (22)	366 (23)
Midland-Odessa	154 (7)	231 (8)	91 (12)	14 (6)	90 (6)
San Angelo	185 (9)	254 (9)	81 (11)	20 (8)	121 (8)
Sherman-Denison	173 (8)	255 (9)	97 (13)	24 (10)	118 (7)
Texarkana	243 (12)	362 (12)	84 (11)	37 (15)	204 (13)
Victoria	123 (6)	178 (6)	47 (6)	8 (3)	92 (6)
Wichita Falls	129 (6)	186 (6)	40 (5)	13 (5)	112 (7)

Table 8 shows the distribution of trips by study area for each driver group of interest. Each group has a fairly similar range. Drivers 16–55 made the highest number of average daily trips at 3.80 trips, whereas younger drivers made the lowest average number of 3.44 trips.

Table 8. Distribution of Trips

Study Area	Older Drivers (N = 2,654)	Younger Drivers (N = 825)	Drivers 16–55 with child passenger(s) (N = 6,106)
	N (%)	**N (%)**	**N (%)**
Abilene	113 (4)	50 (6)	342 (6)
Bryan-College Station	204 (8)	36 (4)	403 (7)
Corpus Christi	97 (4)	27 (3)	273 (4)
El Paso	267 (10)	122 (15)	872 (14)
Houston	422 (16)	173 (21)	1,454 (24)
Midland-Odessa	379 (14)	54 (7)	356 (6)
San Angelo	273 (10)	85 (10)	482 (8)
Sherman-Denison	280 (11)	79 (10)	363 (6)
Texarkana	330 (12)	132 (16)	784 (13)
Victoria	149 (6)	21 (3)	332 (5)
Wichita Falls	140 (5)	46 (6)	445 (7)
AVERAGE	3.57	3.44	3.80

Table 9 illustrates that the gender split for each driver group of interest is fairly balanced overall between males and females. Female drivers between 16–55 years made the highest average daily trips at 4.01 trips per day, whereas young male drivers had the lowest average of 3.27 daily trips. Certain study areas do have gender imbalances, such as Bryan-College Station and Corpus Christi.

Table 9. Distribution of Trips by Gender

Study Area	Older Drivers	(N = 2,654)	Younger Drivers	(N = 825)	Drivers 16–55 with child passenger(s)	(N = 6,106)
	N (%)	**N (%)**	**N (%)**	**N (%)**	**N (%)**	**N (%)**
	Female	**Male**	**Female**	**Male**	**Female**	**Male**
Abilene	75 (66)	38 (34)	27 (54)	23 (46)	181 (53)	161 (47)
Bryan-College Station	64 (31)	140 (69)	22 (61)	14 (39)	267 (66)	136 (34)
Corpus Christi	27 (28)	70 (72)	21 (78)	6 (22)	145 (53)	128 (47)
El Paso	106 (40)	161 (60)	53 (43)	69 (57)	511 (59)	361 (41)
Houston	206 (49)	216 (51)	87 (50)	86 (50)	842 (58)	612 (42)
Midland-Odessa	183 (48)	196 (52)	32 (59)	22 (41)	196 (55)	160 (45)
San Angelo	121 (44)	152 (56)	54 (64)	31 (36)	304 (63)	178 (37)
Sherman-Denison	118 (42)	162 (58)	62 (78)	17 (22)	230 (63)	133 (37)
Texarkana	164 (50)	166 (50)	72 (55)	60 (45)	502 (64)	282 (36)
Victoria	61 (41)	88 (59)	6 (29)	15 (71)	216 (65)	116 (35)
Wichita Falls	84 (60)	56 (40)	16 (35)	30 (65)	231 (52)	214 (48)
GROUP TOTAL	**1,209 (46)**	**1,445 (54)**	**452 (55)**	**373 (45)**	**3,625 (59)**	**2,481 (41)**
AVERAGE # OF DAILY TRIPS	**3.51**	**3.61**	**3.59**	**3.27**	**4.01**	**3.54**

Results

This results section provides a summary of select results from the modeling efforts. The results are organized by the driver groups of interest. Sherman-Denison were selected as the statistical referent group since it possessed the least amount of speeding behavior while maintaining statistical stability with substantial sample sizes.

Younger Drivers (Ages 16–24)

Based on the model constructed for younger drivers (ages 16–24), the following variables were associated with speeding. A total of 825 trips were included in the model.

