2015

Development of rural curve driving models using lateral placement and prediction of lane departures using the SHRP 2 naturalistic driving data

Nicole Lynn Oneyear
Iowa State University

Development of rural curve driving models using lateral placement and prediction of lane departures using the SHRP 2 naturalistic driving data

by

Nicole Oneyear

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Civil Engineering (Transportation Engineering)

Program of Study Committee:
Shauna L. Hallmark, Major Professor
Alicia L. Carriquiry
Jing Dong
Omar Smadi
Sri Sritharan

Iowa State University

Ames, Iowa

2015

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

This work was sponsored by the Federal Highway Administration in cooperation with the American Association of State Highway and Transportation Officials, and it was conducted in the Strategic Highway Research Program, which is administered by the Transportation Research Board of the National Academies. In addition I'd like to thank the Federal Highway Administration Dwight David Eisenhower Transportation Fellowship Program for their support of my graduate studies and this research. I would also like to thank the Midwest Transportation Center for their additional financial support in conducting this research.

I'd also like to thank my committee, especially my major professor Shauna Hallmark, for their insight, guidance and constructive comments in the development of this dissertation. Additionally, thank you to everyone who helped in reducing data for this project, specifically Cher Carney at University of Iowa, Bo Wang and Jordan Turner at Iowa State University. Skylar Knickerbocker and Zach Hans also provided help in utilizing the RID data. Finally, thank you to Samantha Tyner who helped answer all of my statistics related questions.

ABSTRACT

Roadway departure crashes are a major cause of fatalities on rural horizontal curves. In 2008, the Federal Highway Administration estimated that 27% of all fatalities occurred on rural highways and that among those 76% were single vehicles leaving the roadway and striking a fixed object or overturning while another 11% were head-on collisions (AASHTO 2008). Addressing crashes on rural two lane curves, specifically run off the road crashes, remains a priority for our local, state and national roadway agencies.

Much research has been conducted to look at what factors affect curve negotiation, and which factors are more likely to contribute to roadway departures. Previous research has studied how roadway factors, such as radius and shoulder width and environmental factors, such as weather affect crashes, yet limited research has been conducted looking at how driver behaviors affect crash risk. Additional research has been conducted on developing curve negotiation trajectories using small sets of curves and without much driver information.

The recent completion of the Strategic Highway Research Program 2 (SHRP 2) Naturalistic Driving Study (NDS) and Roadway Information Database (RID) allows one to expand on gaps in current literature by utilizing data from a wide variety of participants in multiple states across a broad age ranges. It also allows one to include driver factors such as age and gender, as well as drivers glance behavior and presence of distractions.

This dissertation utilizes early data from the SHRP 2 NDS and RID to develop models which provide an additional understanding of rural curve negotiation. Through three papers, two curve driving models were developed as well a model which predicts the likelihood of lane departures based off kinematic vehicle data.

In the first paper (Chapter 2) a model of normal curve driving trajectories on isolated rural two lane curves was developed using generalized least squares with an autocorrelation

structure. This model found that a drivers offset 100 meters upstream of the start of the curve could help predict a vehicles position at various points throughout the curve. Additionally, the model was able to predict the average path a driver would take through seven points in the curve. These estimators suggest that drivers tend to cut the curve and are more susceptible to a lane departure at certain points in the curve.

Chapter 3, the second paper, builds on the model developed in Chapter 2 and includes additional non-isolated curves as well as non-normal driving (i.e. lane encroachments). This linear mixed effects model of curve driving trajectories included random effects for the repeated samples of drivers and drivers within the same curve as well as the same autocorrelation structure. This model was able to determine a difference in the offset at each point in the curve for those traces where a lane departure towards the inside of curve occurred and when it did not. This allowed for a boundary between normal and non-normal driving to be established. A similar correlation between the driver's lane position upstream of the curve and lane position in the curve was also found. Smaller radii, looking down and being distracted were all found to affect trajectories in rural curves.

The final paper, Chapter 4, includes a mixed logistic regression which included a random effect for curve which took into account the repeated samples for the curves. This model produced odds-ratios for the three variables and found that increasing the amount over the advisory speed by 1 mph at the Point of Curvature (PC) of the curve increased odds of a lane encroachment towards the inside of the curve by 1.11. Shifting lane position by 0.1 m towards the inside of the curve at the PC increased odds of an inside lane departure by 1.5. In addition to the logistic regression model, two linear mixed effects models were developed which allow one to predict the speed and offset at the PC using data from 100 m upstream. This allows one to predict the probability of a lane departure 100 m upstream of the curve in addition to at the PC.

CHAPTER 1: INTRODUCTION

1.1 Background

According to the Federal Highway Administration, a horizontal curve is a part of the roadway that changes the alignment or direction of the road. Horizontal curves make up a small portion of our total roadway miles, yet they were the site of 27% of all fatalities in 2008. Of this 27% of total fatal crashes, 76% were single vehicles leaving the roadway and striking a fixed object or overturning. Another 11% were head-on collisions (AASHTO 2008). Therefore, in 2008 approximately 23% of all fatalities were the result of lane departure crashes on horizontal curves.

Due to the small percentage of roadway miles curves represent, yet the large amount of crashes we see, fatal crashes tend to be overrepresented on curves. A study by Glennon et al. (1985), found that the crash rate on curves is approximately three times the rate on tangent sections. Preston (2009) reported that 25% to 50% of severe road departure crashes in Minnesota occurred on curves, even though they only account for 10% of the system mileage. Addressing crashes on rural two lane curves, specifically run off the road crashes, remains a priority for our local, state and national roadway agencies.

Reducing serious injuries and fatalities due to lane departures is an area of focus in the majority of Strategic Highway Safety Plans (SHSP). In addition to the States' SHSP's, FHWA has recently published a Roadway Departure Strategic Plan which hopes to reduce fatalities by half from 17,000 annually to 8,500 by 2030. In order to accomplish this their mission is to develop, evaluate and deploy life-saving countermeasures and promote data-driven application of safety treatments (FHWA 2013).

1.1.1 Background on SHRP 2 Naturalistic Driving Study

The SHRP 2 NDS represents the largest naturalistic driving study to date. The study was conducted by Virginia Tech Transportation Institute (VTTI). Drivers in six states (Florida, Indiana, New York, North Carolina, Pennsylvania and Washington) had their vehicles equipped with a Data Acquisition System (DAS) which collected information such as speed, acceleration, GPS data, and radar, as well as four cameras which collected forward, rear, drivers face and over the shoulder video. These equipment captured all of the trips a driver made over a period of six months up to two years. Males and females ages 16 to 98 and older participated in the study. Over the three years of the study approximately 3,400 participants drove over 30 million data miles during 5 million trips (Antin 2013 and VTTI 2014).

1.1.2 Background on SHRP 2 Roadway Information Database

In conjunction with the SHRP 2 Naturalistic Driving Study, another project was conducted to collect roadway information for the main roads traveled in the NDS. The Center for Research and Education (CTRE) lead the effort which used mobile data collection vans to collect 12,500 center line miles of data across the six states where the NDS was focused. Data collected included information on roadway alignment, signing, lighting, intersection location and types, presence of rumblestrips as well as other countermeasures. In addition to the mobile data collection effort, existing roadway data collected by local agencies were leveraged to increase the data available. Additionally, supplemental data such as crash data, changes to laws, and construction projects were also collected to further strengthen the database (Smadi 2012).

1.2 Previous Research

1.2.1 Factors contributing to run off the road crashes

Previous research has addressed environmental factors, driver factors and to a large extent roadway factors which contribute to run off the road crashes. In the next few sections major research contributions addressing that factors which have been found to affect run off the road crashes and curve negotiation will be addressed. Studies are discussed in chronological order.

1.2.1.1 Roadway

Roadway factors are among the most studied factors affecting roadway departure crashes. This is due to roadway data being largely available and easily accessible. From the literature, it has been found that degree of curve or radius of curve, presence of spirals, distance between curves and shoulder width and type are the most relevant curve characteristics that affect lane negotiation and lane departures.

Zegeer et al. (1991), studied crash rates at 10,900 horizontal rural two lane curves in Washington State. They studied how roadway factors affect these rates and found through their weighted least squares models that crash rates were significantly higher on shaper curves, narrower widths (lane + shoulder), curves without spirals and as the difference between actual super elevation and optimal super elevation increases.

Miaou and Lum (1993) used a Poisson regression model with data on truck crashes from 1985-1989 obtained from five states in the Highway Safety Information System. Models showed a relationship between crash rates the degree of curvature.

Fink and Krammes (1995) found that crash rates increased for curves following long tangent sections as well as very short tangent sections.

Council (1998) used a database containing the same 10,900 curves used by Zegeer et al (1991) and crash data from 1982 to 1986 to model the effect of spirals on curve crash rates. They found based on a logistic regression model using 8,271 records that on level terrain spirals are beneficial on sharper curve (degree of curvature greater than 3 degrees).

Milton and Mannering (1998) used crash frequencies from principal arterials in Washington State for 1992 and 1993 to create a negative binomial regression model to predict crash frequency. A strong relationship between curve radius and crash frequency was found that as radius increases, crash frequency decreases. It was also found that the longer tangent lengths before the curve led to higher crash frequencies.

A study by Caliendo et al (2007) determined using a negative multinomial regression model built on data from 5 years of crashes on a 4 lane median divided motorway in Italy that both total and severe crashes increase with the length, decreases in curvature, pavement friction and longitudinal slope.

Montella (2009) evaluated crashes occurring from 2001-2005, before and after installation of delineation improvements such as (chevron signs, curve warning signs, and sequential flashing beacons or a combination of all three) on 15 curves in Italy using empirical Bayes. All curves were characterized by a small radius (mean = 365 meters), large deflection angle, and sight distance issues. The study found that increasing delineation with all three of the treatments listed reduced crashes by approximately 47.6%. It also found improved delineation was more effective for smaller radii curves.

A Bayesian semi-parametric estimation procedure was used by Shively et al. (2010) to model counts of crashes on rural two lane roads in the Puget Sound region of Washington State in 2002. A relationship between crashes and curve rates once a radius becomes 1400 feet or less was found. Their model found that as degree of curve increased from 4 to 12 degrees the expected number of crashes increased by 0.06 crashes. They also found that as curve length increased, the expected number of crashes would also increase.

Location of a curve in relation to other curves was taken into consideration to evaluate the safety of a curve in this study. Spatial considerations of the curves influence the safety of the curves because of the driver's expectation to encounter additional curves.

A study by Findley et al (2012) highlighted the importance and significance of spatial considerations for the prediction of horizontal curve safety. The study results showed that distance to adjacent curves was a significant factor in estimating the observed collision in a curve. The study revealed that more closely spaced curves had fewer prediction collisions than those curves which were more distant to each other. The study revealed that a series of curves is expected to be safer than a curve which is isolated from other curves.

1.2.1.2 Environmental

Environmental factors, such as the roadway surface condition will also have an impact on a driver's ability to safely negotiate a curve.

Neuman et al. (2003) found using the 1999 statistics from FARS that for two lane undivided, non-interchange, non-junction roadways that 11% of single vehicle ROR crashes were on wet surfaces, and 3% more occurring when snow or ice were present.

Caliendo et al. (2007) found that both total and severe crashes increased significantly during rain by a factor of 2.7 for total and 3.26 for severe compared to dry using models based on data from 5 years of crashes on a 4 lane median divided motorway in Italy.

McLaughlin et al. (2009) evaluated run-off-road crashes (ROR) and near-crashes in the VTTI 100 car study where 30% of all these crash and near crashes occurred on curves. They found that ROR events were 1.8 times more likely on wet roads than dry, 7 times more likely on roads with snow or ice than dry roads, and 2.5 times more likely in nighttime versus daytime conditions.

1.2.1.3 Driver

Research on driver factors and behaviors which affect ROR crashes have found age, speeding and distraction to all be contributing factors.

A study by McGwin and Brown (1998) found that older drivers were less likely to have crashes on curves based on an analysis of 1996 crash data from Alabama.

Driver error on horizontal curves is often due to inappropriate speed selection, which results in an inability to maintain lane position. FHWA estimates that approximately 56% of ROR fatal crashes on curves are speed related. A study by Davis et al. (2006) using two case control analyses of ROR crashes from Australia and Minnesota and Bayesian relative risk regression found that 5 out of 10 fatal crashes in Minnesota which they investigated would have been prevented had the driver adhered strictly to the posted speed limit.

Distracting tasks such as radio tuning or cell phone conversations can draw a driver's attention away from speed monitoring, changes in roadway direction, lane keeping, and detection of potential hazards (Charlton 2007). Other factors include sight distance issues, fatigue, or complexity of the driving situation (Charlton and DePont 2007, Charlton 2007).

McLaughlin et al. (2009) evaluated ROR crashes and near crashes in the Virginia Tech Transportation Institute (VTTI) 100-car naturalistic driving study and found that distraction was the most frequently identified contributing factor, occurring in 40% of all events. Additionally fatigue, impairment, and maneuvering errors also contributed.

1.2.1.4 Exposure

As would be expected, the larger the ADT, the more chances for a lane departure. A study by Caliendo et al (2007) confirmed this with their Negative Multinomial regression model built on data from 5 years of crashes on a 4 lane median divided motorway in Italy that found both total and severe crashes increase as AADT of the curve increases.

1.2.2 Crash Surrogates Related to Roadway Departures

The factors listed above have been determined to affect the crash risk on rural curves. Crashes tend to be rare and the use of crash data to address safety problems is a reactive approach which is not able to take into account events that lead to successful outcomes (Tarko et al., 2009). Consequently, researchers have proposed use of crash surrogates, as a measure of safety. Additionally, the use of surrogates provides an opportunity to study what happens preceding and following an incident or event.

Time to collision is one of the most common lane departure crash surrogates used. The concept is logical and provides a repeatable and easily understood metric to assess level of crash risk. Risk can be measured as a function of TTC, where at TTC = 0, the subject vehicle and another vehicle/object collide. This makes setting boundaries relatively straightforward. However, it requires one to determine the safety critical event which is not easily defined in roadway departures on curves. As a result other surrogates have been utilized in the research of horizontal curves.

Vehicle lateral placement is one of the operating measures identified as a contributing factor to crash risk on horizontal and used quite extensively in the literature available on rural curve negotiation. In the section below studies which have utilized lateral placement as a surrogate on horizontal curves will be discussed.

1.2.3 Vehicle Path Trajectories and Lateral Position within curves

Previous research has been conducted to develop conceptual models of curve driving. These studies had looked at vehicle path trajectories as a means of evaluating the safety of highway alignments and determining how various factors and countermeasures affect safety. Lateral placement or lane position have been utilized in a majority of studies as a safety surrogate to assess the effectiveness of various countermeasures and safety at curves.

Radius and direction of curve were found to affect lateral position in the curve in studies which developed vehicle path trajectories. Additionally, it was found that most drivers tended to move towards the inside of the curve as they approached the center and therefore flattened the path in which they traveled.

Glennon et al. (1971) mounted a video camera to an observation box on the bed of a truck and used it to capture the path of a study vehicle it was following. Each curve studied was marked with strips at twenty foot intervals along the centerline. Five non-spiraled curves ranging from two to five degrees were traversed by approximately 100 vehicles. The lateral placement was used at the twenty foot intervals to calculate the instantaneous vehicle path radius. It was found that most vehicles will have a path radius that is less than the highway curve radius at some point in the curve.

Glennon et al. (1985) furthered the work conducted in ‘71 by evaluating lateral positon at six curves in Ohio and Illinois. Cameras were used to collect data in this study and used pavement reference markers 150 m upstream of the curve as well as at the PC and every 25 feet after. Results from the analysis indicated that drivers drifted towards the inside of the curve as they neared the center.

Spacek (1998) developed a model of curve negotiation behavior based on lateral position across seven points in a curve. The data were collected for two-lane roads for curves at least 200 meters from another curve or traffic control. Cameras were used to at collet data at a point upstream and downstream of curves as well as at five locations within a curve for 12 sites during off peak hours during daylight and with good weather.

Spline interpolation was used to develop six track profiles which were commonly observed in the field. The models disaggregated curve paths to normal behavior, common intentional lane deviations (cutting and swinging), and two profiles that indicated driver adjustments after misjudging a curve (drifting and correcting). The normal behavior found that drivers tended to drive more towards the inside of the lane, effectively flattening their paths. These paths are shown in Figure 1.1.

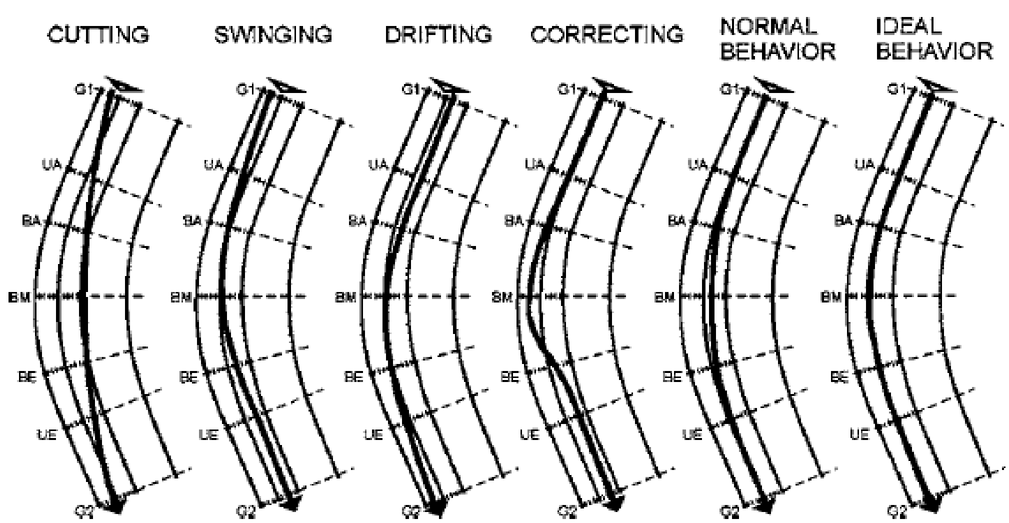


Figure 1.1 Models of Curve Negotiation Developed by Spacek (1998)

Felipe and Navin (1998) also evaluated lateral placement through curves using an instrumented vehicle along a two-lane mountainous road and found that vehicles mostly followed the center of the lane for both directions with large radii. With smaller radii, they found that drivers in both directions followed a flattened path to minimize speed change. They report that variation in path selection was a function of road geometry, surrounding traffic and the driver. They also found that drivers limited speed on curves with small radii based on comfortable lateral acceleration, which corresponded to 0.35 to 0.4g.

A study by Räsänen (2005) used a before and after analysis at a curve in Finland whose pavement markings were worn out and then replaced. Additionally two months after the initial repainting, centerline rumblestrip were also added. Unobtrusive video cameras were used to determine the lateral position through the curve. It was found that oncoming vehicles shifted drivers towards the shoulders by 15-20 cm. Results also indicated that the standard deviation of lateral position decreased from 35 cm to 28 cm with repainting of centerline and 24 cm after the rumble strips were added. Additionally, encroachments decreased from 7.3% to 4.2% and then with rumblestrips to 2.4%.

Levison et al. (2007) developed a driver vehicle module to use with the Interactive Highway Safety Design Model. One component of this model was path selection which assumes the drivers desired path profile is one where drivers drive the curve as if it had a larger radius than it does.

Gunay and Woodward (2007) collected data on traffic flow at five roundabout and three horizontal curve sites in Northern Ireland in 2005 using a camcorder that was hidden from sight as much as possible. Software was used to determine a vehicles lane position from the lane line. They found that on horizontal curves, driver path shifted towards the inside of the curve, with the shift increasing with decreasing radii.

Stodart and Donnell (2008) collected data upstream and within six curves using instrumented vehicles with 16 research participants during nighttime conditions. They used ordinary least squares regression and compared change in lateral position from the upstream tangent to the curve midpoint and found curve radius and curve direction had the largest effect on changes in lateral position between the tangent and midpoint of the curve.

Ben-Bassat and Shinar found similar findings in a study conducted in a driving simulator in 2011. 11 male and 11 female undergraduate students drove through a mixture of tangent and curved sections of differing radii with various shoulder widths and guardrail presence on divided four lane roads. They found as radii of curves decreased, drivers tended to deviate in their lane more than in large radii curves and tangent sections.

Most recently, Fitzsimmons et al. (2014) modeled vehicle trajectories using mixed effects models for a rural and an urban curve in Iowa. Pneumatic road tubes were used to collect lateral position of the vehicles at 5 points throughout each curve. Similar to the Spacek study, it was found that most vehicles tended to traverse the curve as if the radius was larger than the design radius of the curve and therefore tended to travel towards the inside of the curve as they approached the center. The study also found that time of day, direction of curve and vehicle type all affected lateral positon in the curve.

Campbell et al. (2012) also created a model of conceptual curve driving breaking the driving task through a curve into four areas (approach, curve discovery, entry and negotiation, and exit) which require different levels of attention and driving tasks as shown in Figure 1.2. Driving tasks during the approach include scanning for visual cues to locate the curve (i.e. signing), obtaining speed information from signing, and making initial speed adjustments. During this phase, visual demand is low and driver workload to maintain position is low. In curve discovery, drivers use visual and roadway cues (i.e. delineation) to determine the amount sharpness, assess roadway conditions, make necessary speed and steering adjustment to enter curve. At this point, driver workload is moderate but increases to just after the PC. Drivers at the entry and negotiation state use visual and roadway cues (i.e. chevrons) to adjust their speed based on curvature and steering to maintain safe lane position. The primary cues for a driver to adjust speed and position are lateral acceleration and vehicle handling. Driver visual demand and workload are high as drivers adjust speed and trajectory to stay within their lane with higher demands for curves with shorter radii and narrow lane width. At the exit point, drivers use visual and roadway cues (i.e. termination of chevrons) to adjust back to the tangent speed or prepare for negotiation of a subsequent curve. At this point visual demand is low and driver workload is moderate.

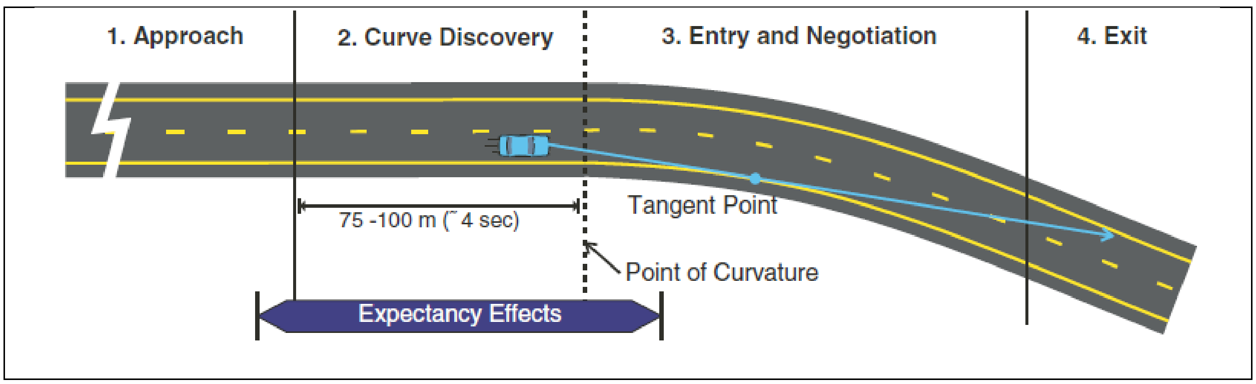


Figure 1.2 Curve Negotiation as Defined by Campbell et al. (2012)

1.2.4 Summary

The studies discussed in this section have provided information regarding what curve characteristics are most relevant and driver behaviors which contribute to crashes on curves, and which factors affect vehicle paths through curves; yet information is lacking. These studies in general have focused on looking at larger samples of traces across a small set of curves to determine how driver's behavior differs across those few curves. Having a limited sample size allows them to determine how drivers path varies based off roadway characteristics such as radius or things such as time of day. They do not however determine the general driving behavior of drivers on curves across various states and curve types and how driver behaviors such as glances and distraction affect negotiation. Having a better understanding of how drivers interact with various roadway feature and countermeasures in different environments in determining vehicle paths will provide information to decision makers in determining how to best allocate limited resources to reduce crashes on curves. The Strategic Highway Research Program 2 (SHPR 2) Naturalistic Driving Study (NDS) and Roadway Information Database (NDS) provide a unique dataset which allow for one to develop models which give insight into how the roadway, environment and driver interact when negotiating horizontal curves.

1.3 Problem statement

The objective of this research is to develop models which provide a better understanding of how drivers traverse curves looking at smaller samples of traces per curve over a larger sample of curves and drivers in order to gain insight into areas which lead to run off the road crashes and ways in which to mitigate these areas. The ultimate goal of this research is to help to reduce fatal crashes on our roads. Roadway departure crashes on curves account for a large percentage of the total fatal crashes, so by reducing these we can help reduce fatal crashes.

Countermeasures such as adding paved shoulders, installing chevrons or rumble strips have been found to help reduce crashes on horizontal curves. In order to be able to efficiently and effectively use countermeasures on horizontal curves, a better understanding of how they affect drivers' negotiation of curves based on roadway, environmental and driver factors so we can tailor the installation of each to situations where they will provide the best safety benefit. Additionally, by having a better understanding of how drivers traverse curves normally and situations which lead to lane departures, technologies that are developed or are being developed can be improved upon by the insight provided. These technologies provide potentially the greatest opportunity to reduce crashes as they remove or reduce the driver decision making. As driver error is a cause in the majority of crashes, removing the chance for driver error should lead to a reduction in crashes.

The models developed will help to address the three research questions outlined below.

1.3.1 Research Question 1: How do drivers normally negotiate a single isolated horizontal curve?

A conceptual model of curve driving will be developed to assess changes in metrics as the driver negotiates the curve. Understanding how a driver normally negotiates a curve provides insight not only into how characteristics of the roadway, driver, and environment influence driving behavior, but also into areas that can lead to roadway departures. Knowing how much drivers normally deviate in their lane as well as how they choose their speed could potentially have implications on policy or design.

A conceptual model will be developed based off past work for isolated curves only (i.e. curves with at least 300 meters between them). The models that were previously modeled differed slightly in approach, but had similar findings. Radius and direction of curve were found to affect lateral position in the curve and models were developed to look at changes in lateral position between upstream and center of the curve or at points (five to seven) within the curve (Spacek 1998, Felipe and Navin, 1998, Stodart and Donnell 2008, Fitzsimmons et al. 2013). These previously developed models of rural curve driving have taken into account roadway, environmental, and to a limited extent driver factors yet none have taken into account driver behavior and how distraction can affect lateral position. This study expands on these previous models by also including additional driver and environmental factors.

A model will be developed for the inside or right curve and outside/left curve to determine lateral position throughout the curve as at points as a driver negotiates their way through using the NDS and RID data. Vehicle offset from the center of the lane will be used as the dependent variable in the model. Key factors which will be used in the analysis include:

- Roadway factors: Curve Radius, length of curve, superelevation, distance between curves, presence of countermeasures (i.e. chevrons, rumble strips, raised pavement markings, curve advisory signs), direction of the curve, and the speed limit upstream and within the curve
- Environmental factors: Time of day, surface condition (wet, dry, snow), pavement condition, lane marking condition, the visibility, if driver is following another vehicle, if driver is passing other vehicles
- Driver factors: age, sex, distractions, glance location, and vehicle type

1.3.2 Research Question 2: How do drivers negotiate horizontal curves?

The second objective of this research is to expand the work from Research Question 1 to include other horizontal curves such as S-curves or other non-isolated curves. Additional data will be incorporated which may strengthen the models and allow for random effects to be captured and results to be applicable to more situations. Additional variables on whether the curve is an S-curve and if so which curve (first encountered or second encountered) will also be included in the analysis. If enough instances of lane departure are present they will also be incorporated into the model to determine how curve negotiation changes in cases of lane departure.

1.3.3 Research Question 3: Which factors increase the likelihood of a lane departure?

The third objective of this research is to develop a model which will determine which driver, roadway and environmental factors affect the probability of a lane departure. This will be accomplished by using the baseline NDS data along with data in which lane departures occur. The following factors will be explored in the analysis:

- <u>common roadway characteristics</u>: radius of curve, length of curve, superelevation, direction of curve, upstream and curve advisory (if present) speed limits, countermeasures(i.e. rumblestrips, chevrons, RPMS, guardrail)
- <u>kinematic driving factors:</u> driver's glance locations, presence of distractions, vehicle offset, speed and acceleration upstream and at various points in the curve
- <u>traditional environmental factors:</u> time of day, weather conditions, and visibility
- <u>exposure factors:</u> presence of oncoming vehicles, if driver is following another vehicle

Additionally, if any kinematic factors are included in the model, an attempt to develop additional models that predict these values based off upstream driving conditions will be developed. These will provide a means of predicting probability of the lane departure upstream from the driver entering the curve thereby leaving time to warn drivers of the potential for the lane departure.

1.4 Study limitations

The author would like to note early on that there were a few major limitation of the research due to the fact that it was being conducted while the NDS and RID data collection were taking place. Among these are data accuracy issues, limited sample size, and use of surrogates.

Data accuracy issues included significant noise being present in variables such as offset, which is expected for large-scale data collection of this nature. It was also due to issues with the machine learning algorithm used in the DAS which depends on lane lines or differences in contrast between the roadway edge and shoulder in order to establish the position. When discontinuities in lane lines occur, offset is reported with less accuracy. Discontinuities occur due to lane lines being obscured or not visible, natural breaks being present in lane lines (e.g., turn lanes, intersections), or visibility being compromised in the forward roadway view. A moving average used to smooth the data helped to reduce some noise, but could not account for large distances of not accurate lane lines. Additionally it should be noted that the fact that offset data were more accurate for highly visible lane lines may lead to some inherent bias in our data samples, which could be addressed with larger samples sizes to include a more equal distribution of highly visible, visible and obscured lane lines.

In other cases, variables of interest were not sufficiently available to be utilized. For instance steering wheel variability would have been helpful for looking at driver's reaction or drowsiness, but was not available for a majority of the data provided. Additionally, although a passive alcohol detector was present, at the time data were collected it did not appear to be reliable enough to identify potential intoxicated drivers. Radar data were also included in the data, but QA/QC had not been conducted, so it could not be included in the analysis.

Additionally, the quality of the driver face video was not always clear enough to be able to see the pupil. This especially occurred at night and when the driver was wearing sunglasses. In these cases driver's head position was used to measure approximate glance location, which may have led to missing some of the more subtle glances such as looking at the rear-view mirror or at the steering wheel. These traces were still included in order to have an adequate sample size and to be able to include night driving as it was thought that missing these subtle glances would not significantly alter the results.

Sample size limitations were due to only one third of the data being available, as well as time and budget constraints limited how much data could be reduced (specifically driver glance data). Accuracy issues with the offset variable, which were described previously, also significantly reduced the samples for these studies as accurate offset was required. Approximately 10% of the data reduced had accurate enough offset to be included in the analysis. The limited sample size also limited the amount of driver and roadway characteristic which could be included. For instance while a large sample of curves with rumblestrips were requested, only two curves which we had reduced data for had rumblestrips. Having a larger sample size would have helped to answer questions that had hoped to be answered in the course of the study but were unable to be determined. For instance with enough data it is thought that the effect of countermeasures such as rumblestrips or chevrons could be determined.

Finally, as crash and near crash data were not available at the time the data for these studies was collected, the use of surrogates was required for the analysis. While surrogates provide some expected correlation with crashes, the exact relationship was not able to be established. Therefore the results of the research cannot be translated to risks of crashes, but to risks of lane encroachments. Having adequate data on the crashes and near crashes would allow one to develop this relationship.

1.5 Study implications

These conceptual models, which will be among the first developed using the SHRP 2 NDS, will advance understanding by providing valuable insight into the interaction and effect that roadway attributes and countermeasures (i.e. chevrons, pavement markings, rumblestrips), driver behaviors and attributes (i.e. distraction, speed and age), and environmental factors (i.e. day vs night or low visibility) have on drivers lateral lane position throughout a curve. It will also provide information on how drivers typically traverse curves. The results of these models can be used by States in developing their performance measures and performance targets in their Strategic Highway Safety Plans by helping to select countermeasures more appropriately and provide areas to target education.

The predictive lane departure model will help gain insight into which driver behaviors are safety critical. The model may also provide data to include in lane departure warning systems or curve speed warning technologies that have not previously been included. Most current lane departure warning systems utilize cameras which track the lane line along with algorithms which predict the likelihood of a lane departure. The model developed as part of question 3 may provide information on how roadway features and driver behavior in the upstream affect the probability of a lane departure and could predict before even entering the curve if the driver is likely to depart their lane in that curve. The long-term impact of these technologies being in passenger cars is that they could result in a large decrease in lane departure resulting in crashes as it takes away opportunities for driver error in deciding their risk of a lane departure.

1.6 Organization of the Dissertation

This dissertation contains five chapters. Chapter 1 introduced the problem of lane departures on rural curves. It also contained the review of existing literature related to curve negotiation and risks associated with lane departures. Chapter 2 addresses research question 1. The development of a conceptual model of rural curve driving on isolated rural curves using the SHRP 2 NDS is represented in this chapter, Chapter 3 expanded on the work conducted in Chapter 2 to include a larger sample size of curves and drivers as well as traces where lane encroachments occur. Chapter 4 presents results of a study that used a slightly expanded data set from chapter 3 to develop a model to predict the likelihood of lane encroachments as well as models to predict input variables to this model. This chapter address research question 3. For the papers contained in Chapters 2-4, Nicole served as the main author and performed the major analysis. The additional authors provided additional expertise in determining and conducting the data reduction process, the statistics to use, and the method for the driver kinematic data reduction. Chapter 5 provides conclusions and main contributions of this dissertation, limitations of the studies and recommendations for future research.

1.7 Additional Contributions

In addition to the work presented in the dissertation, additional contributions were made on the same topic. One of these contributions was second author on an official SHRP 2 report that was peer-reviewed multiple times by a variety of reviewers. The work done as part of this SHRP 2 project has been presented multiple times across the country as well as internationally. Additionally, a paper was accepted to the Journal of Safety Research which will be published in the near future in which I am an author.

1.8 References

AASHTO. *Driving Down Lane-Departure Crashes: A National Priority*. American Association of State Highway and Transportation Officials, Washington, D.C., 2008.

Antin, J. Technical Coordination & Quality Control (S06). Presented at the 8th SHRP 2 Safety Research Symposium, Washington D.C., 2013.

Ben-Bassat, T., and D. Shinar. Effects of shoulder width, guardrail and roadway geometry on driver perception and behavior. *Accident Analysis and Prevention*. Vol 43. Issue 6, 2011, pp 2142-2152.

Campbell, J.L., M.G. Lickty, J.L. Brown, C.M. Richard, J.S. Graving, J. Grahm, M. O'Laughlin, D. Torbic, and D. Harwood. Chapter 6: Curves (Horizontal Alignment). NCHRP Report 600. Human Factors Guidelines for Road Systems, Second Edition. Transportation Research Board of the National Academies. Washington DC. 2012.

Caliendo, C., M, Guida, and A. Parisi. A crash-prediction model for multilane roads. *Accident Analysis and Prevention*. Vol 39. Issue 4, 2007, pp 657-670.

Charlton, S.G. The role of attention in horizontal curves: A comparison of advanced warning, delineation, and road marking treatments. *Accident Analysis and Prevention*. Vol 39, Issue 5. 2007, pp 873-885.

Charlton, S.G., and J.J. DePont. *Curve Speed Management*. Land Transport New Zealand Research Report 323. Land Transport New Zealand, Wellington, New Zealand, 2007.

Council, F.M. Safety Benefits of Spiral Transitions on Horizontal Curves on Two-Lane Rural Roads. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1635, 1998, pp. 10–17.

Davis, G., S. Davuluri, and J. Ping Pei. *A Case Control Study of Speed and Crash Risk, Technical Report 3, Speed as a Risk Factor in Run-off Road Crashes*. Center for Transportation Studies. Minnesota, 2006.

Findley, D.J., J.E. Hummer, W. Rasdorf, C.V. Zegeer and T.J. Fowler. Modeling the impact of spatial relationships on horizontal curve safety. *Accident Analysis and Prevention*. Vol 45, Issue 0, 2012, pp 296-304.

Felipe, E. and F. Navin. Automobiles on Horizontal Curves: Experiments and Observations. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1628, 1998, pp. 50–56.

FHWA. *FHWA Roadway Departure (RwD) Strategic Plan*. Federal Highway Administration, Washington D.C., 2013.

Fink, K.L., and R.A. Krammes. Tangent Length and Sight Distance Effects on Accident Rates at Horizontal Curves on Rural Two-Lane Highways. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1500, 1995, pp. 162–168.

Fitzsimmons, E., V. Kvam, R.R. Souleyrette, S.S. Nambisan, and D.G. Bonett. Determining Vehicle Operating Speed and Lateral Position along Horizontal Curves Using Linear Mixed-Effects Models. *Traffic Injury Prevention*, Vol. 14, Issue, 3, 2013, pp. 309-321.

Glennon, J.C., and G.D. Weaver. *The Relationship of Vehicle Paths to Highway Curve Design*. Texas Transportation Institute. Texas. 1971.

Glennon, J.C., T.R. Neuman, and J.E. Leisch. *Safety and Operational Considerations for Design of Rural Highway Curves.* Report FHWA/RD-86-035. Federal Highway Administration, Washington, D.C., 1985.

Gunay, B. and D. Woodward. Lateral Position of traffic negotiating horizontal bends. *Transport 160.* Issue TRI, 2007, pp 1-11.

Hallmark, S.L., N. Oneyear, S. Tyner, B. Wang, C. Carney and C. McGehee. *SHRP 2 S08D: Analysis of the SHRP 2 Naturalistic Driving Study Data*. Strategic Highway Research Program 2, Transportation Research Board, Washington, D.C., 2015a.

Hallmark, S.L., S. Tyner, N. Oneyear, C. Carney, and D. McGehee. *Evaluation of Driving Behavior on Rural 2-Lane Curves using the SHRP 2 Naturalistic Driving Study Data.* Journal of Safety Research. Manuscript accepted for publication, 2015b.

Lamm, R., E. M. Choueiri, J.C. Hayward, and A. Paluri. Possible Design Procedure to Promote Design Consistency in Highway Geometric Design on Two-Lane Rural Roads. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1195, 1988, pp. 111–122.

Levison, W.H., J.L. Campbell, K. Kludt, A.C. Vittner Jr., I. Potts, D. Harwood, J. Hutton, D. Gilmore, J.G. Howe, J.P. Chrstos, R.W. Allen, B. Kantowitz, T. Robbins, and C. Schreiner. Development of a Driver Vehicle Module (DVM) for the Interactive Highway Safety Design Model (IHSDM). Federal Highway Administration. Report FHWA-HRT-08-019. November 2007.

McGwin G., and D.B. Brown. Characteristics of traffic crashes among young, middle-aged and older drivers. *Accident Analysis and Prevention*. Vol 31. Issue 3, 1995, pp 181-198.

McLaughlin, S.B., J.M. Hankey, S.G. Klauer, and T.A. Dingus. *Contributing Factors to Run-Off-Road Crashes and Near-Crashes.* Report DOT HS 811 079. National Highway Traffic Safety Administration, Washington, D.C., 2009.

Miaou, S.-P., and H. Lum. Statistical Evaluation of the Effects of Highway Geometric Design on Truck Accident Involvements. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1407, 1993, pp. 11–24

.

Milton, J. and F. Mannering. The Relationship among Highway Geometric, Traffic-Related Elements, and Motor-Vehicle Accident Frequencies. *Transportation*, Vol. 25, 1998, pp. 395–413.

Neuman, T.R., R. Pfefer, K.L. Slack, K.K. Hardy, F. Council, H. McGee, L. Prothe, and K. Eccles. Guidance for Implementation of AASHTO Strategic Highway Safety Plan, Volume 6: A Guide for Addressing Run-Off-Road Collisions. NCHRP Report 500. Transportation Research Board of the National Academies, Washington DC. 2003.

Preston, H. *Low-Cost Treatments for Horizontal Curve Safety*. PowerPoint presentation from FHWA webinar, 2009.

Räsänen, M. Effects of a rumble strip barrier line on lane keeping in a curve. *Accident Analysis and Prevention.* Vol 37, Issue 3, 2005, pp 575-581.

Shively, T.S., K. Kockelman, and P. Damien. A Bayesian semi-parametric model to estimate relationships between crash counts and roadway characteristics. *Transportation Research Part B.* Vol 44, pp. 699-715. 2010.

Smadi, O. SHRP 2 S-04A Roadway Information Database Development and Technical Coordination and Quality Assurance of the Mobile Data Collection Project. Presented at the 7th SHRP 2 Safety Research Symposium, Washington D.C., 2012.

Spacek, P. FahrverhaltenunUnfallgeschehen in Kurven, Fahrverhalten in Kurvenbereichen. InStitutFürVerkehrsplanung, Transporttechnick, Strassen- unEisenbahnbaur, ETH Zürich, 1998.

Stodart, B.P. and E.T. Donnell. Speed and Lateral Position Models from Controlled Nighttime Driving Experiment. *ASCE Journal of Transportation Engineering*, Vol. 134, No. 11, 2008, pp. 439–449.

Tarko, A., Davis, G., Saunier, N., Sayed, T., & Washington, S. *Surrogate measures of safety*. TRB Annual Meeting White Paper. Washington, DC. 2009.

VTTI. *InSight Data Access Website SHRP 2 Naturalistic Driving Study*. https://insight.shrp2nds.us/. Accessed July 31st, 2014.

Zegeer, C.V., J.R. Stewart, F.M. Council, and D.W. Reinfurt. *Safety Effects of Geometric Improvements on Horizontal Curves*, University of North Carolina, Chapel Hill, NC, 1991.

CHAPTER 2: DEVELOPMENT OF A CONCEPTUAL MODEL OF CURVE DRIVING FOR ISOLATED RURAL TWO LANE CURVES USING SHRP 2 NATURALISTIC DRIVING DATA

Modified from a paper to be published in the conference proceedings of the *5th International Symposium on Highway Geometric Design*

Nicole Oneyear, Shauna Hallmark, Samantha Tyner, Daniel McGehee and Cher Carney

Abstract

Approximately 27% of all fatalities in 2008 occurred on horizontal curves. Of these, over 80% were run off the road crashes, with the majority of these fatal crashes occurring on rural two lane highways. Consequently, run off the road crashes on rural highway curves present a significant safety concern. Therefore addressing lane-departure crashes on rural curves is a priority for National, State, and local roadway agencies. Much research has been conducted to look at how roadway factors, such as radius and shoulder width and environmental factors, such as weather affect crashes, yet limited research has been conducted looking at how driver behaviors affect crash risk.

This paper utilizes data from the SHRP 2 Naturalistic Driving Study (NDS) and Roadway Information Datasets (RID) to present interim results on the develop a conceptual model of normal curve driving on isolated rural two lane curves that explores how drivers interact with the roadway environment. This includes driver, roadway, and to limited extent environmental conditions. The model helps identify zones where driver are more likely to have lane departures.

Times series data, at the level of 0.1 second were used as the data input. Models were developed using generalized least squares with offset of the center of the vehicle from the center of the lane as the dependent variable. Models for both inside (right-hand curve from the perspective of the driver) and outside (left-hand curve from the perspective of the driver), were developed. Results indicate that lane position within the curve is influenced by lane position

upstream of the curve, drivers glancing down, age, shoulder width, pavement delineation, presence of curve advisory signs, as well as distance into the curve.

2.1 Introduction

Approximately 27% of all fatalities in 2008 occurred on horizontal curves. Of these, over 80% were run off the road crashes, with the majority of these fatal crashes occurring on rural two lane highways *(1)*. Additionally, research has found that the crash rate on curves is approximately three times the rate on tangent sections *(2)*. Consequently, run off the road crashes on rural horizontal curves present a significant safety concern.

The objective of this paper was to understand how a driver negotiates a curve normally. Normal driving is defined as no lane line crossings, crashes, or conflicts. This was done by developing a conceptual model of curve driving on rural two lane curves utilizing the SHRP 2 Naturalistic Driving Study (NDS) and Roadway Information Database (RID).

A better understanding of the interaction between driver characteristics and curve negotiation needs can potentially lead to better design and application of countermeasures. For instance, if older drivers have the hardest time with curve negotiation because they are less likely to see visual cues, the best solution might be larger chevrons. On the other hand, a solution geared towards younger drivers might include more closely spaced chevrons to help drivers gauge the sharpness of the curve. Distracted drivers would perhaps require another solution, such as a tactile cue from transverse rumble strips.

Studies of roadway factors, such as degree of curve *(3,4,5,6)*, presence of spirals *(7)*, or shoulder width and type *(8)*, have provided some information regarding the most relevant curve characteristics, but information is still lacking. In addition, little information is available that identifies driver behaviors that contribute to curve crashes. As a result, a better understanding of how drivers interact with various roadway features and countermeasures may provide valuable information to highway agencies for determining how resources can best be allocated in order to prevent potential lane departures and reduce crashes.

2.1.1 Background on SHRP 2 Naturalistic Driving Study

The SHRP 2 NDS is the largest naturalistic driving study to date. The study was conducted by Virginia Tech Transportation Institute (VTTI). Drivers in six states (Florida, Indiana, New York, North Carolina, Pennsylvania and Washington) had their vehicles equipped with a Data Acquisition System (DAS) which collects information such as speed, acceleration, and GPS data, as well as four cameras which collected forward, rear, drivers face and over the shoulder video. These equipment captured all of the trips a driver made over a period of six months up to two years. Males and females ages 16 to 98 participated in the study. Over the three years of the study approximately 3,300 participants drove over 30 million data miles over 5 million trips *(9,10)*.

2.1.2 Background on SHRP 2 Roadway Information Database

In conjunction with the SHRP 2 Naturalistic Driving Study, another project was conducted to collect roadway information for the main roads traveled in the NDS. The Center for Research and Education (CTRE) led the effort which used mobile data collection to collect 12,500 centerline miles of data across the six states where the NDS was focused. Data collected included information on roadway alignment, signing, lighting, intersection location and types, presence of rumblestrips and other countermeasures. In addition to the mobile data collection effort, existing roadway data collected by local agencies was leveraged to increase the data available. Additionally, supplemental data such as crash data, changes to laws, and construction projects were also collected to further strengthen the database *(11)*.

2.2 Previous Research

Limited research has been conducted to develop conceptual models of curve driving. Models developed differed slightly in approach, but had similar findings. Radius and direction of curve were found to affect lateral position in the curve. Additionally, it was found that most drivers tended to move towards the inside of the curve as they approached the center and therefore flattened the path in which they traveled. The approaches of five models are discussed in further detail.

Spacek (1998) developed a model of curve negotiation behavior based on lateral position across seven points in a curve. Spline interpolation was used to develop six track profiles which were commonly observed in the field. The models disaggregated curve paths to normal behavior, common intentional lane deviations (cutting and swinging), and two profiles that indicated driver adjustments after misjudging a curve (drifting and correcting). The normal behavior found that drivers tended to drive more towards the inside of the lane, effectively flattening their paths *(12)*.

Felipe and Navin (1998) also evaluated lateral placement through curves using an instrumented vehicle along a two-lane mountainous road and found that vehicles mostly followed the center of the lane for both directions with large radii. With smaller radii, they found that drivers in both directions followed a flattened path to minimize speed change. They report that variation in path selection was a function of road geometry, surrounding traffic and the driver *(3)*.

Stodart and Donnell (2008) collected data upstream and within six curves using instrumented vehicles with 16 research participants during nighttime conditions. They used ordinary least squares regression and compared change in lateral position from the upstream tangent to the curve midpoint and found curve radius and curve direction had the largest effect on changes in lateral position between the tangent and midpoint of the curve *(4)*.

Fitzsimmons et al (2014) modeled vehicle trajectories using mixed effects models for a rural and an urban curve in Iowa. Pneumatic road tubes were used to collect lateral position of the vehicles in 5 points throughout each curve. Similar to the Spacek study*(12)*, it was found that most vehicles tended to traverse the curve as if the radius was larger than the design radius of the curve and therefore tended to travel towards the inside of the curve as they approached the center. The study also found that time of day, direction of curve and vehicle type all affected lateral positon in the curve *(13).*

Levison et al. (2007) developed a driver vehicle module to use with the Interactive Highway Safety Design Model. One component of this model was path selection and was assumes the drivers desired path profile is one that drivers the curve as if it had a larger radius than it does *(14)*.

Previously developed models of driving on rural curves have taken into account roadway, environmental, and to a limited extent driver factors yet none have not taken into account driver behavior and how distraction can affect lateral position. This papers hopes to expand on these previous models by also including additional driver and environmental data as well as studying a larger number of curves.

2.3 Methodology

Data were acquired from two main sources, unless noted otherwise. These were the SHRP 2 Naturalistic Driving Study (NDS) and the SHRP 2 Roadway Information Database (RID). The NDS included time series data collected through a data acquisition system (DAS), as well as video data collected from 4 cameras placed in the vehicle which captured the forward view, rear view, driver's face and over the shoulder. As the driver's face and over the shoulder video contained potentially identifying information, these data were viewed and information reduced at the secure enclave housed at VTTI.

2.3.1 Identification of Curves of Interest

At the time this project was conducted, the NDS and RID had not been linked. As a result, the team manually identified curves of interest and then requested any trips on these curves from the NDS. To identify potential curves of interest, the project team made use of weighted trip maps. VTTI prepared trip maps used a subset of trip data in the early stages of the NDS data collection. Trips were overlain with a roadway database and showed an estimate of where trips were likely to have occurred. The trip maps were overlain with the RID and rural 2-lane curves on paved roadways were identified. A one-half mile tangent section upstream and downstream of each curve was also selected. Curves were identified in all states except for Washington since much of the roadway mileage was urban.

A spatial buffer (polygon) was created around each curve. In some cases curves were located near one another and multiple curves were included in a single buffer. The buffers were provided to VTTI and were overlain with the NDS. If a trip fell within a buffer and met certain criteria (i.e. GPS data present, speed data present, etc.) then it became a potential event (one trip through one buffer) to use in the analysis. At the time of the data request, around one-third of the NDS data had been processed and were available. The initial query resulted in around 4,000 traces (one trip through one buffer). Each trace was reviewed and traces where a needed variable was not present or reliable were removed from further consideration. Once these traces were removed, a total of 987 events across 148 curves were selected to represent a good cross-section of curve and driver characteristics. Further details on how the data were requested can be seen in the SHRP2 S08D Final report *(15)*.

2.3.2 Data Collection and Data Reduction

2.3.2.1Roadway Variables

Roadway variables were extracted for the 148 curves using the RID data when available. In some cases a variable was not collected, and in other cases the RID was not available for the study segment because the RID did not cover all roads in the NDS. When the information was not available through the RID, other sources were used to manually extract the data. These additional sources were also used to confirm data collected through the RID, such as speed limit and advisory speed limit.

ArcGIS was used to measure distances between curves using the PC included in the RID. ArcGIS was also used to determine whether the curve was an S-curve or a compound curve based on the distance between curves and direction of curves.