- Contrary to hypothesized associations, gender was not associated with speeding among younger drivers.
- Speeding varied among younger drivers across study area. The odds of speeding were larger among younger drivers in Houston (adjOR = 5.34; 95% CI = 2.27–12.55), Texarkana (adjOR = 2.51; 95% CI = 1.03-6.12), and San Angelo (adjOR = 3.38; 95% CI = 1.23-9.33) compared to drivers in Sherman-Denison.
- There is some evidence to suggest that younger drivers speed less when there is an adult passenger in the vehicle. The odds of speeding were 77% less among those with an adult passenger in the vehicle, but this finding was not statistically significant (adjOR = 0.23; 95% CI = 0.02-3.22). However, this dataset did not include a large number of young drivers, and therefore this analysis should be rerun in the future with a larger sample size. Statistical significance may not have been reached due to a lack of power.
- The variable for a young passenger in the vehicle was not statistically significant and the adjusted odds ratio was very close to 1.00, indicating no association.

All Drivers (Ages 16–55) with Child Passenger(s) (Ages 0–15)

Based on the model constructed for adult drivers (ages 16–55), the following variables were associated with speeding. The number of trips included in the model was 6,106.

- The odds of speeding were 27% lower among those with a child passenger (0–15 years) in the vehicle compared to those without a child passenger (adjOR = 0.73; 95% CI = 0.60–0.87).
- Speeding varied among adult drivers across study areas. The odds of speeding were increased among drivers in Corpus Christi (adjOR = 1.63; 95% CI = 1.04–2.55), El Paso (adjOR = 1.64; 95% CI = 1.16–2.31), Houston (adjOR = 2.64; 95% CI = 1.92–3.63), and Texarkana (adjOR = 1.55; 95% CI = 1.09–2.19), compared to Sherman-Denison.
- The odds of speeding were 77% greater among employed adult drivers compared to unemployed adult drivers (adjOR = 1.77; 95% CI = 1.48–2.12).
- The odds of speeding decreased by 2% for each year increase in drivers' age (adjOR = 0.98; 95% CI = 0.97–0.99)

Older Drivers (Ages 65+)

Based on the model constructed for older drivers (ages 65+), the following variables were associated with speeding. A total of 2,654 trips were included in the model.

- The odds of speeding were 69% greater among black older drivers compared to white older drivers (adjOR = 1.69; 95% CI = 1.07–2.65).
- The odds of speeding were 41% greater among employed older drivers compared to unemployed older drivers (adjOR = 1.41; 95% CI = 1.08–1.83).
- The odds of speeding were 31% greater among male older drivers compared to female older drivers (adjOR = 1.31; 95% CI = 1.02–1.69).
- The assumption of linearity was met given the observed pattern of decreasing odds as age categories increased. Age entered as a continuous variable was statistically significant in the model.
- In its linear form, for each year increase in age, the odds of speeding decreased by 6% (adjOR = 0.94; 95% CI = 0.91–0.96).
- Having an adult occupant aged 25 or older in the vehicle was not statistically significant based on a strict alpha of 0.05 (the upper bound of the 95% CI is 1.00), but this finding suggests that the odds of speeding may be less when at least one adult occupant is in the vehicle (adjOR = 0.73; 95% CI 0.54–1.00).

Discussion

A large portion of this research effort focused on the methodology and processes to make it possible to model the data used in this novel approach. Thus, further model development is an area with rich potential for future research using the project's derived dataset. This may include further binomial logistic regression modeling, or the use of more sophisticated techniques. However, even the initial modeling that was performed as part of this research effort helps provide insights into driver and passenger characteristics that may be linked to speeding. One key takeaway is that drivers speed less when traveling with a passenger, especially if the passenger is a child.

Limitations

This research project represents a novel approach to safety research and eliminates much of the guesswork that has traditionally been associated with the study of how passengers affect driving behavior. However, it has some limitations. While one of the most critical driver groups of interest was parents, there was not a straight-forward method of coding the parent-child relationship. Thus, the "parents" category was recast as adult drivers of a specified age traveling with children. This slight distinction does not allow for strong enough conclusions to be drawn for this user group. Further coding could be performed to allow for parent-child relationships to be detected and used in developing a better-defined parent category.

An additional limitation lies in the data cleaning algorithms. Although great effort and logic were put into cleaning the data and ensuring that specified data standards were met during the sample development process, the process is imperfect and some anomalies and errors may still be present within the GPS and survey data. Additionally, although HERE network data containing speed limit information was purchased and used for this project, it is possible that some speed limit discrepancies still existed or that speed limits could have changed between the study area survey year and the HERE speed limits supplied for the year 2018.

Additionally, as indicated in the Results section, some of the associations found were not significant within the current sample sizes used in modeling. However, continuing to expand the dataset using data from additional household study areas obtained from TxDOT in the future may help to increase the modeling power in future studies.