Google Earth was used to extract the roadway features not included in the RID. It was also used to collect countermeasures before the forward video was available, such as chevrons and RPMs, which were later confirmed with the NDS forward video. Radius was provided for most curves in the RID and was reported as radius by lane. When RID data were not available, which only included a few curves in Florida, radius was measured using aerial imagery and the chord-offset method. This method was verified using curves with known radii. NDS forward video was used to determine subject measures for delineation, pavement condition, roadway lighting, and roadway furniture (which describes objects around the road that provide some measure of clutter). Variables collected are shown in Table 2.1.

Table 2.1 Roadway Variables Extracted and Main Source

Feature	ArcGIS	SHRP2 RID	Google Earth	SHRP 2 NDS Forward Video
Curve radius		✓		
Distance between curves	✓			
Type of curve (isolated, S, compound)	✓			
Curve length	✓	✓		
Super elevation		✓		
Presence of rumble strips		✓		✓
Presence of chevrons		✓	✓	✓
Presence of w1-6 signs			✓	✓
Presence of paved shoulders		✓		✓
Presence of raise pavement markings (rpm)			✓	✓
Presence of guardrail			✓	✓
Speed limit		✓		
Advisory sign speed limit		✓	✓	✓
Curve advisory sign/W1-6		✓	✓	✓
Pavement condition				✓
Delineation				✓
Sight distance				✓
Roadway furniture				✓
Direction of curve				✓
Shoulder width and type		✓		

2.3.2.2 Vehicle, Traffic, Static Driver and Environmental Variables

Each of the traces or events represents one driver trip through a selected roadway segment. One spreadsheet (containing DAS data), one forward video, and one rearview video were provided by VTTI for each trace. Each row of data represents 0.1 seconds, and spatial location was provided at one-second intervals. A time stamp was also provided to link the various videos with the DAS data. A list of the main DAS variables provided and used in the analysis include the following:

- Acceleration, x-axis: vehicle acceleration in the longitudinal direction vs. time
- Acceleration, y-axis: vehicle acceleration in the lateral direction vs. time
- Lane markings, probability, left/right: Probability that vehicle based machine vision lane marking evaluation is providing correct data for the left/right side lane markings

- Lane position offset in meters: Distance to the left or right of the center of the lane based on machine vision
- Lane width (m): Distance between the inside edge of the innermost lane marking to the left and right of the vehicle
- Spatial position: Latitude and Longitude
- Speed : Vehicle speed indicated on speedometer collected from network
- Timestamp Integer used to identify one time sample of data. Arbitrary counter that is unique for each data row in each file. Used by the community viewer.
- Yaw rate, z-axis: Vehicle angular velocity around the vertical axis.

Vehicles traces were overlain with the RID curve, the nearest GPS points to the PC or PT was found and the position of the PC/PT was located within the time series data using interpolation. Once PC/PT were established, vehicle position upstream or downstream of the curve was calculated using speed. For some traces, there were multiple curves, so the PC/PT and upstream/downstream distances were determined for each curve. In some cases, speed was missing for multiple time stamps. In these cases, speed was interpolated assuming a constant increase or decrease.

The static driver and vehicle characteristics were merged with each trace. The characteristics used include driver age and gender and vehicle class and track width.

The forward video was used to reduce the environmental and other variables. The variables collected included the following:

- Surface condition (i.e., dry, wet, snow, etc.)
- Lighting conditions (i.e., day, dawn, dusk, night with no lighting, night with lighting)
- Visibility (i.e. high visibility (clear), low visibility (foggy))

- Locations of vehicles in the opposite direction passing the driver's vehicle
- Locations where the driver's vehicle was following another car
- Presence of curve advisory signs
- Presence of chevrons

2.3.2.3 Kinematic Driver Characteristics

Driver attention was measured by the location where a driver was focused for each sampling interval. Scan position, or eye movement, has been used by several researchers to gather and process information about how drivers negotiate curves *(16*). The majority of studies have used simulators to collect eye tracking information. Because eye tracking is not possible with NDS data, glance location was used as a proxy. Glance locations, represent practical areas of glance locations for manual eye glance data reduction. Glance locations were coded using the camera view of the driver's face, with a focus on eye movements, but taking into consideration head tilt when necessary. Glances were coded as one of 11 potential locations which can be seen below:

- Front
- Left
- Right
- Down
- Steering Wheel
- Center Console
- Rearview Mirror
- Up
- Over the Shoulder
- Missing (due to glare or problems with camera)
- Other Glance

Potential distractions were determined by examining both the view of the driver's face and the view over the driver's right shoulder, which showed hands on/off the steering wheel.

Distractions were identified when drivers took their eyes off the forward roadway. Potential distractions include the following:

- Route planning (locating, viewing, or operating)
- Moving or dropped object in vehicle
- Cell phone (locating, viewing, operating)
- IPod/MP3 (locating, viewing, operating)
- Personal hygiene (i.e. makeup application, brushing hair, etc.)
- Passenger
- Animal/insect in vehicle
- In-vehicle controls
- Drinking/eating
- Smoking

Glance location and distractions were coded for 200 meters upstream and throughout each curve for only 515 of the events due to time constraints. Glance location and distractions were manually merged with the event files using time stamp as a reference. Once this was completed, glance location was indicated for each row in the DAS event file.

There were times in the manual reduction of the glance and distraction reduction when eye movements were obscured due to such things as glare, the driver wearing sunglasses, nighttime. When this occurred, head movement was used to estimate glance. This may have caused minor glances, such as at the steering wheel to have been missed. It should be noted that glance and distraction were more likely to have been accurately coded for traces with clearer views of the face and eyes. However, discarding data where head movements were used instead of eye movements would have entailed removing almost all nighttime data and significantly reducing sample size.

Glance location was further reduced to indicate time spent in “eyes-off-roadway” engaged in roadway-related tasks or “eyes-off-roadway” engaged in non-roadway-related tasks based on data coding used by Angell et al. (2006). The authors define roadway-related glances or situation awareness (SA) as glances to any mirror or speedometer. Glances to other locations are defined as not roadway-related (NR). Roadway-related glances (SA) included left mirror, steering wheel, and rear-view mirror *(17)*.

It was not possible to distinguish between a glance to the right mirror and a glance to the right for other reasons (e.g., to converse with passenger). Additionally, on a two-lane roadway, glances to the right mirror are not likely to be as common because drivers are not expecting vehicles to the right. Consequently, all glances to the right were considered to be non-roadway-related.

Additionally, when glances to roadway-related locations were also associated with a distraction, it was decided that these glances were likely to be non-roadway-related. For instance, a driver who was texting and glancing at the steering wheel was likely to be looking at the cell phone rather than the speedometer. As a result, non-roadway-related glances included center console, up, right, or down.

2.3.2.4 Data smoothing

Smoothing of the DAS data was necessary because a certain amount of noise in the data resulted in improbable data points. These points would be data points that would jump for 0.1 seconds out of a range of what was probable and then continue following the previously seen trend. Several different methods to smooth the data were investigated. The Kalman filter estimates the optimum average factor for each subsequent state using information from past states. It was determined that, although the Kalman filter was appropriate, developing a model for multiple variables for over all of the vehicle traces was overly complicated and time consuming.

A moving average method was selected because it is able to reduce random noise while retaining a sharp step response. Each of the variables listed above was smoothed over 5 data points (0.5 second) using a moving average method. This method involved averaging the data from the 0.2 seconds before the point of interest, the 0.1 second of interest and the 0.2 seconds after the point of interest.

2.3.3 Data Sampling

The sampling plan for the curve model can be seen in Figure 2.1. Data were sampled at each point shown (e.g., PC), and locations for sampling were determined after consulting previous research *(12,13)* as well as plotting events and determining which sampling scheme picked up common patterns. Sampling in the tangent section was based on distance. Sampling within the curve was at equidistant points rather than at a specified distance because the curves have varying lengths.

The points sampled within the curve were the PC, PT, and then five equally spaced points (C2, C3, CC (curve center), C4, and C5), as shown in Figure 2.1. Upstream data were collected every 50 meters up to 300 meters. These locations were chosen in order to capture driving upstream of where drivers react to the curve (i.e., normal tangent driving) along with the reaction and approach areas. Because the data sampling plan required 300 meters of upstream data, the analysis only included isolated curves (i.e., no S-curves or compound curves) and only included curves with a tangent section that was at least 300 meters from the nearest upstream curve.

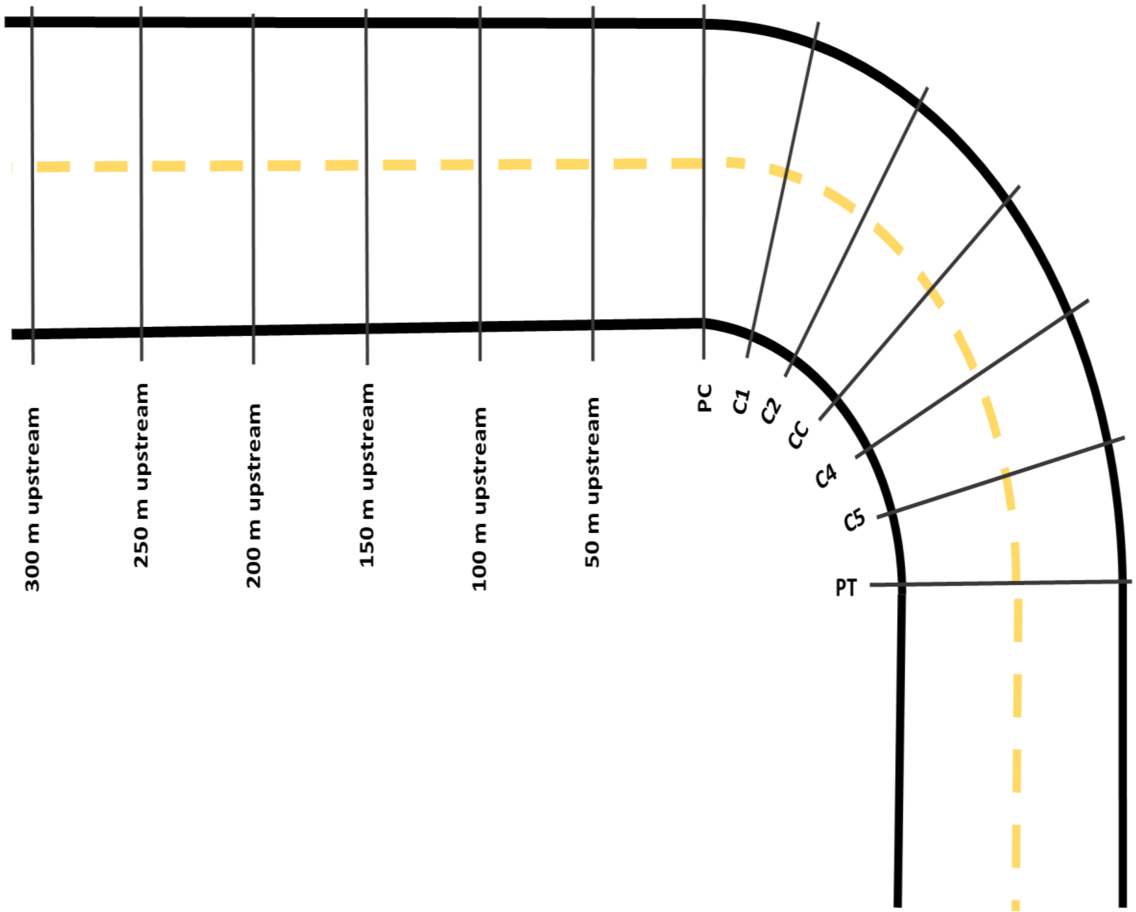


Figure 2.1 Data Sampling Layout for Curve Driving Model for Right-Handed Curve

The DAS and distraction data described previously were sampled at each point in the curve shown. Data collected for the upstream area included the offset and speed at each sample point, along with driver glance location and distractions. These data were merged with environmental, driver, and vehicle data. The summary statistics for the variables used in the final models are listed in Table 2.2, with the offset for the sampled points in the curve being presented separately as they are utilized in the model through the position in curve indicators. A complete list of variables collected, calculated and attempted in the model analysis are included in Table 2.3. For some of the variables, (i.e. surface) only those conditions which were present in the data were included. Therefore since none of the samples occurred when it was currently raining, that was not included as a condition. In other cases groupings were decided based on the samples available. While looking at the difference between a four foot shoulder and an eight foot shoulder would be helpful, not enough data were available to be able to look at this.

Table 2.2 Summary Statistics for Select Variables

Right-handed curves (inside)		
Variable	**Description**	**Mean (std dev) or %**
Offset 100	Distance offset from centerline 100m upstream of curve (m)	-0.01923 (0.34589)
Offset at PC	Distance offset from centerline at PC (m)	-0.04291(0.29099)
Offset at C1	Distance offset from centerline at C1 (m)	0.09273 (0.23994)
Offset at C2	Distance offset from centerline at C2 (m)	0.15614 (0.28088)
Offset at CC	Distance offset from centerline at CC (m)	0.22914 (0.28923)
Offset at C4	Distance offset from centerline at C4 (m)	0.32480 (0.29694)
Offset at C5	Distance offset from centerline at C5 (m)	0.14434 (0.32050)
Offset at PT	Distance offset from centerline at PT (m)	0.10612 (0.27209)
Down	Indicator that driver is glancing down (0: glance not down, 1: glance is down)	1.4%
Under 30	Indicator that driver is under 30 years old (0:30 and over, 1: under 30	18.75%
Curve Advisory Sign	Indicator for presence of curve advisory sign (0: not present, 1: present)	6.67%
Left-handed curves (outside)		
Variable	**Description**	**Mean (std dev) or %**
Offset 100	Distance offset from centerline 100m upstream of curve (m)	-0.05389 (0.25358)
Offset at PC	Distance offset from centerline at PC (m)	-0.2168 (0.31088)
Offset at C1	Distance offset from centerline at C1 (m)	-0.14853 (0.36442)
Offset at C2	Distance offset from centerline at C2 (m)	-0.21050 (0.25853)
Offset at CC	Distance offset from centerline at CC (m)	-0.28222 (0.24645)
Offset at C4	Distance offset from centerline at C4 (m)	-0.15048 (0.27812)
Offset at C5	Distance offset from centerline at C5 (m)	-0.06188 (0.27577)
Offset at PT	Distance offset from centerline at PT (m)	-0.0158 (0.30251)
Delineation	Delineation condition (0: highly visible, 1:visibile)	72%
4’>Shoulder	Paved shoulder greater than 4’ indicator (0: paved shoulder less than 4’, 1: paved shoulder >=4’)	20%

Table 2.3 Variables Explored in Analysis

Variable	Description
CurveID	Unique identifier for each curve including an identifier for each, state, buffer and curve
EventID	ID given by VTTI to uniquely identify each trace through a buffer
Curve Point	Factored variable which indicates the position in the curve where data are sampled from (PC, C1, C2, CC, C4, C5 or PT)
Radius	Radius of the curve (m)
Length	Length of curve (m)
Deflection Angle	Deflection angle for full circular curve measured from tangent at PC or PT
LaneWidth	Width of the travel lane (m)
SuperElevation	Average Cross Slope of the segment (%)
Chevrons	Indicator variable for chevrons (0: not present, 1:present)
Rumblestrips	Indicator variable for rumble strips (0: not present, 1:present)
Guardrail	Indicator variable for guardrail (0: not present, 1:present)
RPM	Indicator variable for raised pavement markings (0: not present, 1:present)
AdvisSign	Indicator variable for curve advisory sign (0: not present, 1:present)
Nighttime indicator	Indicator variable for nighttime (0: daytime or dawn/dusk, 1:nighttime)
SpeedUp	Speed limit in upstream (mph)
AdvisorySpeed	Speed limit in curve when advisory speed is present
Over300	Amount over the speed limit at 300 m upstream of curve (mph)
OverSpeed	Amount over the speed limit at point in curve (mph)
Speed (mph)	Speed at point in the curve (mph)
Offset	Distance offset from centerline in points throughout curve (m)
Offset300	Distance offset from centerline 300 m upstream of curve (m)
Offset250	Distance offset from centerline 250 m upstream of curve (m)
Offset200	Distance offset from centerline 200 m upstream of curve (m)
Offset150	Distance offset from centerline 150 m upstream of curve (m)
Offset100	Distance offset from centerline 100 m upstream of curve (m)
Offset50	Distance offset from centerline 50 m upstream of curve (m)
Distracted	Visual distraction at curve point indicator (1:distraction present, 0: no distraction)
DistractedBefore	Visual distraction between curve points indicator (1: distraction present, 0: no distraction)
Forward	Forward glance at point in curve indicator (1: glance is forward, 0: glance away)
Down	Glance is down indicator (1: glance is down, 0: glance is anywhere but down)
SA	Roadway-related glance (1: roadway-related glance, 0: otherwise)
NR	Non-roadway-related glance at point in curve indicator (1: present, 0: not present)
NRBefore	Non-roadway-related glance between curve points indicator (1: present, 0: not present)
NRup	Non-roadway-related glance in 200 m upstream of curve indicator (1: present, 0: not present)
NRcurve	Non-roadway-related glance in curve indicator (1:present, 0: not present)
Visibility	Visibility indicator (1:low visibility due to fog or glare, 0:otherwise)
Surface	Surface condition (0:dry, 1:pavement wet but not currently raining, 2: snow present, but roadway is bare)
PaveCond	Pavement condition (0: normal surface condition, 1: moderate damage, 2:severe damage)
Delineation	Delineation condition (0: highly visible, 1:visibile, 2:obscured)
Shoulder	Paved shoulder width (1: less than 1’, 2: 1’ to less than 2’, 3: 2’ to less than 4’ 4: greater than or equal to 4’
LargeShoulder	Paved shoulder greater than or equal to 4 feet indicator (0:not present, 1:present)
Gender	Gender Indicator (0:Female, 1: Male)
Under25	Age under 25 indicator (0:over 25, 1: under 25)
Under30	Age under 30 indicator (0:over 30, 1: under 30)
Age	Age of driver at time of first drive
LargeVeh	Large Vehicle (i.e., truck or SUV) indicator (0:car, 1:truck or SUV)

Vehicle offset was the metric used to determine normal driving on the curve as suggested by Hallmark et al, 2011 *(18)*. Due to this, it was required that the offset data be quite accurate, as small discrepancies in the offset could drastically skew the results of the model. This was assessed using the lane markings probability variables in the DAS data. After conferring with VTTI, who collected the data, a threshold was set for the probability which they deemed the data to be accurate and only those samples that were above this threshold were included. Additionally the offset data sampled at 0.1 seconds were plotted to identify outliers. Time series data for curves that had accurate offset at the sampling points, were isolated and then checked to make sure a lane departure did not occur within the curve. Then all of the data including the glance and distraction were merged. Data were ultimately available for 12 unique curves. Thirty traces were available for the inside (right-hand curve) model, and twenty-five were available for the outside (left-hand curve) model. This sample was small, which does limit the applicability of the results, and was due to the inaccuracy in the offset data for the majority of samples. Approximately 10% of the samples examined contained accurate enough offset data to include in the analysis and some of those had to be thrown out as lane departures occurred in these curves. Drivers were distributed by age and gender, as shown in Table 2.4.

Tahle 2.4 Driver Characteristics

Sex	**Age**			**Total**
	16 to 25	**26 to 50**	**50-90**	
Inside curve (right-hand)				
Male	0	2	4	7
Female	4	1	2	7
Outside curve (right-hand)				
Male	0	1	3	4
Female	4	1	6	11

2.4 Analysis

Models for lane position were developed with offset of the center of the vehicle from the center of the lane as the dependent variable for both inside (right-hand curve from the perspective of the driver) and outside (left-hand curve from the perspective of the driver) curves.

A generalized least squares (GLS) model was utilized. A panel data model was tested due to the time-series and cross-sectional nature of the data, with “EventID” as the individual and “Point in Curve” as the time setting. The Breusch-Pagan Lagrange multiplier test found that no panel effect was present, and therefore an ordinary least squares (OLS) model was appropriate. After running the OLS models, it was determined that there were problems with autocorrelation due to the time series nature of the data. A GLS model was then utilized as it is similar to OLS except that it allows models to be fit with a correlated-error structure as seen in our data.

The GLS function in the NLME package of R was used to develop the models. Models were selected to minimize Akaike information criterion (AIC) and Bayesian information criterion (BIC), while including significant variables (α=.05) from the list in Table 2.2. Correlation between the dependent variable and independent variables as well as the correlation between independent variables were examined to determine which variables should potentially be included in the model. The order of autoregression parameter was tested using an analysis of variance (ANOVA) test. The correlation structure of the model took into account the grouping across each event through each unique curve. The grouping factor allows for the correlation structure to be assumed to apply only to observations within the same unique event and curve.

2.5 Results

The results for the two models developed can be seen in the sections below. Neither of the best fit models included the majority of roadway factors which have been cited in the literature. Curve radius, curve length, super elevation, or deflection angle were not found to be significant factors. Additionally other factors cited in the literature such as time of day or vehicle type were also not found to be significant. This may be due to the small sample sizes that were available for this study.

2.5.1 Results for Inside of Curve

The best fit model for lane position for right (inside) curves was developed using 210 observations and contained 10 variables. The list of variables and parameter estimates is shown in Table 2.5. The model suggests an association that as drivers tend to the right (towards the edge line) in the upstream, the offset in the curve also shifts to the right, or near the outside of the lane. It also found that the presence of a curve advisory sign corresponds to drivers shifting 0.22 meters to the right. This would be expected as advisory signs are usually placed on sharper curves where drivers are more likely to flatten their path.

A driver glancing down at a particular point in the curve is associated with the driver's lane position shifting to the right near the outside of the lane 0.30 meters more than if they were not glancing down. The model also found a correlation between age and lane position. Drivers under 30 years were associated with a shift 0.21 meters towards the left (more towards the roadway center).

Finally, the model includes indicator variables relating to the position in the curve. At position C1 (as shown in Figure 2.1), which is just past the point of curvature, the average position is 0.14 meters to the right of the center of the lane, and at position C2 the average position is 0.21 meters. As the driver gets to the center of the curve (position CC), the average lane position is 0.28 meters to the right. Drivers then shift even more right at position C4 to 0.38 meters. Then drivers move back towards the center of the lane at positions C5 and the PT (0.20 and 0.15 meters, respectively). As indicated, a driver's drift to the outside lane edge near the center of the curve suggests that the driver may be most vulnerable to a right-side roadway departure near the center of the curve or just past it. These followed the trends of the input data.

These parameters support the idea that drivers do not maintain a smooth path through the curve. The first-order autoregression parameter phi was found to be 0.59, and the second-order was -0.33.

Table 2.5 Significant Variables for Right Curve Lane Position Model

Variable	**Parameter Estimate**	***p*-value**
Constant	0.02468	0.5711
Offset at 100 feet upstream of curve	0.38240	0.0000
Driver's glance is down indicator (0: if drivers glance is not down, 1: if drivers glance is down	0.29650	0.0047
Under 30 indicator (0: driver's age is 30 or older, 1:driver's age is under 30)	-0.21177	0.0000
C1 position indicator (0:not C1, 1:C1)	0.13564	0.0015
C2 position indicator (0:not C2, 1:C2)	0.20893	0.0004
CC position indicator (0:not CC, 1:CC)	0.28193	0.0000
C4 position indicator (0:not C4, 1:C4)	0.37759	0.0000
C4 position indicator (0:not C5, 1:C5)	0.19713	0.0006
PT position indicator (0: not PT, 1:PT)	0.14903	0.0080
Curve Advisory sign indicator (0: sign no present, 1: sign present)	0.21890	0.0089
First-order autoregression disturbance parameter (phi 1)	0.59334	
Second-order autoregression disturbance parameter (phi 2)	-0.32594	
Number of Observations	210	

2.5.2 Results for Outside of Curve

The best fit model for lane position for left (outside) curves was developed using 175 observations and included 9 variables, as shown in Table 2.6. The parameter for offset at 100 meters is similar to that in the right curve lane position model. The model suggests that if a driver tends to drive to the right of the lane center upstream of the curve, the driver also tends to drive to the right of the lane center within the curve.

The presence of a large paved shoulder (>=4 feet) correlates to the driver moving towards the right (towards the edge line) by 0.21 meters, which is expected because the driver has more space than when no paved shoulder is present. Less visible delineation, when lane lines are harder to see (examples in Appendix A), associates to drivers shifting to the left and towards the center line by 0.16 meters.

Indicator parameters for position in the curve were also included. While the parameters for indicators C4, C5 and PT were not significant, they were still included because they give some information on the change in position throughout the curve. The parameters were similar to what was seen in the input data.

As drivers enter the curve and move to the center of the curve (position C1 to CC, as shown in Figure 2.1), they tend to be positioned around 0.13 to 0.6 meters to the left of the center of the lane (towards the centerline). As drivers moves to the end of the center of the curve (position C4, C5 and the PT), they shift back towards the center of the lane. This suggests that drivers may be most likely to cross the roadway centerline in the first half of the curve.

Table 2.6 Significant Variables for Left Curve Lane Position Model

Variable	**Parameter Estimate**	***p*-value**
Constant	0.07476	0.2312
Offset at 100 feet upstream of curve	0.37602	0.0001
Delineation indicator (0: highly visible, 1:visible)	-0.16487	0.0016
Paved shoulder greater than 4’ indicator (0: paved shoulder less than 4’, 1: paved shoulder >=4’)	0.21265	0.0005
C1 position indicator (0:not C1, 1:C1)	-0.12685	0.0098
C2 position indicator (0:not C2, 1:C2)	-0.18881	0.0075
CC position indicator (0:not CC, 1:CC)	-0.26054	0.0008
C4 position indicator (0:not C4, 1:C4)	-0.12880	0.0851
C5 position indicator (0:not C5, 1:C5)	-0.0402	0.5711
PT position indicator (0: not PT, 1:PT)	0.00588	0.9321
First-order autoregression disturbance parameter (phi 1)	0.70482	
Second-order autoregression disturbance parameter (phi 2)	-0.35961	
Number of Observations	175	

2.6 Summary and Conclusions

The objective of this research was to develop a model of normal curve driving. Understanding how a driver normally negotiates a curve during various situations provides insight into not only how characteristics of the roadway, driver, and environment potentially influence how a driver drives, but also the areas that can lead to lane departures. Knowing how much drivers normally deviate in their lane could potentially have implications on policy or design such as determining lane widths and shoulder widths.

Conceptual models of curve driving were developed to assess changes in lane position as the driver negotiates the curve and interim results were reported. Data for several positions upstream and along the curve were sampled from the time series data. Models were developed using GLS for lane position for both inside (right-hand curve from the perspective of the driver) and outside (left-hand curve from the perspective of the driver), resulting in two models. Lane position was modeled as the offset of the center of the vehicle from the center of the lane.

Results indicate that lane position within the curve is correlated to lane position upstream of the curve. The models developed for offset of lane centerline in this study found that drivers who glanced down from the roadway were associated with a shift away from the center of the lane towards the inside of the curve. When driving on the inside lane, a driver who looked down at a particular point within the curve shifted 0.30 meters to the right compared to if they had not been looking down. This supports the role of attention in lane keeping.

Additionally, the models found that drivers on the inside of a curve tended to move more to the right at just past the center of the curve, while drivers on the outside of a curve were at the furthest point from the centerline at the center of the curve. This suggests that drivers may be particularly vulnerable to roadway departures at certain points in the curve negotiation process and supports previous findings *(3,4,13,14).*

Down glances and position within the curve indicate that drivers may be more vulnerable to a lane departure at certain points within the curve. As a result, countermeasures such as rumble strips, paved shoulders, and high-friction treatments may reduce the consequences of variations in lane position through the curve. Additionally, large paved shoulders were associated with drivers shifting towards the outside of the lane more than small paved shoulders in left-hand curves. Finally, lower visibility delineation was correlated to drivers driving more towards the center of the roadway on left-handed curves. This potential relationship supports the idea that poor delineation affects curve negotiation and better delineation through new paint or use of RPMs could help improve this negotiation.

2.6.1 Limitations

The main limitation of this analysis was sample size. Reliable offset data were only available in a subset of the vehicle traces that were reduced. As a result, the number of driver types and roadway features that could be modeled was limited. Consequently, the results are not transferable to all curves or situations. Adding more data to these models may draw out more relationships or strengthen those already found. A more robust data set could also allow for a mixed effects model to be performed, which would allow the findings to be applied towards all curves and not just those examined.

The face and in-cabin video at times had to be coded based solely on head movements as eyes were obscured due to the drivers wearing sunglasses or poor quality and grainy video. This may have resulted in minor glances such as rear-view mirror or steering wheel being missed. It was decided to include these in the analysis in order to be able to include nighttime driving and have as much data as possible. While these minor glances may have been missed, major distractions and glances which are associated with a head movement were picked up. Throughout the analysis it was found that the subtle glances were not significant, so the fact that they were not able to be discerned in some cases should not have been a problem.

2.7 Acknowledgements

This work was sponsored by the Federal Highway Administration in cooperation with the American Association of State Highway and Transportation Officials, and it was conducted in the Strategic Highway Research Program, which is administered by the Transportation Research Board of the National Academies.

2.8 References

1. Cheung, J. Horizontal Curve Safety. 2010. http://safety.fhwa.dot.gov/roadway_dept/horicurves/cmhoricurves/horiz_curve.pdf. Accessed January 2012.

2. Glennon, J.C., T.R. Neuman, and J.E. Leisch. *Safety and Operational Considerations for Design of Rural Highway Curves*. Report FHWA/RD-86-035. Federal Highway Administration, Washington, D.C., 1985.

3. Felipe, E. and F. Navin. Automobiles on Horizontal Curves: Experiments and Observations. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1628, 1998, pp. 50–56.

4. Stodart, B.P. and E.T. Donnell. Speed and Lateral Position Models from Controlled Nighttime Driving Experiment. *ASCE Journal of Transportation Engineering*, Vol. 134, No. 11, 2008, pp. 439–449.

5. Lamm, R., E. M. Choueiri, J.C. Hayward, and A. Paluri. Possible Design Procedure to Promote Design Consistency in Highway Geometric Design on Two-Lane Rural Roads. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1195, 1988, pp. 111–122.

6. Miaou, S.-P., and H. Lum. Statistical Evaluation of the Effects of Highway Geometric Design on Truck Accident Involvements. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1407, 1993, pp. 11–24.

7. Council, F.M. Safety Benefits of Spiral Transitions on Horizontal Curves on Two-Lane Rural Roads. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1635, 1998, pp. 10–17.

8. Zegeer, C.V., J.R. Stewart, F.M. Council, and D.W. Reinfurt. *Safety Effects of Geometric Improvements on Horizontal Curves*, University of North Carolina, Chapel Hill, NC, 1991.

9. Antin, J. Technical Coordination & Quality Control (S06). Presented at the 8th SHRP 2 Safety Research Symposium, Washington D.C., 2013.

10. VTTI. *InSight Data Access Website SHRP 2 Naturalistic Driving Study*. https://insight.shrp2nds.us/. Accessed July 31st, 2014.

11. Smadi, O. SHRP 2 S-04A Roadway Information Database Development and Technical Coordination and Quality Assurance of the Mobile Data Collection Project. Presented at the 7th SHRP 2 Safety Research Symposium, Washington D.C., 2012.

12. Spacek, P. FahrverhaltenunUnfallgeschehen in Kurven, Fahrverhalten in Kurvenbereichen. InStitutFürVerkehrsplanung, Transporttechnick, Strassen-unEisenbahnbaur, ETH Zürich, 1998.

13. Fitzsimmons, E., V. Kvam, R.R. Souleyrette, S.S. Nambisan, and D.G. Bonett. Determining Vehicle Operating Speed and Lateral Position along Horizontal Curves Using Linear Mixed-Effects Models. *Traffic Injury Prevention*, Vol. 14, Issue, 3, 2013, pp. 309-321.

14. Levison, W.H., J.L. Campbell, K. Kludt, A.C. Vittner Jr., I. Potts, D. Harwood, J. Hutton, D. Gilmore, J.G. Howe, J.P. Chrstos, R.W. Allen, B. Kantowitz, T. Robbins, and C.

Schreiner. Development of a Driver Vehicle Module (DVM) for the Interactive Highway Safety Design Model (IHSDM). Federal Highway Administration. Report FHWA-HRT-08-019. November 2007.

15. Hallmark, S.L., N. Oneyear, S. Tyner, B. Wang, C. Carney and C. McGehee. *SHRP 2 S08D: Analysis of the SHRP 2 Naturalistic Driving Study Data*. Strategic Highway Research Program 2, Transportation Research Board, Washington, D.C., 2015. ‘

16. Shinar, D., E.D. McDowell, and T.H. Rockwell. Eye Movements in Curve Negotiation. *Human Factors*, Vol. 19, 1977, pp. 63–71.
17. Angell, L., J. Auflick, P.A. Austria, D. Kochhar, L. Tijerina, W. Biever, T. Diptiman, J. Hogsett, and S. Kiger. *Driver Workload Metrics*. National Highway Traffic Safety Administration, US Department of Transportation, Washington, D.C., 2006.

18. Hallmark, S., Y. Hsu, L. Boyle, A. Carriquiry, Y. Tian and A. Mudgal. *Evaluation of Data Needs, Crash Surrogates, and Analysis Methods to Address Lane Departure Research Questions using Naturalistic Driving Study Data.* Strategic Highway Research Program Report S2-S01E-RW-1. Transportation Research Board, Washington D.C., 2011.

CHAPTER 3 - CONCEPTUAL LINEAR MIXED EFFECTS MODEL OF RURAL TWO LANE CURVE DRIVING USING SHRP 2 NATURALISTIC DRIVING DATA

A paper to be submitted to *Accident Analysis and Prevention*

Nicole Oneyear, Shauna Hallmark, Cher Carney, and Dan McGehee

Abstract

Rural curves pose a significant safety problem due to the higher rate of crashes on curves than tangent sections. Run off the road crashes on horizontal curves are a particular problem as they accounted for approximately 27% of all fatalities in 2008; the majority of which took place on rural curves. Addressing lane-departure crashes on rural curves is a priority for National, State, and local roadway agencies. Much research has been conducted to look at how roadway factors, like radius and shoulder width and environmental factors, such as weather affect crashes, yet limited research has been conducted looking at how driver behaviors affect crash risk.

This paper utilizes data from the SHRP 2 Naturalistic Driving Study (NDS) and Roadway Information Datasets (RID) to present results on the development of a conceptual model of curve driving on rural two lane curves that explores how drivers interact with the roadway environment. The model helps identify zones where driver are more likely to have lane departures and defines boundaries between lane encroachment events and normal driving.

A Linear Mixed Effects Model with offset from the center of the lane as the dependent variable was developed using times series data, at the level of 0.1 second as the data input. The model provides a means to predict drivers offset at seven positions in the curve with and without lane departures towards the inside. Lateral position upstream of the curve, the direction of the curve (inside/right, outside/left) and driver factors such as sex, downward glance or distraction in the section prior were found to be significant factors which affect offset from the center of the lane.

3.1 Introduction

Rural curves pose a significant safety problem due to the three times higher rate of crashes on curves than tangent sections *(1).* Lane departure crashes on these rural curves are especially of concern due to the fact that approximately 27% of all fatalities in 2008 occurred on horizontal curves and over 80% of these were run off the road crashes, with the majority of these fatal crashes occurring on rural two lane highways *(2)*.

The objective of this paper was to understand how drivers negotiate curves. This was done by building on a previous paper (*3*) where conceptual models of isolated curve driving on rural two lane curves utilizing data from the SHRP 2 Naturalistic Driving Study (NDS) and Roadway Information Database (RID) were developed by including additional data and non-isolated curves such as S curves.

A better understanding of the interaction between driver characteristics and curve negotiation needs can potentially lead to better design and application of countermeasures. For instance, if older drivers have the hardest time with curve negotiation because they are less likely to see visual cues, the best solution might be larger chevrons. On the other hand, a solution geared towards younger drivers might include more closely spaced chevrons to help drivers gauge the sharpness of the curve. Distracted drivers would perhaps require another solution, such as a tactile cue from transverse rumble strips.

Studies of roadway factors, such as radius *(4,5,6,7),* presence of spirals *(8)*, or shoulder width and type *(9)*, have provided some information regarding the most relevant curve characteristics, but information is still lacking. In addition, little information is available that identifies driver behaviors that contribute to curve crashes and curve negotiation. As a result, a better understanding of how drivers interact with various roadway features and countermeasures may provide valuable information to highway agencies for determining how resources can best be allocated in order to prevent potential lane departures and reduce crashes.

3.1.1 Background on SHRP 2 Naturalistic Driving Study

The SHRP 2 NDS is the largest naturalistic driving study to date. The study was conducted by Virginia Tech Transportation Institute (VTTI). Drivers in six states (Florida, Indiana, New York, North Carolina, Pennsylvania and Washington) had their vehicles equipped with a Data Acquisition System (DAS) which collects information such as speed, acceleration, and GPS data, as well as four cameras which collected forward, rear, drivers face and over the shoulder video. These equipment captured all of the trips a driver made over a period of six months up to two years. Males and females ages 16 to 98 participated in the study. Over the three years of the study approximately 3,300 participants drove over 30 million data miles over 5 million trips *(10,11)*.

3.1.2 Background on SHRP 2 Roadway Information Database

In conjunction with the SHRP 2 Naturalistic Driving Study, another project was conducted to collect roadway information for the main roads traveled in the NDS. The Center for Research and Education (CTRE) led the effort which used mobile data collection to collect 12,500 centerline miles of data across the six states where the NDS was focused. Data collected included information on roadway alignment, signing, lighting, intersection location and types, presence of rumblestrips as well as other countermeasures. In addition to the mobile data collection effort, existing roadway data collected by local agencies was leveraged to increase the data available. Additionally, supplemental data such as crash data, changes to laws, and construction projects were also collected to further strengthen the database *(12)*.

3.2 Previous Research

Limited research has been conducted to develop models of curve driving. Models developed differed slightly in approach, but had similar findings. Radius and direction of curve were found to affect lateral position in the curve. Additionally, it was found that most drivers tended to move towards the inside of the curve as they approached the center and therefore flattened the path in which they traveled. The approaches of five models are discussed in further detail.

Spacek (1998) developed a model of curve negotiation behavior based on lateral position across seven points in a curve. Spline interpolation was used to develop six track profiles which were commonly observed in the field. The models disaggregated curve paths to normal behavior, common intentional lane deviations (cutting and swinging), and two profiles that indicated driver adjustments after misjudging a curve (drifting and correcting). The normal behavior found that drivers tended to drive more towards the inside of the lane, effectively flattening their paths *(13)*.

Felipe and Navin (1998) also evaluated lateral placement through curves using an instrumented vehicle along a two-lane mountainous road and found that vehicle path tended to differ based on the radius of the curve. Vehicles mostly followed the center of the lane for curve with large radii; however with smaller radii curve, they found that drivers in followed a flattened path to minimize speed change. They report that variation in path selection was a function of road geometry, surrounding traffic and the driver *(4)*.

Stodart and Donnell (2008) also found curve radius and curve direction to significantly impact lateral position using data collected upstream and within six curves using instrumented vehicles with 16 research participants during nighttime conditions. They used ordinary least squares regression and compared change in lateral position from the upstream tangent to the curve midpoint *(5)*.

Fitzsimmons et al (2014) modeled vehicle trajectories using mixed effects models for a rural and an urban curve in Iowa. Pneumatic road tubes were used to collect lateral position of the vehicles in 5 points throughout each curve. Similar to the Spacek study*(13)*, it was found that most vehicles tended to traverse the curve as if the radius was larger than the design radius of the curve and therefore tended to travel towards the inside of the curve as they approached the center. The study also found that time of day, direction of curve and vehicle type all affected lateral positon in the curve *(14).*

Levison et al. (2007) developed a driver vehicle module to use with the Interactive Highway Safety Design Model. One component of this model was path selection and was assumes the drivers desired path profile is one that drivers the curve as if it had a larger radius than it does *(15)*.

These model of curve driving have taken into account roadway, environmental, and to a limited extent driver factors yet none have not taken into account driver behavior and how distraction can affect lateral position. A previous study, using a small sample of the SHRP 2 data by Oneyear et al. (2015) for isolated curves used generalized least squares regression to create models for curve driving for inside and outside curves. This model also found that drivers flatten their path as they traverse the curve. It however was not able to find common factors in previous research such as radius to be significant. This study hopes to expand on the work started in Oneyear et al. to create a model which includes additional data as well as non-isolated curves and is transferable to all curves and drivers by including random effects *(3)*.

3.3 Methodology

Data were acquired from two main sources, unless noted otherwise. These were the SHRP 2 Naturalistic Driving Study (NDS) and the SHRP 2 Roadway Information Database (RID). The NDS included time series data collected through a Data Acquisition System (DAS), as well as video data collected from 4 cameras placed in the vehicle which captured the forward view, rear view, driver's face and over the shoulder view. As the driver's face and over the shoulder video contained potentially identifying information, these data were viewed at the secure enclave housed at VTTI.

3.3.1 Identification of Curves of Interest

At the time this project was conducted, the NDS and RID had not been linked. As a result, the team manually identified curves of interest and then requested any trips on these curves from the NDS. To identify potential curves of interest, the project team made use of weighted trip maps prepared by VTTI using a subset of trip data in the early stages of the NDS data collection. The trip maps were overlain with the RID and rural 2-lane curves on paved roadways were identified. A one-half mile tangent section upstream and downstream of each curve was also selected. Curves were identified in all states except for Washington since much of the roadway mileage was urban.

A spatial buffer (polygon) was created around each curve. In some cases curves were located near one another and multiple curves were included in a single buffer. The buffers were provided to VTTI and were overlain with the NDS. If a trip fell within a buffer and met certain criteria (i.e. GPS data present, speed data present, etc.) then it became a potential event (one trip through one buffer) to use in the analysis. At the time of the data request, around one-third of the NDS data had been processed and were available. The initial query resulted in around 4,000 traces (one trip through one buffer). Each trace was reviewed and traces where a needed variable was not present or reliable were removed from further consideration. Once these traces were removed, a total of 987 events across 148 curves were selected to represent a good cross-section of curve and driver characteristics. Further details on how the data were requested can be seen in the SHRP2 S08D Final report *(16)*.

3.3.2 Data Collection and Data Reduction

3.3.2.1 Roadway Variables

Roadway variables were extracted for the 148 curves using the RID data when available. In some cases a variable was not collected, and in other cases the RID was not available for the study segment because the RID did not cover all roads in the NDS. When the information was not available through the RID, other sources were used to manually extract the data. These additional sources were also used to confirm data collected through the RID, such as speed limit and advisory speed limit.

ArcGIS was used to measure distances between curves using the PCs and PTs included in the RID. ArcGIS was also used to determine whether the curve was an S-curve or a compound curve based on the distance between curves and direction of curves.

Google Earth was used to extract the roadway features not included in the RID. It was also used to collect countermeasures before the forward video was available, such as chevrons and RPMs, which were later confirmed with the NDS forward video. Radius was provided for most curves in the RID and was reported as radius by lane. When RID data were not available, which only included a few curves in Florida, radius was measured using aerial imagery and the chord-offset method. This method was verified using curves with known radii. NDS forward video was used to determine subject measures for delineation, pavement condition, roadway lighting, and roadway furniture (which describes objects around the road that provide some measure of clutter). Variables collected are shown in Table 3.1.

Table 3.1 Roadway Variables Extracted and Main Source

Feature	ArcGIS	SHRP2 RID	Google Earth	SHRP 2 NDS Forward Video
Curve radius		✓		
Distance between curves	✓			
Type of curve (isolated, S, compound)	✓			
Curve length	✓	✓		
Super elevation		✓		
Presence of rumble strips		✓		✓
Presence of chevrons		✓	✓	✓
Presence of w1-6 signs			✓	✓
Presence of paved shoulders		✓		✓
Presence of raise pavement markings (rpm)			✓	✓
Presence of guardrail			✓	✓
Speed limit		✓		
Advisory sign speed limit		✓	✓	✓
Curve advisory sign/W1-6		✓	✓	✓
Pavement condition				✓
Delineation				✓
Sight distance				✓
Roadway furniture				✓
Direction of curve				✓
Shoulder width and type		✓		

3.3.2.2 Vehicle, Traffic, Static Driver and Environmental Variables

Each of the traces or events represents one driver trip through a selected roadway segment. One spreadsheet (containing DAS data), one forward video, and one rearview video were provided by VTTI for each trace. Each row of data represents 0.1 seconds, and spatial location was provided at one-second intervals. A time stamp was also provided to link the various videos with the DAS data. A list of the main DAS variables provided and used in the analysis include the following:

- Acceleration, x-axis: vehicle acceleration in the longitudinal direction vs. time
- Acceleration, y-axis: vehicle acceleration in the lateral direction vs. time

- Lane markings, probability, left/right: Probability that vehicle based machine vision lane marking evaluation is providing correct data for the left/right side lane markings
- Lane position offset (m) : Distance to the left or right of the center of the lane based on machine vision
- Lane width (m): Distance between the inside edge of the innermost lane marking to the left and right of the vehicle
- Spatial position: Latitude and Longitude
- Speed : Vehicle speed indicated on speedometer collected from network
- Timestamp Integer used to identify one time sample of data. Arbitrary counter that is unique for each data row in each file. Used by the community viewer.
- Yaw rate, z-axis: Vehicle angular velocity around the vertical axis.

Vehicles traces were overlain with the RID curve, the nearest GPS points to the PC or PT was found and the position of the PC/PT was located within the time series data using interpolation. Once PC/PT were established, vehicle position upstream or downstream of the curve was calculated using speed. For some traces, there were multiple curves, so the PC/PT and upstream/downstream distances were determined for each curve. In some cases, speed was missing for multiple time stamps. In these cases, speed was interpolated assuming a constant increase or decrease.

The static driver and vehicle characteristics were merged with each trace. The characteristics used include driver age and gender and vehicle class and track width.

The forward video was used to reduce the environmental and other variables. Appendix A includes information on how these data were collected. The variables collected included the following:

- Surface condition (i.e., dry, wet, snow, etc.)
- Lighting conditions (i.e., day, dawn, dusk, night with no lighting, night with lighting)
- Visibility (i.e. high visibility (clear), low visibility (foggy))
- Locations of vehicles in the opposite direction passing the driver's vehicle
- Locations where the driver's vehicle was following another car
- Presence of curve advisory signs
- Presence of chevrons

Information on whether there was a lane encroachment, defined as a right or left vehicle edge lane line crossing was also gathered using the forward video and kinematic vehicle data. For the purpose of this research an encroachment was determined to have occurred when two of the following criteria were present:

- vehicle edge is 0.2 meter beyond lane line
- 0.2 g lateral acceleration is present
- a lane crossing is visually confirmed using the forward view

3.3.2.3 Kinematic Driver Characteristics

Driver attention was measured by the location where a driver was focused for each sampling interval. Scan position, or eye movement, has been used by several researchers to gather and process information about how drivers negotiate curves *(17)*. The majority of studies have used simulators to collect eye tracking information. Because eye tracking is not possible with NDS data, glance location was used as a proxy. Glance locations, represent practical areas of glance locations for manual eye glance data reduction. Glance locations were coded using the camera view of the driver's face, with a focus on eye movements, but taking into consideration head tilt when necessary. Glances were coded as one of 11 potential locations which can be seen below:

- Front
- Left
- Right
- Down
- Steering Wheel
- Center Console
- Rearview Mirror
- Up
- Over the Shoulder
- Missing (due to glare or problems with camera)
- Other Glance

Potential distractions were determined by examining both the view of the driver's face and the view over the driver's right shoulder, which showed hands on/off the steering wheel. Distractions were identified when drivers took their eyes off the forward roadway. Potential distractions include the following:

- Route planning (locating, viewing, or operating)
- Moving or dropped object in vehicle
- Cell phone (locating, viewing, operating)
- IPod/MP3 (locating, viewing, operating)
- Personal hygiene (i.e. makeup application, brushing hair, etc.)
- Passenger
- Animal/insect in vehicle
- In-vehicle controls
- Drinking/eating
- Smoking

Glance location and distractions were coded for 200 meters upstream and throughout each curve for only 515 of the events due to time constraints. Glance location and distractions were merged with the event files using time stamp as a reference. Once this was completed, glance location was indicated for each row in the DAS event file.

There were times in the manual reduction of the glance and distraction reduction when eye movements were obscured due to such things as glare, the driver wearing sunglasses, or darkness. When this occurred, head movement was used to estimate glance. This may have caused minor glances, such as at the steering wheel to have been missed. It should be noted that glance and distraction were more likely to have been accurately coded for traces with clearer views of the face and eyes. However, discarding data where head movements were used instead of eye movements would have entailed removing almost all nighttime data and significantly reducing sample size.

Glance location was further reduced to indicate time spent in “eyes-off-roadway” engaged in roadway-related tasks or “eyes-off-roadway” engaged in non-roadway-related tasks based on data coding used by Angell et al. (2006). The authors define roadway-related glances or situation awareness (SA) as glances to any mirror or speedometer. Glances to other locations are defined as not roadway-related (NR). Roadway-related glances (SA) included left mirror, steering wheel, and rear-view mirror *(18)*.

It was not possible to distinguish between a glance to the right mirror and a glance to the right for other reasons (e.g., to converse with passenger). Additionally, on a two-lane roadway, glances to the right mirror are not likely to be as common because drivers are not expecting vehicles to the right. Consequently, all glances to the right were considered to be non-roadway-related.

Additionally, when glances to roadway-related locations were also associated with a distraction, it was decided that these glances were likely to be non-roadway-related. For instance, a driver who was texting and glancing at the steering wheel was likely to be looking at the cell phone rather than the speedometer. As a result, non-roadway-related glances included center console, up, right, or down.

3.3.2.4 Data smoothing

Smoothing of the DAS data was necessary because a certain amount of noise in the data resulted in improbable data points. These points would be data points that would jump for 0.1 seconds out of a range of what was probable and then continue following the previously seen trend. Several different methods to smooth the data were investigated. The Kalman filter estimates the optimum average factor for each subsequent state using information from past states. It was determined that, although the Kalman filter was appropriate, developing a model for multiple variables for over all of the vehicle traces was overly complicated and time consuming.

A moving average method was selected because it is able to reduce random noise while retaining a sharp step response. Each of the variables listed above was smoothed over 5 data points (0.5 second) using a moving average method. This method involved averaging the data from the 0.2 seconds before the point of interest, the 0.1 second of interest and the 0.2 seconds after the point of interest.

3.3.3 Data Sampling

The sampling plan for the curve model can be seen in Figure 3.1. Data were sampled at each point shown (e.g., PC), and locations for sampling were determined after consulting previous research *(13,14)* as well as plotting events and determining which sampling scheme picked up common patterns. Sampling in the tangent section was based on distance. Sampling within the curve was at equidistant points rather than at a specified distance because the curves have varying lengths.

The points sampled within the curve were the PC, PT, and then five equally spaced points (C2, C3, CC (curve center), C4, and C5), as shown in Figure 3.1. Upstream data were collected at 100 and 50 meters. These locations were chosen based on a preliminary study conducted on isolated rural curves which found any distance upstream beyond these to be less significant. Because the data sampling plan required 100 meters of upstream data, the analysis did not include the second curve in a compound curve nor the second curve in closely spaced S-curves and only included curves with a tangent section that was at least 100 meters from the nearest upstream curve.