Conclusions and Recommendations

The work performed as part of this research project not only produced preliminary results using a novel dataset and approach to the area of safety, but also laid the groundwork for several additional research avenues. The cleaning and algorithm development processes were time intensive, resulting in a rich dataset that could be expanded and utilized in answering a number of additional research questions. Though not an exhaustive list by any means, the following list provides a good start for brainstorming future avenues of research that could build on the progress associated with this research project.

- **Expanded Sample Scope:** The cleaning and processing algorithms developed as part of this project could be rapidly applied to household travel survey information from additional Texas study areas, expanding the sample size of the current project. It may also be possible to expand the sample size even more by applying the analysis techniques to household travel survey data collected in an area outside of Texas. While the TxDOT TSP is among the most extensive travel survey programs in the country, other areas may have household travel survey data that could be used to supplement the current research project and expand its scope and sample size. National Household Travel Survey data (past, current, or future) may be a good place to start.
- **Additional Model Development:** Given more time, additional, more complex models could be developed. While the focus of this project was on the relationship of speeding and vehicle occupancy for the driver groups of interest, any number of other angles could be taken to assess additional characteristics that may impact speeding. Examples may include focusing on trip time of day, type of trip, type of roadway, and vehicle type. There is also a need for developing an optimal modeling approach that includes the definition of outcome variables (e.g., examining other definitions of speeding behavior beyond a threshold of 20%).

- **Additional Speed Study Approaches:** The approach used in defining speeding for this research project is just one study design among several. Future research could take a different approach to studying speeding, such as a longitudinal speed study.
- **Link to Crash Records:** Speeding information could be linked to crash records, such as the Crash Records Information System, to study the link between not only vehicle occupancy and speeding, but also vehicle occupancy, speeding, and crashes.
- **Comparison to Naturalistic Driving Study:** Future research efforts may also include teaming with naturalistic driving study experts and comparing this project's methods and results more directly to naturalistic driving study methods and results.
- **Data Cleaning Sensitivity Analysis:** The cleaning and algorithm development process used for this project could be reviewed to see how the modeling results change based on the cut-off criteria used to determine which GPS trip traces are adequate for inclusion in the analysis. This could be viewed as a type of sensitivity analysis.

The wide array of potential future research topics that could be pursued using the techniques, coding, and algorithms developed from this novel approach to using household travel survey data in the area of safety speaks to the value of this research project. This project, and the research that will stem from it, are a true contribution to safety through disruption.

Additional Products

The Education and Workforce Development (EWD) and Technology Transfer (T2) products created as part of this project can be downloaded from the project page on the Safe-D website. The final project dataset is located on the Safe-D Dataverse.

Education and Workforce Development Products

A PowerPoint presentation was developed for use as a learning module for a graduate level transportation engineering class. It provides some background on this research project, the datasets used, and an overview of the data cleaning and merging processes. It also provides a small sample example of how to use the data in developing binomial logistic regression models in R. An overview of how to interpret model results is provided.

Slides to be used in the transportation planning session of the AVID middle school presentation were also developed. They provide a high-level overview of the datasets and research performed and help to emphasize that different areas of transportation research (such as planning and safety) are interrelated.

In addition, a brown bag seminar/webinar describing this project and its findings was held on November 29, 2018. Key TxDOT personnel were invited to participate.

Technology Transfer Products

Efforts to publish the results of this research in an academic journal are underway, based on material largely derived from this report. The journal article was submitted to the *Transportation Research Record* journal as part of the 2019 Annual Meeting call for papers and was selected for both poster presentation and publication (#20-04603, *Vehicle Occupants and Driver Behavior: An Assessment of Vulnerable User Groups*). The authors are currently communicating with the publisher and conducting manuscript edits. Poster presentation will be held the 2020 TRB Annual Meeting in Washington, D.C. on January 13th in the *Transportation Safety Management from Start to Finish* session (#1284).

Data Products

The model data sets resulting from the cleaning and merging process performed on the TxDOT household travel survey data are provided on Safe-D Dataverse (https://doi.org/10.15787/VTT1/YRTS1Z). The final data products for this project consists of three CSV files representing each of the driver groups of interest:

- all_driver_data
- older_driver_data
- younger_driver_data

Each file contains the following variables: trip ID, unique driver ID, study area code, speeding 20% of free flow trip duration indicator, driver's gender, driver's age, driver's ethnicity, driver's employment status, occupancy type code, child occupancy type code, driver with child indicator, and driver with adult indicator.

References

[1] U.S Department of Transportation Federal Highway Administration. (2000). Speeding and Highway Safety-The U.S. Department of Transportation's Policy and Implementation Strategy. Retrieved June 28, 2017, https://safety.fhwa.dot.gov/speedmgt/policy/spdppolicybro.cfm

[2] Dingus, T. A., Hankey, J. M., Antin, J. F., Lee, S. E., Eichelberger, L., Stulce, K. E., . . . Stowe, L. (2015). Naturalistic Driving Study: Technical Coordination and Quality Control: Virginia Tech Transportation Institute.