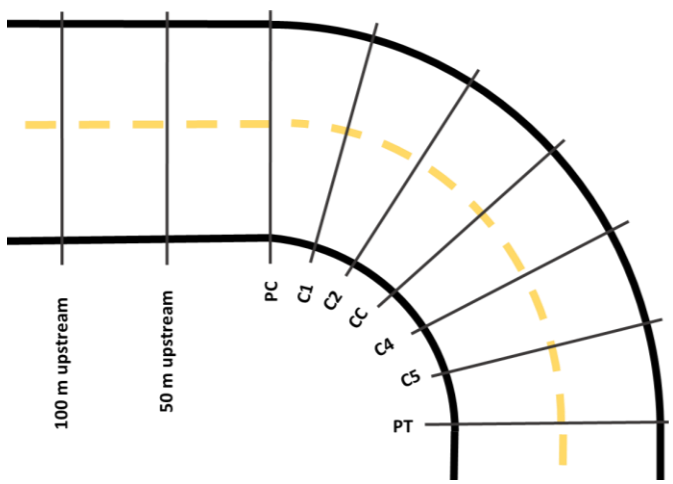


Figure 3.1 Data Sampling Layout for Curve Driving Model for Right-Handed Curve

The DAS and distraction data described previously were sampled at each point in the curve shown. Data collected for the upstream area included the offset and speed at each sample point, along with driver glance location and distractions. These data were merged with environmental, driver, and vehicle data. The summary statistics for the variables used in the final model are listed in Table 3.2, with the offset for the sampled points in the curve being presented separately as they are utilized in the model through the position in curve indicators. A complete list of variables collected, calculated and attempted in the model analysis are included in Table 3. 3. For some of the variables, (i.e. surface) only those conditions which were present in the data were included. Therefore since none of the samples occurred when it was raining heavily, that was not included as a condition. In other cases groupings were decided based on the samples available. While looking at the difference between a four foot shoulder and an eight foot shoulder would be helpful, not enough data were available to be able to look at this. Additional groupings not listed in the tables below were also tried such as only looking at effects for drivers under 25.

Vehicle offset was the metric used as a crash surrogate as suggested by Hallmark et al, 2011 *(19)*. A crash surrogate was necessary as the data received from VTTI contained only road departure crash. Due to offset being used as the main metric, it was required that the offset data be quite accurate, as small discrepancies in the offset could drastically skew the results of the model. This was assessed using the lane markings probability variables in the DAS data. After conferring with VTTI, who collected the data, a threshold was set for the probability which they deemed the data to be accurate and only those samples that were above this threshold were included. Additionally the offset data sampled at 0.1 seconds were plotted to ensure additional bad data did not exist. Then all of the data including the glance and distraction were merged.

Data were ultimately available for 323 traces across 98 unique curves with 68 unique drivers. This sample was relatively small compared to the size of the SHRP 2 NDS database, which does limit the applicability of the results, and was due to the inaccuracy in the offset data for the majority of samples. Approximately 10% of the samples examined contained accurate enough offset data to include in the analysis. Drivers were distributed by age and gender, as shown in Table 3.4 and curve and traces were distributed by radius as shown in Table 3.5.

Table 3.2 Summary Statistics for Select Variables

Variable	Description	Mean (std dev) or %
Offset 100	Distance offset from centerline 100 m upstream of curve (m) (+) value is in direction of inside of curve (-) is toward outside of curve	-0.01811 (0.33731)
Offset at PC	Distance offset from centerline at PC in meters (+) value is toward inside of curve (-) is toward outside of curve	-0.01648 (0.35527)
Offset at C1	Distance offset from centerline at C1 in meters (+) value is toward inside of curve (-) is toward outside of curve	0.06484 (0.35944)
Offset at C2	Distance offset from centerline at C2 in meters (+) value is toward inside of curve (-) is toward outside of curve	0.16127 (0.34662)
Offset at CC	Distance offset from centerline at CC in meters (+) value is toward inside of curve (-) is toward outside of curve	0.21790 (0.38034)
Offset at C4	Distance offset from centerline at C4 in meters (+) value is toward inside of curve (-) is toward outside of curve	0.19490 (0.35364)
Offset at C5	Distance offset from centerline at C5 in meters (+) value is toward inside of curve (-) is toward outside of curve	0.10676 (0.35670)
Offset at PT	Distance offset from centerline at PT in meters (+) value is toward inside of curve (-) is toward outside of curve	0.04563 (0.32685)
Down	Indicator that driver is glancing down (0: glance not down, 1: glance is down)	2%
Sex	Indicator for gender (0: Female, 1: Male)	39.6%
Direction	Indicator for direction of curve (0: outside or left, 1: inside of right)	5.0%
Distracted in section prior	Indicator for distraction between points in the curve (0: not distracted, 1: distracted)	8.5%
Lane Encroachment Inside (LEI)	Indicator that a lane encroachment towards the inside occurred within the curve (0: no inside lane encroachment 1: inside lane encroachment)	6.9%
Offset at PC with LEI	Distance offset from centerline at PC in meters if an inside lane encroachment occurred in the curve (+)value is toward inside of curve (-) is toward outside of curve	0.21907 (0.30935)
Offset at C1 with LEI	Distance offset from centerline at C1 in meters if an inside lane encroachment occurred in the curve (+)value is toward inside of curve (-) is toward outside of curve	0.41047 (0.26192)
Offset at C2 with LEI	Distance offset from centerline at C2 in meters if an inside lane encroachment occurred in the curve (+)value is toward inside of curve (-) is toward outside of curve)	0.55578 (0.27539)
Offset at CC with LEI	Distance offset from centerline at CC in meters if an inside lane encroachment occurred in the curve (+)value is toward inside of curve (-) is toward outside of curve	0.70485 (0.28014)
Offset at C4 with LEI	Distance offset from centerline at C4 in meters if an inside lane encroachment occurred in the curve (+)value is toward inside of curve (-) is toward outside of curve	0.60502 (0.41481)
Offset at C5 with LEI	Distance offset from centerline at C5 in meters if an inside lane encroachment occurred in the curve (+)value is toward inside of curve (-) is toward outside of curve	0.38831 (0.55347)
Offset at PT with LEI	Distance offset from centerline at PT in meters if an inside lane encroachment occurred in the curve (+)value is toward inside of curve (-) is toward outside of curve	0.22022 (0.49410)

Table 3.3 Variables Explored in Analysis

Variable	Description
CurveID	Unique identifier for each curve including an identifier for each, state, buffer and curve
EventID	ID given by VTTI to uniquely identify each trace through a buffer
DriverID	Unique identifier given to each driver
Curve Point	Factored variable which indicates the position in the curve where data are sampled (PC, C1, C2, CC, C4, C5 or PT)
Radius	Radius of the curve (m)
Length	Length of curve (m)
DefllectAngle	Deflection angle for full circular curve measured from tangent at PC or PT
LaneWidth	Width of the travel lane (m)
SuperElevation	Average Cross Slope of the segment (%)
Chevrons	Indicator variable for chevrons (0: not present, 1:present)
Rumblestrips	Indicator variable for rumble strips (0: not present, 1:present)
Guardrail	Indicator variable for guardrail (0: not present, 1:present)
RPM	Indicator variable for raised pavement markings (0: not present, 1:present)
AdvisSign	Indicator variable for curve advisory sign (0: not present, 1:present)
SpeedUp	Speed limit in upstream (mph)
AdvisorySpeed	Speed limit in curve when advisory speed is present
Speed (mph)	Speed at point in the curve (mph)
Offset	Distance offset from centerline in points throughout curve (m)
Offset100	Distance offset from centerline 100 m upstream of curve (m)
Offset50	Distance offset from centerline 50 m upstream of curve (m)
GyroZ	Vehicle angular velocity around the vertical axis (yaw rate)
AccelX	Vehicle acceleration in the longitudinal direction versus time
Accel Y	Vehicle acceleration in the lateral direction versus time
Distracted	Visual distraction at curve point indicator (1:distraction present, 0: no distraction)
DistractedBefore	Visual distraction between curve points indicator (1: distraction present, 0: no distraction)
Forward	Forward glance at point in curve indicator (1: glance is forward, 0: glance away)
Down	Down glance indicator (1: glance is down, 0: glance is anywhere but down)
SA	Roadway-related glance (1: roadway-related glance, 0: otherwise)
NR	Non-roadway-related glance at point in curve indicator (1: present, 0: not present)
NRBefore	Non-roadway-related glance between curve points indicator (1: present, 0: not present)
NRup	Non-roadway-related glance in 200 m upstream of curve indicator (1: present, 0: not present)
NRcurve	Non-roadway-related glance in curve indicator (1:present, 0: not present)
Visibility	Visibility indicator (1:low visibility due to fog or glare, 0:otherwise)
Surface	Surface condition (0:dry, 1:pavement wet but not currently raining, 2: wet and light rain, 4: snow present, but roadway is bare, 5:snow along road edge and/or centerline)
PaveCond	Pavement condition (0: normal surface condition, 1: moderate damage, 2:severe damage)
Delineation	Delineation condition (0: highly visible, 1:visibile, 2:obscured)
Lighting	Light condition (0:daytime, 1:dawn/dusk, 2:nighttime, no lighting, 3:nighttime, lighting present)
Shoulder	Paved shoulder width (1: < 1’, 2: 1’ to <2’, 3: 2’ to < 4’ 4: greater than or equal to 4’
Gender	Gender Indicator (0:Female, 1: Male)
Age	Age of driver at time of first drive
Track	Vehicle track width in meters
VehClass	Class of vehicle (1:Car, 2:SUV Crossover, 3: Pickup Truck
LaneEncroach	Indicator variable for if a lane encroachment occurred in the curve (0: did not occur, 1: occurred)
LEI	Indicator variable for lane encroachment towards inside of curve (0: did not occur, 1: occurred)
LEO	Indicator variable for lane encroachment towards outside of curve (0: did not occur, 1: occurred)
DistUp	The distance from the PT of the previous curve to the PC of the current curve in meters
SightDist	The estimated sight distance of the curve in meters
Oncoming	Indicator variable for oncoming vehicle in other lane (0:no vehicle present, 1: vehicle oncoming)
Following	Variable for following another vehicle (0: not following, 1: following, 2: following closely)

Table 3.4 Driver Characteristics

Sex	Age 16 to 25	26 to 50	51 to 90	Total
Male	6	13	18	37
Female	15	8	8	31

Table 3.5 Curves and Traces by Curve Radius

	R< =750' (~230 m)	R >750' (~230 m) to <=1500' (~460 m)	R>1500 (~460 m) to <=2250 (~690 m)	R>2250 (~690 m)	Total
Number of Curves	7	19	28	44	98
Number of Traces	16	46	84	177	323

3.4 Analysis

A Linear mixed effects (LME) model was utilized to create a model which predicts a drivers offset of the center of the vehicle from the center of the lane at the seven points in the curve based on the drivers offset 100 meters upstream of the PC. Offset at 100 meters upstream was used instead of the 50 meters upstream offset based on data from previous research *(3)* as well as the fact that the 50 meters upstream data was less accurate for some of the traces. The LME model was chosen as it allows one to account for random effects due to repeated measures from including multiple traces by the same driver in the same curve. The general form of a LME model with random effects at two levels (nested) can be written as *(20)*:

$$y_{ijk} = \beta_j + b_i + b_{ij} + \epsilon_{ijk} \quad i = 1, \dots, n_i \quad j = 1, \dots, n_j, \quad k = 1, .. n_k$$

$$b_i \sim N(0, \sigma_1^2), \quad b_{ij} \sim N(0, \sigma_2^2), \qquad \epsilon_{ijk} \sim N\left(0, \sigma^2\right)$$

The LME function in the NLME package of R was used to develop the model. The best fit model was selected by finding the model which minimized Akaike information criterion (AIC) and Bayesian information criterion (BIC), while including significant variables (α=.05) from the list in Table 3.2. Correlation between the dependent variable and independent variables as well as the correlation between independent variables were examined to determine which variables should potentially be included in the model.

Due to the data being of a time series nature, a correction for the autocorrelation was required. The order of the autoregression parameter was tested using the acf() function in R and the analysis of variance (ANOVA) test. The correlation structure of the model took into account the grouping across each driver and each event through each unique curve. The grouping factor allows for the correlation structure to be assumed to apply only to observations within the same unique event, driver and curve.

CurveID nested within DriverID was used as the random variable in the model as repeated samples were taken for drivers with some drivers having repeated samples in certain curves. Cross random effects which would take into account the random effects due to CurveID and Driver ID separately may have been a better fit for the model, however due to limitations of the software this was not feasible. NLME requires that the correlation structure and random effects structures are similar; crossed random effects are not able to be used due to this. Another package (lme4) is available in R which allows one to easily incorporate cross random effects, however it does not allow one to incorporate a correlation structure which is required for this data set.

The basic assumptions of a LME model are that within-group errors are independent and ~N(0, σ^2) and are independent of the random effects and that random effects are normally distributed around 0 and covariance matrix Ψ and are independent for different groups *(20)*. Once the model was developed, these assumptions were tested. Two violations of the assumptions were found. The within-group errors were found to be dependent and the AR(2) correlation structure helped to address this. Plots also showed a potential problem with the constant variance assumption. To help address this problem models were tested assuming a variance structure with unequal variances for certain conditions. The heteroskedastic model was the best fit model and incorporates a weighted variance structure which takes into account the different variance structures with respect to when a lane encroachment occurs in the curve, when a non-roadway related glance occurs in the curve, or a combination of the two.

The output from R for random intercepts for the best fit are presented in Appendix 3.

3.5 Results

The results for the best fit model can be seen in the Table 3.6. The best fit model did not included the majority of roadway factors which have been cited in the literature. Curve radius, curve length, super elevation, or deflection angle were not found to be significant factors. Additionally other factors cited in the literature such as time of day or vehicle type were also not found to be significant. The most significant factors were found to be those related to the driver's position in the curve.

Table 3.6 Best fit model

Variable	Estimate	P value	95% Lower	95% upper
Intercept	-0.039	0.049	-0.079	-0.0002
Offset at 100 m upstream	0.438	<0.001	0.374	0.502
Small Radius (R<460m~1500')	0.067	0.050	0.000	0.134
Glancing down	0.080	0.016	0.015	0.146
Distracted in prior section	0.045	0.035	0.003	0.087
C1	0.074	<0.001	0.047	0.101
C2	0.172	<0.001	0.134	0.209
CC	0.223	<0.001	0.180	0.266
C4	0.205	<0.001	0.160	0.249
C5	0.124	<0.001	0.079	0.169
PT	0.066	0.004	0.021	0.111
PC : Inside lane encroachment	0.233	0.005	0.071	0.395
C1 : Inside lane encroachment	0.362	<0.001	0.202	0.523
C2 : Inside lane encroachment	0.407	<0.001	0.247	0.570
CC : Inside lane encroachment	0.482	<0.001	0.321	0.642
C4 : Inside lane encroachment	0.406	<0.001	0.247	0.565
C5 : Inside lane encroachment	0.280	<0.001	0.123	0.437
PT : Inside lane encroachment	0.162	0.037	-0.010	0.315
σ Driver random effect	0.026			
σ Curve in Driver random effect	0.096			
σ Residual	0.3382			
Phi 1	0.770		0.758	0.775
Phi 2	-0.197		-0.246	-0.147

The best fit model was developed using 2261 observations and included 18 variables. The model suggests an association that as drivers tend to the inside direction of the curve in the upstream, the offset in the curve also shifts to the inside. It also found a correlation between curves with a radius less than 460 meters shifting 0.067 meter towards the inside of the curve.

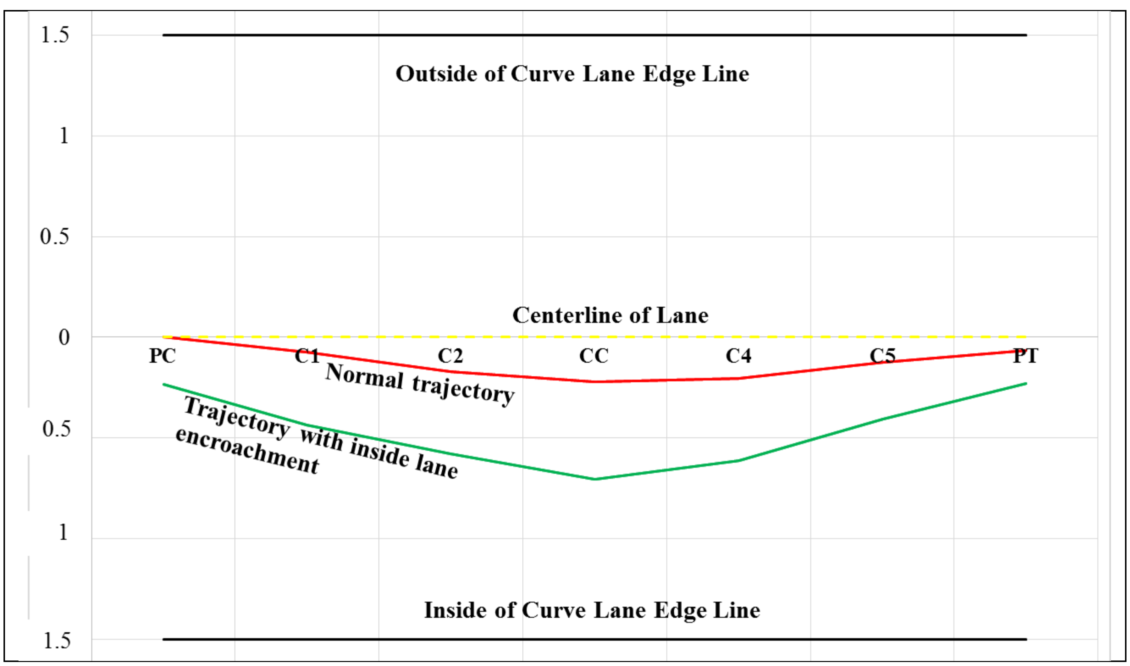
A driver glancing down at a particular point in the curve is associated with the driver's lane position shifting towards the inside of the curve by approximately 0.08 meters. A similar correlation was found if the driver was distracted in the prior section. Therefore if they were distracted between the PC and C1 their position at C1 would be 0.045 meters more towards the inside of the curve than if they were not distracted.

Next, the model includes indicator variables relating to the position in the curve. At position C1 (as shown in Figure 3.1), which is just past the point of curvature, the average position is 0.074 meters towards the inside of the curve, and at position C2 the average position is 0.172 meters towards in the inside. As the driver gets to the center of the curve (position CC), the average lane position is 0.223 meters to the inside. Drivers then begin shifting slightly away from the inside direction of the curve at position C4 to 0.205 meters towards the inside of the curve from the center of the lane. Then drivers continues moving back towards the center of the lane at positions C5 and the PT (0.124 and 0.066 meters toward inside from the center of the lane, respectively). As indicated, drivers drift to the inside of curve near the center of the curve suggests that the driver may be most vulnerable to a right-side roadway departure near the center of the curve for the inside lane or for a lane departure into the other lane for an outside curve. These followed the trends of the input data.

Finally the model includes interaction indicator variables for the position in the curve when there is an inside lane encroachment that occurs in the curve. These parameters present the path a vehicle who has a lane encroachment towards the inside of the curve would see. The parameters indicate that when a lane encroachment occurs towards the inside of the curve it generally occurs near the CC where the parameters estimate the offset is shifted an additional 0.482 m towards the inside of the curve than when a lane encroachment does not occur.

The confidence intervals for both the point in curve and point in curve when there is an inside lane encroachment parameters do not overlap except at the PT and therefore a threshold can potentially be identified at which lane encroachments occur.

These parameters demonstrate that the path generally taken through a curve tends to be a flattened path with the driver being near the centerline of the lane at the beginning and end of the curve, but moving towards the inside of the curve as they reach the center. The path drivers follow when a lane encroachment towards the inside of the curve occurs is shifted significantly towards the inside of the curve throughout the whole curve. Figure 3.2 illustrates these paths.



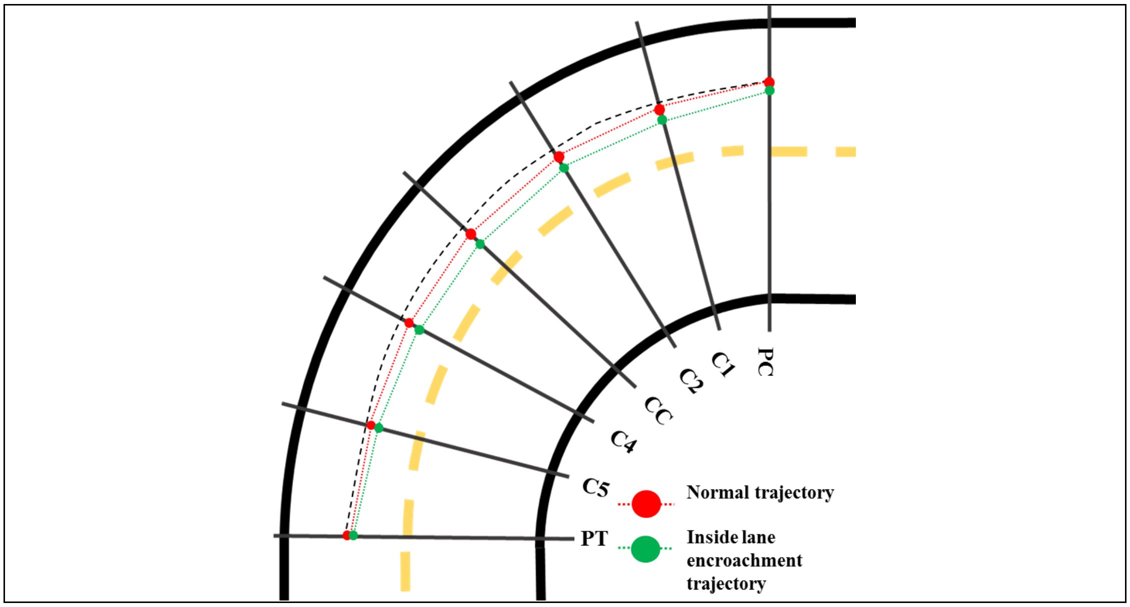


Figure 3.2 Parameter estimates of vehicle trajectories

3.6 Summary and Conclusions

The objective of this research was to develop a conceptual model of curve driving. Understanding how a driver negotiates a curve during various situations provides insight into not only how characteristics of the roadway, driver, and environment potentially influence how a driver drives, but also the areas that can lead to lane departures. Knowing how much drivers normally deviate in their lane could potentially have implications on policy or design such as determining lane widths and shoulder widths.

A linear mixed effects model was developed to assess changes in lane position as the driver negotiates the curve and results were reported. Data for several positions upstream and along the curve were sampled from the time series data. Lane position was modeled as the offset of the center of the vehicle from the center of the lane.

The model found a correlation between small radius curves and shifts towards the inside of the curve, which had been seen previously in the research *(4,5,6,7)*, Results indicate that lane position within the curve is correlated to lane position upstream of the curve. The model also found that drivers who glanced down from the roadway were associated with a shift away from the center of the lane towards the inside of the curve. A driver who looked down at a particular point within the curve shifted 0.08 meters towards the inside of the curve compared to if they had not been looking down. Additionally if the driver was distracted in the prior section it also correlated to a shift towards the inside of the curve by approximately 0.05 meters. This supports the role of distraction in lane keeping.

Additionally, the model found a large shift (from 0.16m to 0.48m depending on curve position) towards the inside of the curve when a lane encroachment towards the inside occurred in the curve, compared to when one does not occur. The larger shifts occurred in the first half and just past the center of the curve, with the largest shift occurring at the center of the curve (CC). This suggests that drivers may be particularly vulnerable to roadway departures at certain points in the curve negotiation process and supports previous findings *(5,13,14,15).*

Downward glances, distractions and position within the curve indicate that drivers may be more vulnerable to a lane departure at certain points within the curve. As a result, countermeasures such as rumble strips, paved shoulders, and high-friction treatments may reduce the consequences of variations in lane position through the curve.

Similar to the models developed in Chapter 2, this model found similar magnitude for the effect of offset 100 m upstream. Driver’s downward glance was found to have a smaller affect in this model than the once in Chapter 2, but still a change to the offset in the same direction. The offsets at each point in the curve followed a similar path as those in the models developed previously; however, the changes between offset at each point in the curve were found to be quite smaller than in the model developed in Chapter 2. This may be due to having more data and being able to determine more accurate estimates. Some of the roadway characteristics which were found to be significant in the models developed in Chapter 2 were not found to be significant here which may be due to the larger sample of curves and drivers which would make it harder to pick out specific variables as well as the inclusion of random effects for drivers and curves which may have influenced these some in those previously developed models.

3.6.1 Limitations

The main limitation of this analysis was sample size. Reliable offset data were only available in a subset of the vehicle traces that were reduced. As a result, the number of driver types and roadway features that could be modeled was limited. Increasing the sample size and focusing on including curves with the roadway features of interest could potentially lead to a relationship being established. Additionally, for this study only up to 100 m of upstream data were included as opposed to 300 m in Chapter 2 which helped to increase the sample size as well by not excluding those with inaccurate offset data in the upstream areas which were not utilized in the model.

The face and in-cabin video at times had to be coded based solely on head movements as eyes were obscured due to the drivers wearing sunglasses or poor quality and grainy video. This may have resulted in minor glances such as rear-view mirror or steering wheel being missed. It was decided to include these in the analysis in order to be able to include nighttime driving and have as much data as possible. While these minor glances may have been missed, major distractions and glances which are associated with a head movement were picked up and these minor glances were not found to be significant anyway.

3.7 Acknowledgements

This work was sponsored by the Federal Highway Administration in cooperation with the American Association of State Highway and Transportation Officials, and it was conducted in the Strategic Highway Research Program, which is administered by the Transportation Research Board of the National Academies.

3.8 References

1. Glennon, J.C., T.R. Neuman, and J.E. Leisch. *Safety and Operational Considerations for Design of Rural Highway Curves.* Report FHWA/RD-86-035. Federal Highway Administration, Washington, D.C., 1985.

2. Cheung, J. Horizontal Curve Safety. 2010. http://safety.fhwa.dot.gov/roadway_dept/horicurves/cmhoricurves/horiz_curve.pdf. Accessed January 2012.

3. Oneyear, N., S. Hallmark, S. Tyner, D. McGehee, C. Carney. Development of a Conceptual Model of Curve Driving for Rural Two Lane Curves Using SHRP 2 Naturalistic Driving Data. Conference proceedings from the 5th International Symposium on Highway Geometric Design, Vancouver BC, 2015.

4. Felipe, E. and F. Navin. Automobiles on Horizontal Curves: Experiments and Observations. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1628, 1998, pp. 50–56.

5. Stodart, B.P. and E.T. Donnell. Speed and Lateral Position Models from Controlled Nighttime Driving Experiment. *ASCE Journal of Transportation Engineering*, Vol. 134, No. 11, 2008, pp. 439–449.

6. Lamm R., E. M. Choueiri, J.C. Hayward, and A. Paluri. Possible Design Procedure to Promote Design Consistency in Highway Geometric Design on Two-Lane Rural Roads. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1195, 1988, pp. 111–122.

7. Miaou, S.-P., and H. Lum. Statistical Evaluation of the Effects of Highway Geometric Design on Truck Accident Involvements. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1407, 1993, pp. 11–24.

8. Council, F.M. Safety Benefits of Spiral Transitions on Horizontal Curves on Two-Lane Rural Roads. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1635, 1998, pp. 10–17.

9. Zegeer, C.V., J.R. Stewart, F.M. Council, and D.W. Reinfurt. *Safety Effects of Geometric Improvements on Horizontal Curves*, University of North Carolina, Chapel Hill, NC, 1991.

10. Antin, J. Technical Coordination & Quality Control (S06). Presented at the 8th SHRP 2 Safety Research Symposium, Washington D.C., 2013.

11. VTTI. *InSight Data Access Website SHRP 2 Naturalistic Driving Study*. https://insight.shrp2nds.us/. Accessed July 31st, 2014.

12. Smadi, O. SHRP 2 S-04A Roadway Information Database Development and Technical Coordination and Quality Assurance of the Mobile Data Collection Project. Presented at the 7th SHRP 2 Safety Research Symposium, Washington D.C., 2012.

13. Spacek, P. FahrverhaltenunUnfallgeschehen in Kurven, Fahrverhalten in Kurvenbereichen. InStitutFürVerkehrsplanung, Transporttechnick, Strassen-unEisenbahnbaur, ETH Zürich, 1998.

14. Fitzsimmons, E., V. Kvam, R.R. Souleyrette, S.S. Nambisan, and D.G. Bonett. Determining Vehicle Operating Speed and Lateral Position Along Horizontal Curves Using Linear Mixed-Effects Models. *Traffic Injury Prevention*, Vol. 14, Issue, 3, 2013, pp. 309-321.

15. Levison, W.H., J.L. Campbell, K. Kludt, A.C. Vittner Jr., I. Potts, D. Harwood, J. Hutton, D. Gilmore, J.G. Howe, J.P. Chrstos, R.W. Allen, B. Kantowitz, T. Robbins, and C. Schreiner. Development of a Driver Vehicle Module (DVM) for the Interactive Highway Safety Design Model (IHSDM). Federal Highway Administration. Report FHWA-HRT-08-019. November 2007.

16. Hallmark, S.L., N. Oneyear, S. Tyner, B. Wang, C. Carney and C. McGehee. *SHRP 2 S08D: Analysis of the SHRP 2 Naturalistic Driving Study Data*. Strategic Highway Research Program 2, Transportation Research Board, Washington, D.C., 2014.

17. Shinar, D., E.D. McDowell, and T.H. Rockwell. Eye Movements in Curve Negotiation. *Human Factors*, Vol. 19, 1977, pp. 63–71.

18. Angell, L., J. Auflick, P.A. Austria, D. Kochhar, L. Tijerina, W. Biever, T. Diptiman, J. Hogsett, and S. Kiger. *Driver Workload Metrics*. National Highway Traffic Safety Administration, US Department of Transportation, Washington, D.C., 2006.

19. Hallmark, S., Y. Hsu, L. Boyle, A. Carriquiry, Y. Tian and A. Mudgal. *Evaluation of Data Needs, Crash Surrogates, and Analysis Methods to Address Lane Departure Research Questions using Naturalistic Driving Study Data.* Strategic Highway Research Program Report S2-S01E-RW-1. Transportation Research Board, Washington D.C., 2011.

20. Pinheiro, J. C., and D.M. Bates. *Mixed-Effects Models in S and S-PLUS*. Springer Verlag, New York, 2000.

Appendix 3: Random Intercepts

Random Effects

Table A3.1 DriverID

Driver ID	Intercept	Driver ID	Intercept
3	0.003933	489515	-0.00272
5	-0.0007	489604	-0.00146
173	0.002814	489784	0.00422
179	0.007266	494390	0.000439
204	-0.00904	494464	-0.00587
229	-0.0039	495440	0.002223
314	0.004188	495466	0.004762
601	-0.00106	495497	-0.00236
820	0.000884	495876	0.003751
935	-0.01339	495990	0.000634
1163	0.002611	496523	-0.00287
1414	-0.00329	496528	-0.00117
1654	-0.00331	496852	-0.003
13633	-0.00432	497016	-0.0023
13647	0.00395	497061	0.005361
13921	0.001716	497104	0.001237
14102	0.00848	497111	0.001136
14664	0.008106	497185	-0.00291
15142	0.006588	497227	0.000416
15285	-0.0044	497650	-0.00165
15519	-0.00703	497781	-0.00591
16070	-0.00159	502097	0.011846
16260	0.003734	502640	0.01486
16863	-0.00233	502879	-0.00575
17653	0.003988	502931	0.011425
368046	0.006152	505061	-0.00392
368199	-0.00134	505211	0.000119
368513	0.002088	505247	-0.00961
368717	-0.00779	5080259	-0.00279
368799	-0.00184	5080732	0.005842
368822	-0.00588	5080779	-0.00818
368948	-0.01637	5080845	0.00728
489058	0.009482	5081247	0.001999
489073	-0.0013	5081802	-0.00217

Table A3.2 CurveID in DriverID

Driver ID/CurveID	Intercept	Driver ID/CurveID	Intercept	Driver ID/CurveID	Intercept
3/NY46A	0.015298	368822/NY52D	-0.06643	497781/NY46A	-0.00321
3/NY51A	-0.02111	368822/NY62A	-0.02523	497781/NY46C	-0.19941
3/NY55A	0.061235	368948/PA16A	-0.11897	497781/NY48A	-0.00447
5/NY23A	-0.00989	368948/PA16G	-0.01542	497781/NY52C	0.003174
173/NY69A	0.039653	368948/PA29A	0.045146	497781/NY52D	0.120653
179/FL11a	0.10238	368948/PA29B	-0.05737	502097/IN27A	0.024108
204/NY17A	-0.06421	368948/PA29C	-0.08402	502097/IN44A	0.040062
204/NY17C	-0.15072	489058/IN44C	0.069637	502097/IN44C	0.009778
204/NY18A	-0.04063	489058/IN44E	0.046832	502097/IN44D	0.028763
204/NY18B	0.128183	489058/IN44G	0.037617	502097/IN44E	0.026714
229/FL12a	-0.0549	489058/IN44I	-0.00942	502097/IN44F	0.001381
314/NY18A	0.059008	489058/IN44J	-0.02956	502097/IN44G	-0.01672
601/FL4a	-0.0149	489058/IN44K	0.018499	502097/IN44H	0.065801
820/FL1A	0.012457	489073/NC20A	-0.04273	502097/IN44I	0.006717
935/PA16A	-0.07195	489073/NC20B	0.024437	502097/IN44J	0.021521
935/PA16D	0.008276	489515/NY23A	-0.02063	502097/IN44K	-0.0412
935/PA16E	0.014687	489515/NY32A	-0.0177	502640/IN11A	0.008177
935/PA16G	0.038526	489604/PA29A	-0.01766	502640/IN11B	0.078014
935/PA29A	-0.06607	489604/PA29C	-0.00297	502640/IN11C	0.06478
935/PA29B	-0.10667	489784/PA29A	0.020208	502640/IN11D	0.015458
935/PA29C	-0.00549	489784/PA29C	0.03925	502640/IN11G	0.082757
1163/IN13B	0.036787	494390/NC17A	0.006189	502640/IN11H	0.003361
1414/IN27A	-0.04633	494464/PA1A	-0.14262	502640/IN11I	-0.10481
1654/IN15C	-0.04662	494464/PA1B	0.025911	502640/IN11K	-0.01394
13633/NY23A	-0.06086	494464/PA1C	0.018422	502640/IN11L	0.075584
13647/NC17A	0.055663	494464/PA1D	0.019006	502879/NY63A	-0.08107
13921/PA16B	-0.00582	494464/PA1E	-0.00339	502931/NY51A	0.038341
13921/PA16E	0.0175	495440/NY17B	0.03133	502931/NY51B	0.052748
13921/PA16F	0.031625	495466/NY60A	0.067102	502931/NY51C	-0.02343
13921/PA16H	-0.01912	495497/PA29A	-0.03186	502931/NY52C	0.012717
14102/NC3A	0.119481	495497/PA29C	-0.0014	502931/NY52D	0.030321
14664/NY13A	0.114524	495876/NY32A	0.052848	502931/NY55A	0.050284
14664/NY13B	-0.00031	495990/NC7E	0.008938	505061/NY65A	-0.0553
15142/NY6B	0.042953	496523/NY69A	-0.04039	505211/NY51A	0.065609
15142/NY6C	0.049873	496528/NY48A	-0.02712	505211/NY51C	-0.01189
15285/NY69A	-0.06202	496528/NY62A	0.010649	505211/NY52C	-0.06635
15519/IN11A	0.014705	496852/IN44A	-0.0423	505211/NY52D	0.03494
15519/IN11B	0.014606	497016/NY61A	-0.03239	505211/NY69A	-0.02063
15519/IN11C	0.019976	497061/IN11A	0.02078	505247/NY15A	-0.03351
15519/IN11D	-0.00979	497061/IN11D	0.054753	505247/NY17A	-0.03739
15519/IN11G	-0.07666	497104/NY69A	0.017433	505247/NY17C	-0.07295

15519/IN11H	-0.04076	497111/IN44C	-0.0626	505247/NY69A	0.008476
15519/IN11I	-0.00236	497111/IN44E	0.080668	5080259/PA29C	-0.03937
15519/IN11K	-0.01445	497111/IN44G	-0.06373	5080732/IN13A	0.06549
15519/IN11L	-0.00439	497111/IN44I	0.027262	5080732/IN13B	0.001596
16070/NY69A	-0.02244	497111/IN44J	-0.00631	5080732/IN77A	0.010381
16260/NY65B	0.052611	497111/IN44K	0.040721	5080732/IN77B	0.004856
16863/NY69A	-0.03285	497185/PA24A	0.02014	5080779/IN13A	-0.0934
17653/NY41A	0.056188	497185/PA24C	0.058667	5080779/IN1A	-0.02212
368046/PA29A	-0.00142	497185/PA29A	-0.03108	5080779/IN1B	0.018697
368046/PA29C	0.088107	497185/PA29C	-0.06756	5080779/IN3A	0.066231
368199/NC16D	-0.01895	497185/PA30D	-0.02116	5080779/IN3D	-0.06119
368513/IN27A	0.029426	497227/IN44F	-0.01146	5080779/IN3E	-0.0052
368717/PA29B	-0.03724	497227/IN44I	0.090379	5080779/IN77A	-0.01613
368717/PA29C	-0.07247	497227/IN44J	-0.07306	5080779/IN77B	0.022737
368799/NY14A	0.046741	497650/NC7A	-0.00666	5080779/IN77D	-0.04614
368799/NY62A	-0.07266	497650/NC7B	0.029116	5080779/IN8A	0.021264
368822/NY32A	-0.01502	497650/NC7C	0.021533	5080845/IN1A	0.113979
368822/NY32B	0.014838	497650/NC7D	-0.11952	5080845/IN3A	-0.0145
368822/NY46A	0.006971	497650/NC7E	0.032879	5080845/IN3E	0.003103
368822/NY46B	-0.00256	497650/NC7F	0.019341	5081247/NY64C	0.028173
368822/NY51A	0.004557			5081802/NY69A	-0.03058

CHAPTER 4: PREDICTION OF LANE ENCROACHMENT ON RURAL TWO LANE CURVES USING THE SHRP 2 NATURALISTIC DRIVING STUDY DATA

A paper to be submitted to the *Transportation Research Record*

Nicole Oneyear, Shauna Hallmark, Cher Carney, and Dan McGehee

Abstract

Lane departure crashes on horizontal curves accounted for approximately 28% of all fatal crashes in 2008. Curves have been found to have a three times higher crash rate than tangent sections. Therefore addressing crashes on rural two lane curves, specifically run off the road crashes, remains a priority for our local, state and national roadway agencies. Previous research has been conducted looking at roadway and environmental factors and to a limited extent driver factors in lane departure crashes. However almost no research has addressed the interaction of these three variables and the risk of lane departure.

This study utilized data from the SHRP 2 naturalistic driving study and roadway information database to develop a mixed effect logistic regression model to predict the likelihood of a lane encroachment towards the inside of the curve based on driver, environmental and roadway factors. The model found that direction of the curve, vehicle offset from the center of the lane and amount over the advisory speed limit all increased odds of a lane departure crash.

Additionally two other models were developed using linear mixed effects models which predicted speed and offset at the point of curvature using the roadway, driver and environmental factors. The model to predict speed at the PC found the drivers speed and acceleration at 100 m upstream of the curve to be significant factors, as well as the recommended speed of the curve (advisory speed or speed limit) and a driver’s age (> 60 years). The model for offset at the PC found the driver offset at 100 m upstream of the curve to be significant. Presence of an oncoming vehicle at 100 m upstream and whether it was dawn/dusk were also significant. The results of the

speed and offset model could potentially be used in the lane encroachment model to predict the likelihood of a lane departure from 100 m upstream of the curve.

4. 1 Introduction

Roadway departure crashes account for approximately 87% of all curve related crashes with 76% being due to drivers leaving the roadway and striking a fixed object or over turning and the other 11% being head-on collisions (AASHTO 2008). Due to the small percentage of roadway miles curves represent, yet the large amount of crashes seen, fatal crashes tend to be overrepresented on curves. A study by Glennon et al. (1985), found that the crash rate on curves is approximately three times the rate on tangent sections. Addressing crashes on rural two lane curves, specifically run off the road crashes, remains a priority for our local, state and national roadway agencies. For instance, reducing serious injuries and fatalities due to lane departures is an area of focus in the majority of state's Strategic Highway Safety Plans (SHSP).

Previous research has addressed this topic, mainly looking at the role roadway factors affect crash risk. Radius or degree of curve **(**Felipe and Navin 1998, Stodart and Donnell 2008, Lamm et al. 1988, Miaou and Lum 1993**)**, length of curve, lane and shoulder width **(**Zegeer et al. 1991), preceding tangent length **(**Milton and Mannering 1998) and required speed reduction between tangent and curve have been found be correlated with crash risk. Environmental factors have also been studied found to play a role in roadway departure crashes. Using crash and near crash data from the VTTI 100 car study, McLaughlin et al. (2009) found that wet roads saw lane departure risk increase by 1.8 time on wet compared to dry roads, 7 times on roads with snow or ice than on dry roads, and 2.5 times more in nighttime versus daytime conditions.

Some driver behaviors have also been identified which affect roadway departure risk. These include speed selection and distractions. FHWA estimates that approximately 56% of run-off-road (ROR) fatal crashes on curves are speed related. Distracting tasks such as radio tuning or cell phone conversations can draw a driver's attention away from speed monitoring, changes in roadway direction, lane keeping, and detection of potential hazards (Charlton, 2007).

Additionally, Hallmark et al (2015a) developed logistic regression models to predict the odds of a right or left side lane encroachment on rural curves based on a variety of roadway, driver and environmental factors using the larger SHRP 2 dataset that this paper is based on. They found that the proportion of time a driver is glancing forward in the 200 m upstream of a curve, driver's gender, the curve direction, curve radius, guardrail and curve warning sign presence all affected the odds of a lane encroachment.

4.1.1 Objective

Rural curves pose a significant safety problem, especially in regards to roadway departure crashes. Research has been completed which has examined roadway factors role in rural curve safety. Additional research has been completed which studies driver and environmental roles yet it is limited. Little has been done to study the interaction of driver, environmental and roadway factors in roadway departures. The objective of this research was to first assess the relationship between driver behavior, roadway factors, environmental factors, and the likelihood of lane encroachments on rural two-lane curves. This will differ from the research previously conducted by Hallmark et al (2015a) by only including trips with accurate offset data which allows for the inclusion of additional kinematic data such as offset. More detailed driver data, such as the length of glances will also be studied. Finally, lane encroachments will be towards the inside of the curve or outside of the curve instead of right or left side. The second objective was to develop models which would predict the factors found to affect the likelihood of a lane encroachment based on driver's behavior in the upstream tangent area.

In order to accomplish these objectives, data from the second Strategic Highway Research Program (SHRP 2) naturalistic driving study (NDS) and roadway information database (RID) were utilized as they provided the necessary information on driver behavior, environmental characteristics and roadway factors.

The authors note that there is no established relationship between a lane encroachment and crash risk. Additionally, while it is generally believed that a strong correlation exists between speed and crash risk, the exact relationship is not well quantified. As a result, while both encroachment and speed are used as surrogates for crash risk, the authors understand that the safety risk is unknown.

4.2 Data

4.2.1 Data Sources

Data for this study came from two main sources. The SHRP 2 Naturalistic Driving Study and the SHRP 2 Roadway Information Database. In 2005 congress passed the second Strategic Highway Research Program (SHRP2) whose research fell into four main areas: capacity, renewal, reliability, and safety (TRB, 2015). The majority of the safety research focused on developing the largest Naturalistic Driving Study done to date along with a Roadway Information Database to complement the NDS.

4.2.1.1 SHRP 2 Naturalistic Driving Study

The study was conducted by Virginia Tech Transportation Institute (VTTI) from 2011-2014. Male and female drivers with ages ranging from 16 to 98 in six states (Florida, Indiana, New York, North Carolina, Pennsylvania and Washington) had their vehicles equipped with a data acquisition system (DAS) which collected information on trips they made over a period of six months up to two years. The DAS collected information such as speed, acceleration, and location. Additionally, four cameras which collected forward, rear, drivers face and over the shoulder video were also placed in each vehicle. Over the three years of the study approximately 3,300 participants drove over 30 million data miles or 5 million trips (Antin, 2013 and VTTI, 2014).

4.2.1.2 SHRP 2 Roadway Information Database

In conjunction with the SHRP 2 Naturalistic Driving Study, another project was conducted to collect roadway information for the main roads traveled in the NDS. The Center for Research and Education (CTRE) led the effort which used mobile data collection to collect 12,500 centerline miles of data across the six states where the NDS was focused. Data collected included information on roadway alignment, signing, lighting, intersection location and types, presence of rumblestrips and other countermeasures. In addition to the mobile data collection effort, existing roadway data collected by local agencies was leveraged to increase the data available. Additionally, supplemental data such as crash data, changes to laws, and construction projects were also collected to further strengthen the database (Smadi 2012).

4.2.2 Data Request

At the time this study was conducted, the NDS and RID were still in progress. Due to this fact there were some constraints on the data available. For instance, only about a third of the NDS data were available. Additionally some data had not been processed such as the radar. The crashes and near crashes had not been identified, and therefore surrogates needed to be used in the analysis. Finally, the RID and NDS had not been linked. Therefore data had to be manually requested. Curves were identified using the RID and then overlain with maps of initial trip locations provided by VTTI. GIS buffers were created around curves of interest and then sent to VTTI to request data. Approximately 700 curves were included in this data requested. Data were requested from all of the states in the study except WA as the bulk of their trips appeared to be urban.

Data requested included time series data for the curves as well as a tangent section 0.5 miles upstream of the point of curvature (PC) and 0.5 miles downstream of the point of tangent (PT). In some cases, the tangent distance and subsequent curves overlapped.

Over 4,000 traces were originally identified and then through a series of steps the sample was reduced to approximately 787 traces. Of these only a subset had driver glance and distraction data due to time constraints. A more detailed description of the data request process can be found in Hallmark et al. 2015b.

4.2.3 Data Reduction

Data used in the study fell into four main categories: roadway, vehicle, driver and environmental. A brief description of the data collected in each category is summarized below. A more detailed summary of the data reduction process can be found in Hallmark et al. 2015b and Appendix A.

4.2.3.1 Roadway

Roadway data were gathered primarily from the Roadway Information Database. Data for curves not collected as part of the SHRP2 RID or for data not included in the RID were collected using Google Earth and verified using the forward NDS video. Roadway data collected included information on curve alignment (length, radius), cross-section (lane width, presence and type of shoulder, super elevation), countermeasure presence (rumblestrips, raised pavement markings, guardrail, curve advisory signs, chevrons), type of curve (S-curve, compound curve) along with other pertinent information (speed limits, curve advisory speeds, pavement and pavement marking conditions, distance between curves, a measure of roadway furniture and approximate sight distance).

4.2.3.2 Vehicle

Time series data at a sampling of 0.1 second were provided for each event requested. These data provided information on the vehicles speed, acceleration (lateral and longitudinal), offset from the center of the vehicle lane, the yaw rate as well as GPS coordinates for each second which allowed us to geo-locate each trace and pick out when the driver was at the PC, PT, and other distances within the curve as well as the distance upstream. Additional information on the vehicle type and track width were also provided.

4.2.3.3 Lane encroachment

Due to the fact that the crash-near crash data were not available at the time this study was conducted, a surrogate measure was utilized. While time to collision is one of the most widely used surrogates, it was not able to be utilized in this study with the NDS data in its current form. Lane deviation has been used as a crash surrogate for both road departure crashes and crashes due to distraction (Donmez et al. 2006). Previous studies have often used lateral placement or encroachment to evaluate rumble strips (Porter et al 2004, Hallmark et al 2011 and Taylor et al 2002).

Lane deviation was provided in the DAS time series data as offset from the lane center. Other metrics such as distance from the left or right lane line could also be calculated using additional lane position variables such as lane width. However there were a number of issues that limited the number of traces where lane position was viable throughout the entire curve. This was due to noise being present in the data, which is expected with data collection efforts of this scale as well as due to the machine visioning algorithm in the DAS. It depends on lane lines or differences in contrast between the roadway edge and shoulder in order to establish position so when discontinuities (such as breaks in the lines due to intersection or lane lines being obscured) in lane lines occur, offset is reported with less accuracy. As a result, lane offset could not be reliably used as a surrogate and therefor it was determined that encroachments, or a lane line crossing would be used instead.

For the likelihood prediction model "encroachment" was used as the dependent variable. A right-side encroachment was defined as the right side of the vehicle crossing the right edge line (when present) or the estimated boundary between the lane and shoulder (when lane lines were not present). A left-side encroachment is defined as the left side of the vehicle crossing the centerline. In all cases, the centerline was visible. An encroachment was determined to have occurred when at least two of the following criteria were present:

- Vehicle edge is 0.2 meters beyond edge line/centerline/lane–shoulder boundary
- >= 0.2 g lateral acceleration is present
- Edge line/centerline/lane–shoulder boundary crossing is visually confirmed using the forward view.

These right and left-side encroachments were then redefined into inside encroachments and outside encroachments. An inside encroachment was when the encroachment was towards the inside of the curve. Therefore for right-handed (inside) curves it would be a right-side encroachment and for left-handed (outside) curves it would a left-side encroachment. For outside encroachments, the opposite was true.

4.2.3.4 Driver

The age of the driver at the time of the trip as well as the driver's sex was provided along with the time series data for each trip. Additionally kinematic driver data were collected including approximate glance location as well as any visual distraction. These kinematic data were reduced at the VTTI secure data enclave using a tool they developed which allowed for the analyst to code the glance location and distractions while viewing the various camera views simultaneously.

Driver attention was measured by the location where a driver was focused for each sampling interval. Scan position, or eye movement, has been used by several researchers to gather and process information about how drivers negotiate curves (Shinar 1977). The majority of studies have used simulators to collect eye tracking information. Because eye tracking is not possible with NDS data, glance location was used as a proxy.

Glance locations, shown in Figure 4.1, represent practical areas of glance locations for manual eye glance data reduction. Note that Figure 4.1 does not show “over the shoulder”, “missing”, and “other” eye glance locations. “Missing” was used when a driver’s face was obscured due to glare or when a glance was not able to be determined. These were determined based on the University of Iowa team members’ extensive eye glance reduction experience. Glance locations were coded using the camera view of the driver’s face, with a focus on eye movements, but taking into consideration head tilt when necessary.