[3] Threewitt, C., & Vincent, J. M. (2018). How Do Those Car Insurance Tracking Devices Work? https://cars.usnews.com/cars-trucks/how-do-those-car-insurance-tracking-devices-work

[4] Fleiter, J. J., Lennon, A., & Watson, B. (2010). How do other people influence your driving speed? Exploring the 'who' and the 'how' of social influences on speeding from a qualitative perspective. *Transportation Research Part F: Traffic Psycology and Behavior, 13*, 49-62. doi: 10.1016/j.trf.2009.10.002

[5] Lee, C., & Abdel-Aty, M. (2008). Presence of passengers: Does it increase or reduce driver's crash potential? *Accident Analysis and Prevention, 40*, 1703-1712. doi: 10.1016/j.aap.2008.06.006

[6] Vollrath, M., Meilinger, T., & Krüger, H.-P. (2002). How the presence of passengers influences the risk of a collision with another vehicle. *Accident Analysis & Prevention, 34*(5), 649-654. doi: https://doi.org/10.1016/S0001-4575(01)00064-1

[7] Preusser, D. F., Ferguson, S. A., & Williams, A. F. (1998). The effect of teenage passengers on the fatal crash risk of teenage drivers. *Accident Analysis & Prevention, 30*(2), 217-222. doi: https://doi.org/10.1016/S0001-4575(97)00081-X

[8] Doherty, S. T., Andrey, J. C., & MacGregor, C. (1998). The situational risks of young drivers: the influence of passengers, time of day and day of week on accident rates*,* 45.

[9] Aldridge, B., Himmler, M., Aultman-Hall, L., & Stamatiadis, N. (1999). Impact of Passengers on Young Driver Safety. *Transportation Research Record: Journal of the Transportation Research Board, 1693*, 25-30. doi: 10.3141/1693-05

[10] Fu, H., & Wilmot, C. (2008). Effect of Passenger Age and Gender on Fatal Crash Risks of Young Drivers. *Transportation Research Record: Journal of the Transportation Research Board, 2078*, 33-40. doi: 10.3141/2078-05

[11] Lam, L. T., Norton, R., Woodward, M., Connor, J., & Ameratunga, S. (2003). Passenger carriage and car crash injury: a comparison between younger and older drivers*,* 861.

[12] Rosenbloom, T., & Perlman, A. (2016). Tendency to commit traffic violations and presence of passengers in the car. *Transportation Research Part F: Traffic Psychology and Behaviour, 39*, 10-18. doi: https://doi.org/10.1016/j.trf.2016.02.008

[13] Keall, M. D., Frith, W. J., & Patterson, T. L. (2004). The influence of alcohol, age and number of passengers on the night-time risk of driver fatal injury in New Zealand. *Accident Analysis & Prevention, 36*(1), 49-61. doi: https://doi.org/10.1016/S0001-4575(02)00114-8

[14] Chen, L. H., Baker, S. P., Li, G., & Braver, E. R. (2000). Carrying passengers as a risk factor for crashes fatal to 16- and 17- year-old drivers. *Journal of the American Medical Association, 283*(12), 1578-1582. doi: 10.1001/jama.283.12.1578

[15] State of California. (2018). *California Driver Handbook*. Retrieved from https://www.dmv.ca.gov/web/eng_pdf/dl600.pdf.

[16] Texas Department of Public Safety. (2017). *Texas Driver Handbook*. Retrieved from https://driving-tests.org/wp-content/uploads/2017/10/TX_DL-7.pdf.
[17] Shope, J. T. (2007). Graduated driver licensing: Review of evaluation results since 2002. *Journal of Safety Research, 38*, 165-175. doi: 10.1016/j.jsr.2007.02.004
[18] Koppel, S., Charlton, J., Kopinathan, C., & Taranto, D. (2011). Are child occupants a significant source of driving distraction? *Accident Analysis & Prevention, 43*(3), 1236-1244. doi: https://doi.org/10.1016/j.aap.2011.01.005
[19] Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., & Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proceedings of the National Academy of Sciences, 113*(10), 2636-2641. doi: 10.1073/pnas.1513271113
[20] Hing, J. Y., Stamatiadis, N., & Aultman-Hall, L. (2003). Evaluating the impact of passengers on the safety of older drivers*,* 343.
[21] Braver, E. R., & Trempel, R. E. (2004). Are older drivers actually at higher risk of involvement in collisions resulting in deaths or non-fatal injuries among their passengers and other road users? *Injury Prevention, 10*(1), 27-32. doi: 10.1136/ip.2003.002923