Potential distractions were determined by examining both the view of the driver’s face and the view over the driver’s right shoulder, which showed hands on/off the steering wheel. Distractions were identified when drivers took their eyes off the forward roadway. Potential distractions included the following:

- Route planning (locating, viewing, or operating)
- Moving or dropped object in vehicle
- Cell phone (locating, viewing, operating)

- IPod/MP3 (locating, viewing, operating)
- Personal hygiene
- Passenger
- Animal/insect in vehicle
- In-vehicle controls
- Drinking/eating
- Smoking

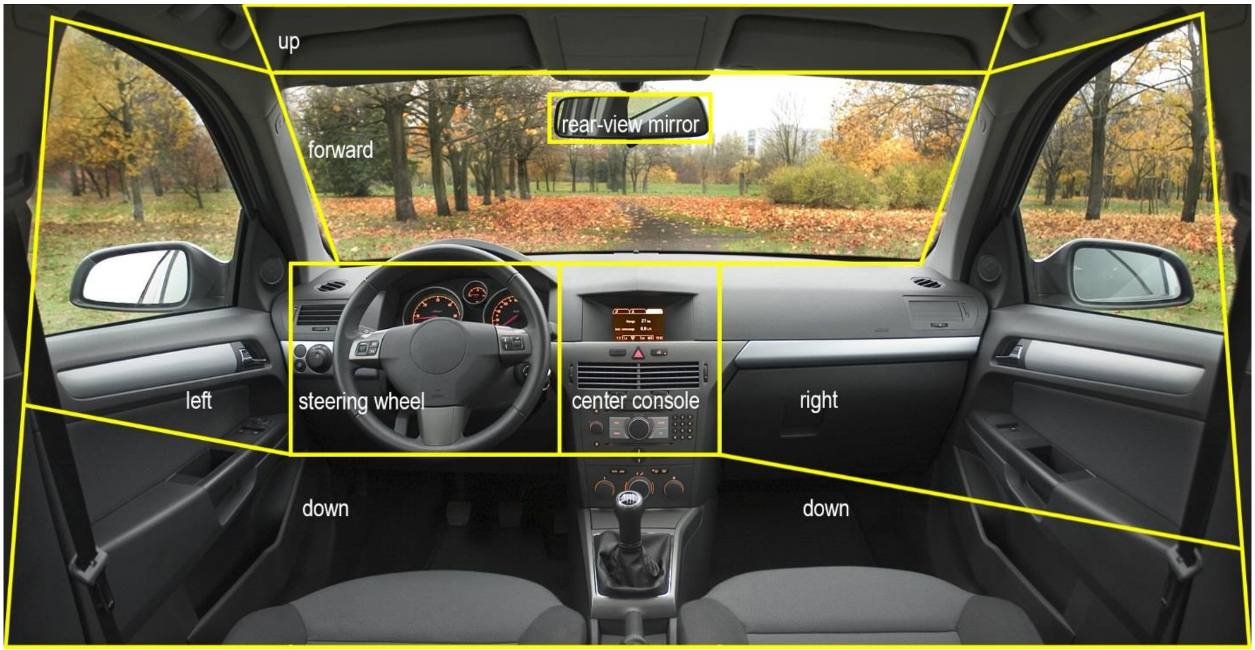


Figure 4.1 Glance Locations

Glance location and distractions were coded for each trace. The data reductionist indicated each time the glance location changed, and the data reduction tool recorded the time stamp. Similarly, the start and end times for distractions were also recorded.

Glance location was further reduced to indicate time spent in “eyes-off-roadway” while engaged in roadway-related tasks or “eyes-off-roadway” engaged in non-roadway-related tasks based on data coding used by Angell et al. (2006). Roadway-related glances or situation awareness (SA) included glances to the left mirror, steering wheel, and rear-view mirror. Angell et al (2006) included glances to the right mirror. However, glances to the right mirror are not likely to be as common because drivers are not expecting vehicles to the right and it was difficult to distinguish glances to the right mirror from other right locations. Consequently, all glances to the right were considered to be non-roadway-related.

Glances to other locations are defined as non-roadway-related (NR). Additionally, when glances to roadway-related locations were also associated with a distraction, it was determined that these glances were likely to be non-roadway-related and were coded as such. For instance, a driver who was texting and glancing at the steering wheel was likely to be looking at a cell phone being held on or near the steering wheel rather than at the speedometer.

The drivers glance location and the presence of a distraction at 100 m upstream and at the CC were coded for use in the study. Additionally it was coded if the driver was distracted or had a non-roadway related glace at any time in the 100 m upstream or in the curve.

4.2.3.5 Environmental

Information on the environmental data were collected mainly through the forward video of each trace. Data collected included the presence of other vehicles (oncoming or following), the roadway surface condition (dry, wet and raining, snowy), the lighting (day, dusk/dawn, nighttime with no lights, nighttime with roadway lighting) and visibility (high and low).

4.2.4 Data Sampling

Data were aggregated in this study by trace. A trace was one trip through one curve. Roadway and environmental data were sampled once per trip. The driver and vehicle data were sampled at multiple places: 100 meters upstream of the curve, at the PC and at the CC. These locations were chosen based off previous research. The upstream distance of 100 meters was chosen as it was right the approximate boundary between the approach and the curve discovery area as defined by Campbell et al (2012). The PC and CC were used as they are commonly used data points in curve modeling. For all of the time series and driver glance and distraction data were smoothed as there was quite a bit of noise present. These data were smoothed using a moving average over 0.5 seconds.

For the 100 m upstream location data on the acceleration, speed and offset were collected along with the drivers glance location and if they were distracted. At the PC and CC data on the vehicles offset, speed, acceleration and yaw rate, glance location and presence of a distraction were sampled. Additionally if the driver was distracted or had a non-roadway related glance at all in the upstream or curve were also sampled. Finally data were sampled on if a lane departure towards the inside or outside of the curve occurred anywhere within the curve to use as the dependent variable in our analysis.

As the analysis was including the potential effect of offset on lane encroachments, the offset data for the points selected needed to be accurate. As mentioned previously, the offset data was not always reliable. The NDS time series data included a statistic on the reliability of the offset at each reading, and VTTI provided a threshold to use to assess the accuracy. This requirement severely limited the amount of data available for the analysis as only a small portion of the data had accurate offset at the points in question. Other factors such as a limited number of samples with driver glance and distraction behavior (due to time and funding) also limited the final sample size. Additionally, some of the traces with accurate data were removed as they featured a driver who repeatedly intentionally cut the curve, often driving down the middle of the roadway.

A total of 327 trips over 95 curve driven by 68 unique drivers were included in the analysis. 32 inside lane encroachments and 8 outside lane encroachments were also included in the analysis. A summary of the roadway characteristics and driver characteristics can be seen in Tables 4.1 and 4.2. Tables 4.3 and 4.4 list a description of all of the dependent variables included in the analysis.

Table 4.1 Distribution of Curve Characteristics

radius (m)	< 500	500 to < 1000	1000 to < 1500	1500 to < 2000	2000 +	total
chevrons	7	3	0	0	0	10
some paved shoulder	17	37	11	4	4	73
rumble strips	0	3	0	0	0	3
RPM	11	24	2	2	1	40
markings obscured or not present	1	1	0	0	0	2
lighting	1	1	2	0	1	4
guardrail	6	7	3	0	0	16
total	27	46	14	4	4	95

Table 4.2 Distribution Driver Age and Gender

Age	Male	Female
16-17	6.7%	0.0%
18-20	11.6%	5.2%
21-25	6.1%	8.6%
26-30	0.6%	8.6%
30-35	5.5%	2.8%
36-40	0.0%	0.6%
41-45	6.7%	0.0%
46-50	1.8%	1.5%
51-55	3.7%	1.8%
56-60	4.3%	0.0%
61-65	0.3%	2.4%
66-70	0.0%	2.4%
71-75	0.9%	1.5%
76-80	0.6%	1.5%
80+	11.9%	2.1%

Table 4.3 Environmental, Driver, and Other Factors

	Variable	Measure	Range
Environmental/Other Factors	UpOncom, CurveOncom, PCOncom, 100Oncom	presence an oncoming vehicle is present in 100 m upstream, in curve, at Pc or at 100 m upstream of curve	0 = not present; 1 = present
	UpFollow & CurveFollow	Indicator for if driver is following another vehicle in upstream or curve	0: not following; 1= following
	UpFollowclose & CurveFollowclose	Indicator for if driver is closely following another vehicle in upstream or curve	0: not closely following; 1= closely following
	AccelX100, AccelXPC	The longitudinal acceleration (in g's) at 100 m upstream of curve and at PC	-0.10 to 0.16; -0.17 to 0.08
	UpSpeed and Upoverspeed	the speed and amount over the speed limit (mph) at 100 m upstream of curve	36.72 to 70.84 mph; -22.39 to 49.45 mph
	SpeedPC and overadvisPC	the speed and amount over the advisory speed limit (mph) at the PC	30.76 to 71.46 mph; -23.65 to 35.22
	SpeedCC and overadvisCC	the speed and amount over the advisory speed limit (mph) at the CC	9.32to 71.87 mph; -35.68 to 16.87
	Offset100	Offset from center of curve at 100 m upstream of curve (+ towards inside of curve, - towards outside)	-0.7819 to 0.7699 m
	Surface	roadway surface condition	0 = dry; 1 = wet
	Lighting	lighting conditions	0 = daytime; 1 = dawn/dusk; 2 = nighttime/no lighting; 3 = nighttime/with lighting
	Visibility	measure of visibility of forward view	0 = clear; 1 = reduced visibility; 2 = low visibility
Driver/Vehicle Factors	SubjectID	ID for driver	17 to 86 years
	Gender	Drivers gender	0 = male; 1 = female
	Age	Drovers age at time of trip	
	Forward	Indicator if glance at PC is forward	0:other glance; 1: forward glance
	SA	Indicator if situational awareness glance at PC	0: other glance; 1:SA glance
	UpNR, NR, CurveNR	Indicator if non-roadway glance in upstream, at PC and in curve	0: other glance; 1: NR glance
	DistractUp, DistractCurve	Indicator if visual distraction is present in upstream, curve	0:no distraction; 1:distraction
	DistractUp.1, DistractCurve.1	Indicator if visual distraction greater than 1 second is present in upstream, curve	0:no distraction; 1:distraction
	DistractPC	Indicator if visual distraction is present at PC	0:no distraction; 1:distraction
	Track	Vehicle track width in m	1.6 to 2.02 m
	VehClass	Class of the vehicle	1=Car; 2=Pickup, 3=SUV Crossover

Table 4.4 Roadway Factors

	Variable	Measure	Range
Roadway Factors	CurveID	ID number unique for each curve	
	Direction	curve direction from driver perspective	0 = inside(right); 1 = outside (left)
	Markings	visibility of pavement markings	0 = pavement markings visible; 1 = obscure
	PaveCond	pavement condition	0 for normal; 1 = moderate pavement; 2 = severe pavement damage
	Radii	curve radius	35.51 to 2244 meters
	Chevron	presence of chevrons	0 = no chevrons; 1 = chevrons
	PvdShd	presence of paved shoulders through curve	0 = not present; 1 = present
	RS	presence of rumble strips through curve	0 = not present; 1 = present
	RPM	raised pavement markers	0 = not present; 1 = present
	Guardrail	presence of guardrail through curve	0 = not present; 1 = present
	CurveWarn	presence of curve warning sign	0 = not present; 1 = present
	CAdvSpd	curve advisory speed if present	9 to 22 mps (20 to 50 mph)
	Speedlimitup	tangent speed limit	18 to 27 mps (40 to 60 mph)
	Curvespeed	Curve advisory speed if present, otherwise tangent speed	9 to 27 mps (20 to 60 mph)
	CurveType	type of curve	0 = normal; 1 = S-curve; 2 = compound
	SecondcurveS	Indicator of second curve encountered in an S-curve	0=not 2^{nd} S-curve, 1=2^{nd} S-curve
	UpDist	distance to nearest upstream curve	42 to 9,915 meters
	Super	super elevation of curve (%)	1.5 to 10.6%
	Length	Length of curve in m	56 to 797 m
	Markings	condition of pavement markings	0 = highly visible; 1 = visible; 2 = obscured or not present
	LaneWidth	The width of the lane in m	2.3 to 3.8 m

4.3 Analysis

4.3.1 Lane Encroachment Probability

Logistic regression was used to model the probability (odds) of having an inside lane encroachment for each trace, indexed by i as a random variable Y_i, which follows a Bernoulli distribution with probability of departure, p_i .

Logistic regression was used as it evaluates the association between a binary response, in this case whether a lane departure occurred or not, and explanatory variables. The output of the model are easily interpreted odds ratios. Odds ratios are the probability that an event happens in relation to the probability that it does not happen.

Due to the limited number of traces with a lane encroachments towards the outside of the curve this was not modeled, and only inside lane encroachments were. The glmer() function in the lme4 package in R was used to model a mixed logistic regression. A mixed model was used as we have multiple samples from some drivers and for each curve, which can be accounted for as random effects. The model was fit utilizing the Alkaline Information Criteria (AIC) statistic to determine the best fit model for the data as well as making sure parameters were significant. Additionally, ANOVA tests were used to determine if inclusion of a parameter or random effect significantly improved the model.

4.3.2 LME models

The logistic regression model found that both offset at the PC as well as the amount over the speed limit were significant factors in the probability of a lane encroachment towards the inside of the curve. Having models to predict these two values based on variables from upstream driving as well as roadway and environmental characteristics could help to determine upstream whether a lane departure is likely to occur. This prediction before entering the curve could allow for additional time to make corrections.

Linear mixed effects models were used to develop models for the speed at the PC and the offset at the PC. The lmer() function in the lme4 package in R was used to develop these models. A linear mixed effects model was utilized for this analysis as it allowed for having multiple samples from the same curves and same drivers which were accounted for through random effects. Models were run with variables being manually added and removed using the AIC statistic again to determine the best fit model and making sure variables were statistically significant at a 95% confidence. ANOVA tests were again utilized to determine if the inclusion of a variable significantly improved the models fit. In the case of factor variables however, sometimes levels of the factor were included even if they were not significant as overall they inclusion of the other factors increased the fit. This was true in the best fit offset model. Additionally, other tests were conducted to make sure the model met linear assumptions as well as to make sure there was no multi-collinearity in the variables nor any autocorrelation in the errors.

4.4 Results

4.4.1 Lane Encroachment Logistic Regression Model

The log odds of inside encroachment were modeled as follows. Inside encroachments are encroachments towards the center of the curve; for a right curve the encroachment would be crossing the outside lane line onto the shoulder, while the left curve it would be over the centerline. None of the often cited roadway factors such as radius were found to be significant factors in the model.

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + B_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \gamma_i$$

$$\gamma_i \sim Normal(0, \sigma_c^2)$$

Where:

x_1 = amount over the advisory speed or speed limit (if no advisory speed) at the PC in mph

x_2 = offset from the center of the curve at the PC in meters (+ towards the inside of curve, - towards the outside of curve)

x_3 = dummy variable for the direction of the curve (0 is left (outside); 1 is right (inside))

γ_i = random effect for curve

Parameter estimates, p-values, and 95% Wald confidence intervals are shown in Table 4.5.

Table 4.5 Parameter Estimates for Inside Encroachments

Parameter	Estimate	p-value	2.5%	97.5%
$\boldsymbol{\beta_0}$	-5.8255	0.0001	-8.7816	-2.8695
$\boldsymbol{\beta_1}$	0.1054	0.0363	0.0067	0.2041
$\boldsymbol{\beta_2}$	4.0321	0.0003	1.8463	6.2179
$\boldsymbol{\beta_3}$	1.7174	0.0125	0.3703	3.0645
@$\boldsymbol{\sigma_c^2}$	7.3097	n/a	n/a	n/a

The interpretation of these parameters is as follows: for a 1 unit increase in the value of x_i, the odds of a lane encroachment changes by a factor of e^{β_i}. These can also be scaled to any level, so for instance if you wanted to look at a 10 unit increases effect on the odds of a lane encroachment on would use $e^{10*\beta_i}$. Odds ratios and 95% Wald confidence intervals are shown in Table 4.6.

Table 4.6 Confidence Intervals for Inside Encroachments

Variable	Odds Ratio Est.	2.5%	97.5%
Over advisory speed at PC	1.1112	1.0067	1.2264
Offset at PC	56.3792	6.3363	501.6487
Direction	5.5700	1.4482	21.4237

As noted, for every mph over the curve advisory speed limit a driver is 1.1 times more likely to have an inside encroachment. For every meter away from the center of the lane towards the inside of the curve at the PC increases odds of an inside lane encroachment by 56. Looking at a more realistic shift of 0.1 meters towards the inside direction of the curve from the center of the lane would increase odds of a lane encroachment by 1.5. Shifting 0.1 meters to towards the outside of the curve would decrease odds of an inside lane encroachment by 0.67. Odds of an inside lane encroachment is 5.6 times more likely for right (inside) curves compared to left (outside) curves. An output of the random effects intercepts can be seen in Appendix 4.

Inside encroachments are likely to be drivers who "cut the curve" or drive as though the curve has a larger radius than it actually does. Although it is difficult to determine driver intent, in several cases the driving manner as evidenced in the forward videos strongly suggested that the driver was intentionally crossing the centerline. These observations were removed. However it was not always possible to distinguish between intentional and unintentional lane crossings so some intentional encroachments may be included in the model.

4.4.2 Speed at Point of Curvature Linear Mixed Effects Model

The linear mixed effects model for speed at the PC can be seen below with parameter estimates in Table 4.7. Speed at the PC was used as the dependent variable instead of the amount over the advisory speed at the PC due to a better fit being able to be achieved. If the speed is known along with the advisory speed (or speed limit if no advisory speed is posted) one can then determine the amount over to use in the logistic regression found above.

$$Y_{IJ} = \beta_0 + \beta_1 x_1 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \gamma_i$$

$$\gamma_i \sim Normal(0, \sigma_c^2)$$

Where:

x_1 = speed at 100 meters upstream of curve in mph

x_2 =dummy variable for if the driver is over 60 (0 is 60 and under, 1 is over 60)

x_3 = curve advisory speed (or speed limit if no advisory speed limit exists) in mph

x_4 = Longitudinal acceleration at 100 meters upstream of curve in gs

γ_i = random effect for curve

Table 4.7 Parameter Estimates for Speed at PC

Parameter	Estimate	p-value	2.5%	97.5%
$\boldsymbol{\beta_0}$	-5.3709	0.0004	-8.3257	-2.4153
$\boldsymbol{\beta_1}$	0.9339	<0.0001	0.8955	0.9729
$\boldsymbol{\beta_2}$	-0.6960	0.0135	-1.2580	-0.1455
$\boldsymbol{\beta_3}$	0.1657	<0.0001	0.1157	0.2159
$\boldsymbol{\beta_4}$	16.6563	<0.0001	9.4597	23.7439
$\boldsymbol{\sigma_c^2}$	1.606	n/a	n/a	n/a
$\boldsymbol{\sigma_{residual}^2}$	2.941	n/a	n/a	n/a

The model includes four variables along with random effects for curves as drivers were not found to be significant. The model predicts that the drivers speed at the PC will be approximately 0.934 times that at 100 m upstream. The model also found a correlation that drivers over 60 on average tend to drive approximately 0.7 mph slower than those drivers under 60. The model also predicts that for higher curve advisory speeds (or speed limits if no advisory speed exists) that drivers will have a higher speed entering the curve, which is expected. Finally the model found that if drivers are accelerating at 100 meters upstream of the curve their speed entering the curve will be larger than if they were not. Appendix 4 includes the random intercepts for this model.

4.4.3 Offset at Point of Curvature Linear Mixed Effects Model

The model for offset at the PC can be seen below, with parameter estimates, significance and confidence intervals in Table 4.8. A negative offset corresponds to moving from the center of the lane towards the outside of the curve while a positive offset corresponds to moving from the center of the lane towards the inside of the curve. The best fit model included five variables, two of which are factors, along with an intercept and random effects for curves.

$$Y_{IJ} = \beta_0 + \beta_1 x_1 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \gamma_i$$

$$\gamma_i \sim Normal(0, \sigma_c^2)$$

Where:

x_1 = offset from centerline in meters at 100 meters upstream of curve (+ towards inside of curve – towards outside of curve)

x_2 = dummy variable for the dusk or dawn (0 is day or night; 1 is dusk or dawn)

x_3=

x_{3a} = factor variable for oncoming vehicle at 100 m upstream for outside curve (1: oncoming vehicle present)

x_{3b} = factor variable for oncoming vehicle at 100 m upstream for inside curve (1: oncoming vehicle present)

γ_i = random effect for curve

Table 4.8 Parameter Estimates for Offset at PC

Parameter	Estimate	p-value	5%	95%
$\boldsymbol{\beta_0}$	0.0331	0.1245	-0.0021	0.0686
$\boldsymbol{\beta_1}$	0.6417	<0.0001	0.5637	0.7196
$\boldsymbol{\beta_2}$	-0.1557	0.0011	-0.2339	-0.0770
$\boldsymbol{\beta_{3a}}$	0.03643	0.5692	-0.0684	0.1411
$\boldsymbol{\beta_{3b}}$	-0.1816	0.0380	-0.3249	-0.0376
$\boldsymbol{\sigma_c^2}$	0.0171	n/a	n/a	n/a
$\boldsymbol{\sigma_{residual}^2}$	0.0536	n/a	n/a	n/a

The best fit model found that the drivers offset at 100 m upstream of the curve correlates with the drivers offset at the PC. If the driver is driving towards the direction of the inside of the curve in the upstream, they will be as well entering the curve. The model also predicts that during dawn or dusk hours drivers tend to enter the curve more in the direction of the outside of the curve than they do during the day or at night.

Finally, a factor variable was included in the model which predicted how the presence of an oncoming vehicle at 100 m upstream of the curve affected drivers offset at the PC. A factor variable was used instead of an indicator variable as depending on the direction of the curve, the response to offset is expected to be different. In both cases drivers are expected to shift away from the centerline. With the convention for determining sign of offset in our model, the response would be different. The model found only a significant effect for when drivers on an inside (right) curve encountered an oncoming vehicle at 100 m upstream of the curve. The model predicts the driver's offset at the PC will shift 0.182 meters more towards the outside of the curve (centerline) than if an oncoming vehicle were not present. This response is expected as the oncoming vehicle at 100 meters upstream would have increased their offset at that point as they would most likely shift away from the center line. Appendix 4 includes the random intercepts.

4.5 Discussion and Conclusions

The objective of this research was to assess the relationship between driver, roadway, and environmental factors and probability of a lane departure. The study first modeled the probability of an inside curve encroachment, using logistic regression at the trace level. Then linear mixed effect models were developed to assess the relationships between driving 100 meters upstream of the curve, driver and environmental factors and the lane position and speed at the PC.

The model for probability of an inside lane encroachment indicated three main factors which affect the likelihood. The model indicated that for every mph over the advisory curve speed (or speed limit if an advisory speed was not present) a driver was driving at the PC a drivers odds of an inside lane encroachment increased by 1.11. Therefore a driver exceeding the advisory speed by 5 mph would be 1.7 times more likely to have an encroachment crash than if they were going the suggested speed. The model also found that a shift of 0.1 meters towards the inside of the curve from the center of the lane at the PC would result in the odds of an encroachment increasing by 1.5. It finally noted that drivers driving on right-handed (inside) curve are 5.6 times more likely to have an inside lane encroachment than those drivers in the left-hand (outside) curves.

The author does acknowledge that each state has their own criteria for setting advisory speed limits, so there may be some bias in using this variable, however it was found to be a better predictor than drivers speed or the amount over the speed limit. If enough data were available developing state specific models may help to avoid this potential bias.

Due to both lane position and amount over the advisory speed being significant factors in the logistic regression model, models were developed to predict these based on upstream driving. Instead of modeling amount over the advisory speed, speed at the PC was used instead as a better model resulted. The results of the model could be applied to the logistic regression then if the advisory speed of the curve is known. The model found that speed and acceleration at 100 meters upstream of the curve, the curve advisory speed, and a driver being older all affected speed at the PC.

The linear mixed effects model found that offset at 100 m upstream of the curve, if a driver encountered an oncoming vehicle at 100 m upstream of the curve and if it was dusk all affected offset at the PC. Drivers on average are at 60% of the offset they are at 100 meters upstream of the curve.

The mixed effects models developed could be used in conjunction with the logistic regression model to predict a drivers likelihood of an inside curve encroachment based on their upstream driving behavior.

4.5.1 Limitations

The main limitations of this study are in regards to data. Overall, the most significant limitation is sample size and representation of different curve and driver characteristics. Over 700 potential curves were initially identified. This represented a wide range of roadway characteristics and countermeasures. However, some countermeasures, such chevrons and rumble strips, were not widely available in the study areas, and some countermeasures, such as post-mounted delineators, were not available at all. Additionally, only one-third of the full NDS data set was available for query at the time the data request was made, and data were only found for 110 curves, which reduced the number of roadway characteristics that could be included. If additional data were included representing specific countermeasures of interest as well as more accurate driver samples based off overall countries driving population breakdown, could results in models which would include the countermeasures of interest and be more representative of the population as a whole.

Additionally due to limitations with the data accuracy, specifically with the lane offset, the sample size was severely restricted. A total of 327 observations were included in the analysis. However, only 32 inside curve lane encroachments and 8 outside curve lane encroachments were present. The small number of outside lane encroachments prevented a model from being developed. Also, as the crash/near-crashes were not available, the surrogate of encroachments was used and a relationship between encroachment and roadway departure crash risk could not be established. If these more lane encroachments, crashes or near crashes were available their inclusion could significantly improve the accuracy and applicability of the models. If these data were included, more baseline data would also be needed to help provide additional insight into baseline driving and what behaviors, both kinematic vehicle and driver glance affect the likelihood of a crash.

4.6 References

Angell, L., J. Auflick, P.A. Austria, D. Kochhar, L. Tijerina, W. Biever, T. Diptiman, J. Hogsett, and S. Kiger. *Driver Workload Metrics*. National Highway Traffic Safety Administration, US Department of Transportation, Washington, D.C., 2006.

Antin, J. Technical Coordination & Quality Control (S06). Presented at the 8th SHRP 2 Safety Research Symposium, Washington D.C., 2013.

Campbell, J.L., M.G. Lickty, J.L. Brown, C.M. Richard, J.S. Graving, J. Grahm, M. O'Laughlin, D. Torbic, and D. Harwood. Chapter 6: Curves (Horizontal Alignment). NCHRP Report 600. Human Factors Guidelines for Road Systems, Second Edition. Transportation Research Board of the National Academies. Washington DC. 2012.

Charlton, S.G. The Role of Attention in Horizontal Curves: A Comparison of Advance Warning, Delineation, and Road Marking Treatments. *Accident Analysis and Prevention*, Vol. 39, 2007, pp. 873–885.

Donmez, B., L. Boyle, and J. Lee. The Impact of Driver Distraction Mitigation Strategies on Driving Performance. *Human Factors*, Vol. 48, No. 4, 2006, pp 785-804.

Felipe, E. and F. Navin. Automobiles on Horizontal Curves: Experiments and Observations. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1628, 1998, pp. 50–56.

Glennon, J.C., T.R. Neuman, and J.E. Leisch. *Safety and Operational Considerations for Design of Rural Highway Curves.* Report FHWA/RD-86-035. Federal Highway Administration, Washington, D.C., 1985.

Hallmark, S., T. McDonald, and R. Sperry. Evaluation of Rumble Stripes on Low-Volume Rural Roads in Iowa – Phase 2. InTrans Project 07-305. Institute for Transportation, Ames, IA, 2011.

Hallmark, S.L., S. Tyner, N. Oneyear, C. Carney, and D. McGehee. *Evaluation of Driving Behavior on Rural 2-Lane Curves using the SHRP 2 Naturalistic Driving Study Data.* Journal of Safety Research. Manuscript accepted for publication, 2015a.

Hallmark, S.L., N. Oneyear, S. Tyner, B. Wang, C. Carney and C. McGehee. *Analysis of the SHRP 2 Naturalistic Driving Study Data: Roadway Departures on Rural Two-Lane Curves*. Strategic Highway Research Program 2 Report S2-S08D-RW-1. Transportation Research Board, Washington, D.C., 2015b.

Lamm R., E. M. Choueiri, J.C. Hayward, and A. Paluri. Possible Design Procedure to Promote Design Consistency in Highway Geometric Design on Two-Lane Rural Roads. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1195, 1988, pp. 111–122.

McLaughlin, S. B., J. M. Hankey, S. G. Klauer, and T. A. Dingus. *Contributing Factors to Run-Off-Road Crashes and Near-Crashes*. Report DOT HS 811 079. National Highway Traffic Safety Administration, 2009.

Miaou, S.-P., and H. Lum. Statistical Evaluation of the Effects of Highway Geometric Design on Truck Accident Involvements. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1407, 1993, pp. 11–24.

Milton, J., and F. Mannering. The Relationship Among Highway Geometric, Traffic-Related Elements, and Motor-Vehicle Accident Frequencies. *Transportation*, Vol. 25, 1998, pp. 395-413.

Shinar, D., E.D. McDowell, and T.H. Rockwell. Eye Movements in Curve Negotiation. *Human Factors*, Vol. 19, 1977, pp. 63–71.

Smadi, O. SHRP 2 S-04A Roadway Information Database Development and Technical Coordination and Quality Assurance of the Mobile Data Collection Project. Presented at the 7th SHRP 2 Safety Research Symposium, Washington D.C., 2012

Stodart, B.P. and E.T. Donnell. Speed and Lateral Position Models from Controlled Nighttime Driving Experiment. *ASCE Journal of Transportation Engineering*, Vol. 134, No. 11, 2008, pp. 439–449.

Taylor, M.C., A. Baruya, and J.V. Kennedy. *The Relationship between Speed and Accidents on Rural Single-Carriageway Roads*. TRL Report TRL511. TRL, Berks, United Kingdom, 2002.

Transpiration Research Board (TRB). *Strategic Highway Research Program (SHRP 2*). http://www.trb.org/AboutTRB/SHRP2.aspx. Accessed June 16, 2015.

VTTI. *InSight Data Access Website SHRP 2 Naturalistic Driving Study*. https://insight.shrp2nds.us/. Accessed July 31st, 2014.

Zegeer, C.V., J.R. Stewart, F.M. Council, and D.W. Reinfurt. *Safety Effects of Geometric Improvements on Horizontal Curves*, University of North Carolina, Chapel Hill, NC, 1991.

Appendix 4 – Random Effects Intercepts

A4.1 Logistic Regression

Table A4.1 Logistic Regression Curve Random Intercepts

CurveID	(Intercept)	CurveID	(Intercept)	CurveID	(Intercept)
FL11A	2.2885	IN77A	-0.0649	NY51B	-0.1171
FL12A	5.8764	IN77B	-0.0303	NY51C	-0.1018
FL14B	-0.0474	IN77D	-0.0142	NY52C	3.1998
FL1A	-0.1475	IN8A	-0.0878	NY52D	1.7691
FL4A	-0.2491	NC16D	-0.0303	NY55a	3.9653
IN11A	-0.5499	NC17A	-0.4347	NY55A	1.4703
IN11B	-0.2463	NC20A	-0.0127	NY61A	2.0412
IN11C	-0.2328	NC20B	-0.1433	NY62A	-0.0370
IN11D	-0.2426	NC3A	4.3002	NY63A	3.8225
IN11G	-0.3000	NC7A	-0.3143	NY64C	-0.0408
IN11H	-0.1376	NC7B	-0.1341	NY65B	-0.2089
IN11I	-0.2162	NC7C	3.0290	NY67A	-0.0225
IN11K	-0.0948	NC7D	-0.5724	NY69A	1.4452
IN11L	-0.8541	NC7E	-0.0206	NY6B	-0.0438
IN13A	-0.1458	NC7F	1.9386	NY6C	-0.2260
IN13B	-0.3246	NY13A	-2.1721	PA16A	-0.0264
IN15C	-0.0531	NY14A	-0.0622	PA16B	-0.0834
IN1A	-0.0989	NY15A	-0.0055	PA16D	-0.0773
IN1B	-0.0034	NY17A	-0.4518	PA16E	-0.0800
IN27A	-0.2664	NY17B	-0.0023	PA16F	-0.0921
IN3A	-0.2636	NY17C	-0.0510	PA16G	-0.0584
IN3D	-0.0302	NY18A	-0.3035	PA16H	-0.0608
IN3E	-0.0321	NY18B	-0.5719	PA1B	2.2721
IN44A	-0.3862	NY23A	-0.0842	PA1C	2.5136
IN44C	-0.1234	NY32A	-0.3996	PA1D	-0.0267
IN44D	-0.0581	NY32B	-0.0675	PA1E	-0.1343
IN44E	-0.6224	NY41A	-0.0322	PA24A	-0.4691
IN44F	-0.1506	NY46A	2.3282	PA24C	-0.0536
IN44G	2.8738	NY46B	-0.3840	PA29A	-0.8401
IN44H	-0.1007	NY48A	-0.1584	PA29B	-1.3558
IN44I	-0.4823	NY51A	2.2995	PA29C	1.4290
IN44J	-0.1652			PA30D	-0.1741
IN44K	-0.4783				

A4.2 Linear Mixed Model – Speed

Table A4.2 Speed LME Curve Random Intercepts

CurveID	(Intercept)		CurveID	(Intercept)		CurveID	(Intercept)
FL11A	0.4289		IN77A	0.1305		NY51B	-0.0819
FL12A	-0.4340		IN77B	-0.1524		NY51C	-0.2387
FL14B	0.0067		IN77D	-0.5341		NY52C	0.0372
FL1A	-2.5805		IN8A	0.1966		NY52D	1.3036
FL4A	-0.1827		NC16D	0.6232		NY55a	-0.3050
IN11A	1.6629		NC17A	-0.6548		NY55A	0.0047
IN11B	-0.1778		NC20A	-0.6977		NY61A	-0.5353
IN11C	0.7621		NC20B	-0.4779		NY62A	-1.5572
IN11D	-0.6108		NC3A	0.2349		NY63A	0.4545
IN11G	0.1106		NC7A	-0.0362		NY64C	0.5890
IN11H	-0.0216		NC7B	-0.4062		NY65B	0.5762
IN11I	0.1558		NC7C	-0.8582		NY67A	-3.0351
IN11K	-1.4867		NC7D	0.9808		NY69A	1.7383
IN11L	0.7793		NC7E	-0.4279		NY6B	0.2568
IN13A	0.1679		NC7F	0.5743		NY6C	0.3638
IN13B	0.2275		NY13A	2.3994		PA16A	0.2656
IN15C	-2.1061		NY14A	-0.4831		PA16B	0.8561
IN1A	-0.5099		NY15A	-1.6356		PA16D	-0.2036
IN1B	-0.3255		NY17A	1.5675		PA16E	1.4716
IN27A	0.5643		NY17B	-0.9829		PA16F	0.9462
IN3A	-0.5537		NY17C	0.1136		PA16G	0.8041
IN3D	0.2158		NY18A	-0.8618		PA16H	0.3945
IN3E	-0.7621		NY18B	-0.0140		PA1B	-1.6621
IN44A	0.6708		NY23A	-0.1737		PA1C	-0.3240
IN44C	0.2920		NY32A	0.9031		PA1D	1.2995
IN44D	-1.0897		NY32B	-0.9686		PA1E	1.2294
IN44E	-0.0278		NY41A	0.7701		PA24A	-0.0036
IN44F	0.3611		NY46A	2.0252		PA24C	0.3390
IN44G	-0.1424		NY46B	-0.0919		PA29A	-0.5015
IN44H	-0.4644		NY48A	0.3423		PA29B	0.2850
IN44I	-0.0826		NY51A	0.0322		PA29C	0.3893
IN44J	0.9029					PA30D	0.1102
IN44K	-2.4539						

A4.3 Linear Mixed Model – Offset

Table A4.3 Offset LME Curve Random Intercepts

CurveID	(Intercept)	CurveID	(Intercept)	CurveID	(Intercept)
FL11A	0.1800	IN77A	-0.0499	NY51B	0.0049
FL12A	0.1443	IN77B	-0.0365	NY51C	0.0113
FL14B	0.0039	IN77D	-0.1003	NY52C	-0.0586
FL1A	0.0633	IN8A	0.0225	NY52D	0.2083
FL4A	0.0535	NC16D	-0.0377	NY55a	0.0365
IN11A	-0.0638	NC17A	0.0488	NY55A	0.1962
IN11B	0.0291	NC20A	-0.0477	NY61A	0.0905
IN11C	0.0109	NC20B	0.0616	NY62A	-0.1409
IN11D	-0.0606	NC3A	0.1819	NY63A	-0.0827
IN11G	0.0372	NC7A	-0.0855	NY64C	-0.0769
IN11H	-0.1213	NC7B	0.0114	NY65B	0.0635
IN11I	-0.0398	NC7C	0.1210	NY67A	-0.1474
IN11K	-0.0637	NC7D	-0.0803	NY69A	-0.0394
IN11L	0.1190	NC7E	-0.0443	NY6B	0.0519
IN13A	-0.1052	NC7F	0.0889	NY6C	0.0114
IN13B	-0.0286	NY13A	0.1030	PA16A	-0.1160
IN15C	-0.0192	NY14A	0.0590	PA16B	-0.0068
IN1A	0.0431	NY15A	0.0073	PA16D	0.0178
IN1B	-0.0759	NY17A	0.0667	PA16E	-0.0449
IN27A	-0.0407	NY17B	-0.0754	PA16F	0.0783
IN3A	0.0367	NY17C	-0.1665	PA16G	-0.0069
IN3D	-0.0447	NY18A	-0.1029	PA16H	-0.0419
IN3E	-0.0625	NY18B	0.0624	PA1B	0.0135
IN44A	0.1412	NY23A	-0.1923	PA1C	0.0722
IN44C	-0.1326	NY32A	-0.0593	PA1D	0.0050
IN44D	0.0946	NY32B	-0.0009	PA1E	-0.0161
IN44E	0.1233	NY41A	-0.0046	PA24A	-0.0802
IN44F	-0.0732	NY46A	0.0812	PA24C	-0.0268
IN44G	0.0612	NY46B	0.0128	PA29A	-0.0414
IN44H	0.1275	NY48A	-0.0519	PA29B	-0.1128
IN44I	0.0600	NY51A	0.0787	PA29C	-0.1197
IN44J	-0.0958			PA30D	0.0080
IN44K	0.1177				

CHAPTER 5: CONCLUSIONS AND DISCUSSION

5.1 General Conclusions

Road departure are a leading cause of fatal crashes on rural horizontal curves. Previous research has studied how individual roadway and environmental factors along with driver behaviors contribute to roadway departures on rural curves. Little research has been conducted to study the interaction of these three categories of factors in affecting roadway departures. Through three papers this dissertation set out to better understand how these various factors affect how drivers negotiate curves and to determine which factors may increase the risk of a lane departure.

The paper in Chapter 2 developed basic conceptual models of normal driving curve for a limited sample of rural two lane isolated curves. This analysis, which utilized generalized least squares regression to develop models for right-handed and left-handed curves which predicted a driver's lane position (modeled as offset from the center of the lane in meters). The models found that a drivers offset 100 meters upstream of the start of the curve could help predict a vehicles position at various points throughout the curve. The models were also able to predict the average path a driver would take through seven points in the curve. These estimators suggest that drivers tend to cut the curve and are more susceptible to a lane departure at certain points in the curve. The models also found that things such as glancing down or being younger (under 30) correlated with changes in lane position. The left-handed model also found that the presence of roadway features such as large paved shoulders, poor delineation and curve advisory signs possibly play a role in lane position.

The work conducted in Chapter 2 was expanded in Chapter 3 to include a larger number of curves and drivers as well as traces where lane encroachments occurred. This was accomplished by using up to 100 m of upstream driving which allowed for the inclusion of S curves as well as a larger sample of other non S-curves who had, had bad data in the 150-300 m upstream section which could now be included. A conceptual model of curve driving was developed which included a total of 323 traces for 68 unique drivers on 98 different curves which included 16 lane departures towards the inside of the curve. A single model was developed for this analysis, instead of two like in Chapter 2, as it allowed for a more robust model. The model was able to determine a difference in the offset at each point in the curve for those traces where a lane departure towards the inside of curve occurred and when it did not. The model also found a similar correlation between the driver's lane position upstream of the curve and lane position in the curve. The model also found that smaller radii, looking down and being distracted all also influenced lane position.

Chapter 4 used trace level data from the data in Chapters 2 and 3 along with some additional data to create a mixed logistic regression model which predicts the likelihood of a lane encroachment towards the inside of the curve. This model was based on a sample of 327 traces through 95 curves by 68 unique drivers. The data set included 32 inside lane encroachments and 8 outside lane encroachments. Due to the limited data for the outside lane encroachments, only inside lane encroachments were modeled. The best fit model found that the amount over the curve advisory speed (or speed limit if no advisory speed exists) at the PC, offset from the center of the lane at the PC and direction of curve all affected the likelihood of a lane encroachment.

Additional linear mixed effect regression models were developed in Chapter 4 to predict a drivers offset and speed at the PC based on upstream driving characteristics. The speed model found that a drivers speed at the PC correlates to the drivers speed and acceleration at 100 m upstream of the curve, a driver being older (60+), and the curve advisory speed (or speed limit).

The offset model found a drivers offset at the PC to be correlated to the drivers offset at 100 m upstream, the time of day (specifically if it is dawn/dusk), as well as the presence of an oncoming vehicle 100 m upstream.

5.2 Contribution to State Of The Art

The research conducted for this dissertation contributes to the state of the art by providing new insight into how driver, environmental and roadway factors interact in the negotiation of rural curves. The conceptual models developed in Chapters 2 and 3 provide new understanding of how drivers' path changes as they progress through the curve and how driver behaviors such as glancing down or being distracted affect this path. These models include a large sample of curves with smaller samples of traces through these curves where previous research has mainly looked at larger samples of traces through curves and smaller samples of curves. The paths developed all show that drivers' paths vary as they traverse a curve and are more likely to experience a lane departure near the center of the curve more than at the beginning or end of the curve. As a result, countermeasures such as rumble strips, paved shoulders, and high-friction treatments may reduce the consequences of variations in lane position through the curve.

The models in these two chapters also help to develop a great base model which can be expanded on with the inclusion of additional data to draw out more relationships. The basic framework developed for the models could be used in other studies hoping to gain more insight into how specific roadway features or driver behaviors affect negotiations be looking at more samples traces from a smaller subset of curves.

The offset model developed in Chapter 3 also determined boundaries between normal driving and lane encroachments towards the inside, the beginning of non-normal driving situations. This boundary could be used to identify events of interest (non-normal) more easily in future studies.

The prediction model developed in Chapter 4 provides odds ratios on how speed, lane position and direction of curve affect likelihood of a lane encroachment. Additionally the linear mixed effects regression models provide a means of estimating expected speed and lane position at the PC from 100 meters upstream of the curve. The results from these models can then be plugged into the logistic regression model to predict, based off upstream driving, the probability of the driver having a lane departure towards the inside of the curve. This provides a framework to expand on to develop an advanced lane departure warning system or curve speed warning system.

The insight into how speed increases odds of a lane encroachment determined in Chapter 4 can help target education. Also knowing how increases in speed effect likelihood of a lane encroachment could be used in improving speed thresholds used in dynamic curve warning signs which provide an out-of-vehicle warning.

5.3 Limitations

As mentioned in the papers above, there were a few limitation to the research that was conducted as part of this dissertation. The limitations are summarized below.

5.3.1 Data accuracy

NDS data are collected through uncontrolled field conditions and as a result noise and other data quality issues are inherently present. At the time when this project obtained data, some data had not been quality controlled and some characteristics of the data were not yet well understood. For instance, significant noise was present in variables such as offset, which is expected for large-scale data collection of this nature. It was also due to issues with the machine learning algorithm used in the DAS which depends on lane lines or differences in contrast between the roadway edge and shoulder in order to establish the position. When discontinuities in lane lines occur, offset is reported with less accuracy. Discontinuities occur due to lane lines being obscured, natural breaks being present in lane lines (e.g., turn lanes, intersections), or visibility being compromised in the forward roadway view. A moving average used to smooth the data helped to reduce some noise, but could not account for large distances of not accurate lane lines.

In other cases, variables of interest were not sufficiently available to be utilized. For instance, use of steering wheel variability has been used as an indicator of drowsiness by a number of researchers (Kircher et al, 2002; Liu et al, 2009). Since drowsiness is a likely contributor to roadway departures, ideally, a search algorithm could have utilized to identify potential drowsy driving events using a measure of steering wheel reversal. However, not all variables could be output from the OBD in all vehicles including steering wheel position which was only available for a small subset of vehicles. Additionally, although a passive alcohol detector was present, at the time data were collected it did not appear to be reliable enough to identify potential intoxicated drivers.

Additionally, the quality of the driver face video was not always clear enough to be able to see the pupil. This especially occurred at night and when the driver was wearing sunglasses. In these cases driver's head position was used to measure approximate glance location, which may have led to missing some of the more subtle glances such as looking at the rear-view mirror or at the steering wheel. Initial work by Muñoz et al 2015 using the SHRP 2 data set suggests that head position may provide a reasonable estimate of glance location. The kinematic driver data that was found to be significant in the studies, only included distractions and glancing down, which were generally, or in the case of glancing down, associated with a head movement so they would have been captured.

5.3.2 Limited sample sizes

At the time the data request for this project was made, only around one-third of the full data set was available. Time and budget constraints also limited the amount of traces where kinematic driver characteristics could be reduced. Accuracy issues with offset, which were described previously, also significantly reduced the samples for these studies as accurate offset was required. Approximately 10% of the data reduced had accurate enough offset to be included in the analysis. The limited sample size also limited the amount of driver and roadway characteristic which could be included. For instance while a large sample of curves with rumblestrips were requested, only two curves which we had reduced data for had rumblestrips. Having a larger sample size would help to answer questions that had hoped to be answered in the course of the study but were unable to be determined. For instance with enough data it is thought that the effect of countermeasures such as rumblestrips or chevrons could be determined.

5.3.3 Use of surrogates

As crash and near crash data were not available at the time the data for these studies was collected, the use of surrogates was required for the analysis. While surrogates provide some expected correlation with crashes, the exact relationship was not able to be established. Therefore the results of the research cannot be translated to risks of crashes, but to risks of lane encroachments. Having adequate data on the crashes and near crashes would allow one to develop this relationship.

5.4 Additional Research

5.4.1 Expand current models

As mentioned above, the research in this dissertation was developed using a limited supply of the SHRP 2 NDS data set. At the time the data for this research was requested, only about a third of the data were available. Additionally, the NDS and RID had not been linked, so specific roadway attributes were hard to get adequate data to analyze. Additionally, due to time and budget constraints, driver data reduction was only completed for about half of the data received. The models in Chapters 3 & 4 could be greatly improved by including additional data. With more data, specifically a better sampling of trips through curves with countermeasures of interest, may provide insight into how exactly they affect driver behavior which was a goal of the study, but was unable to be drawn out of the current data set. For instance if we have enough data from the same drivers driving through a variety of similar curves, some with a countermeasure of interest and some without, the effect of the countermeasure on curve negotiation could potentially be determined. If insight into the countermeasures effect on negotiation is able to be determined, a more targeted approach to their use could be a potential benefit.

Additionally if the crash and near-crash information were able to be added to the models, one may also be able to determine boundaries between normal driving, conflicts (lane encroachments), near crashes and crashes. Knowing these boundaries can help in the development and improvement of lane departure warning systems so less type I and type II errors occur.

5.4.2 Develop crash prediction model

As mentioned previously, a large limitation of this study is that it did not include any crash or near crash data and therefore results cannot be used to determine how lane position relates to crash risk, only encroachments. As the crash near-crash data are now available, they could be used to develop models similar to the logistic regression model developed in Chapter 4, but instead of predicting the probability of a lane encroachment, they would predict the probability of a crash or near crash. The results of this research, if robust enough, could then be used to begin developing advanced lane departure warning systems. Models such as the linear mixed effects models in Chapter 4 could then be developed so one could estimate the probability of a lane departure crash upstream of the curve so the warning system could be activated. As vehicle's automation improves, the vehicle could potentially be designed to brake or adjust lane position to reduce their risk of a lane departure before entering the curve.

5.5 References

Kircher, A., M. Uddman, and J. Sandin. *Vehicle Control and Drowsiness*. Swedish National Road and Transport Research Institute. 2002.

Liu, C.C., S.G. Hosking, and M.G. Lenne. "Predicting Driver Drowsiness Using Vehicle Measures: Recent Insights and Future Challenges." Journal of Safety Research. Vol. 40. 2009. pp. 239-245.

Muñoz, M., J. Lee, B. Reimer, B. Mehler, and T. Victor. Analysis of Drivers Head and Eye Movement Correspondence: Predicting Drivers' Glance Location Using Head Rotation Data. Proceedings of the *8th International Driving Symposium on Human Factors in Driver Assessment Training and Vehicle Design.* Snowbird, UT, 2015.

APPENDIX A: DATA EXTRACTION METHODOLOGY

A.1 Roadway Data

The methodology used to reduce various roadway data features is described in the sections below.

Data element: vehicle position within its lane

Need: Lane position may be the best indicator of when a lane departure has occurred. Lane position can also be used to determine the magnitude of the lane departure in terms of departure angle from the roadway and amount that the vehicle encroaches onto the shoulder. Both can be used to set thresholds between different levels of crash surrogates.
Potential source for data element: Data can only be obtained from lane position tracking algorithms and associated data streams such as forward video.
Accuracy: Not yet available from VTTI
Resolution: 10 Hz
Comments: The NDS DAS reports information that can be used to establish lane position. Lane tracking units were reported as centimeters in the data dictionary but a review of the first data set indicated this was erroneous. In a follow-up conversation with VTTI, it was determined that the units initially reported are millimeters. The following variables are used to calculate lane position:

- Lane Position Offset (vtti.lane_distance_off_center): Distance to the left or right of the center of the lane based on machine vision.
- Lane Width (vtti.lane_width): Distance between the inside edge of the innermost lane marking to the left and right of the vehicle. Note that lane width is calculated for each 0.1 second interval and varies somewhat.
- Lane Marking, Distance, Left (vtti.left_line_right_distance): Distance from vehicle centerline to inside of left side lane marker based on vehicle based machine vision.
- Distance from vehicle centerline to inside of left side lane marker based on vehicle based machine vision.
- Lane Marking, Distance, Right (vtti.right_line_left_distance): Distance from vehicle centerline to inside of right side lane marker based on vehicle based machine vision.
- Lane Marking, Probability, Right (vtti.right_marker_probability): Probability that vehicle based machine vision lane marking evaluation is providing correct data for the right side lane markings. Higher values indicate greater probability.
- Lane Markings, Probability, Left (vtti.left_marker_probability): Probability that vehicle based machine vision lane marking evaluation is providing correct data for the left side lane markings.

Offset from lane center and distance from the right (R_D) or left lane (L_D) line are the metrics currently being used as crash surrogates. R_D and L_D are calculated as shown below in meters.

$$L_D = -(L_{CL}) - (T_w/2) \qquad \text{(Eq. A-1)}$$

$R_D = R_{CL} - (T_w/2)$ (Eq. A-2)

Where:

L_D = distance from left edge of vehicle to left edge of lane line, if negative means left edge of car is to the left of the left edge line

R_D = distance from right edge of vehicle to right edge of lane line, if negative, means right edge of car is to the right of the right edge line

T_w = vehicle track width

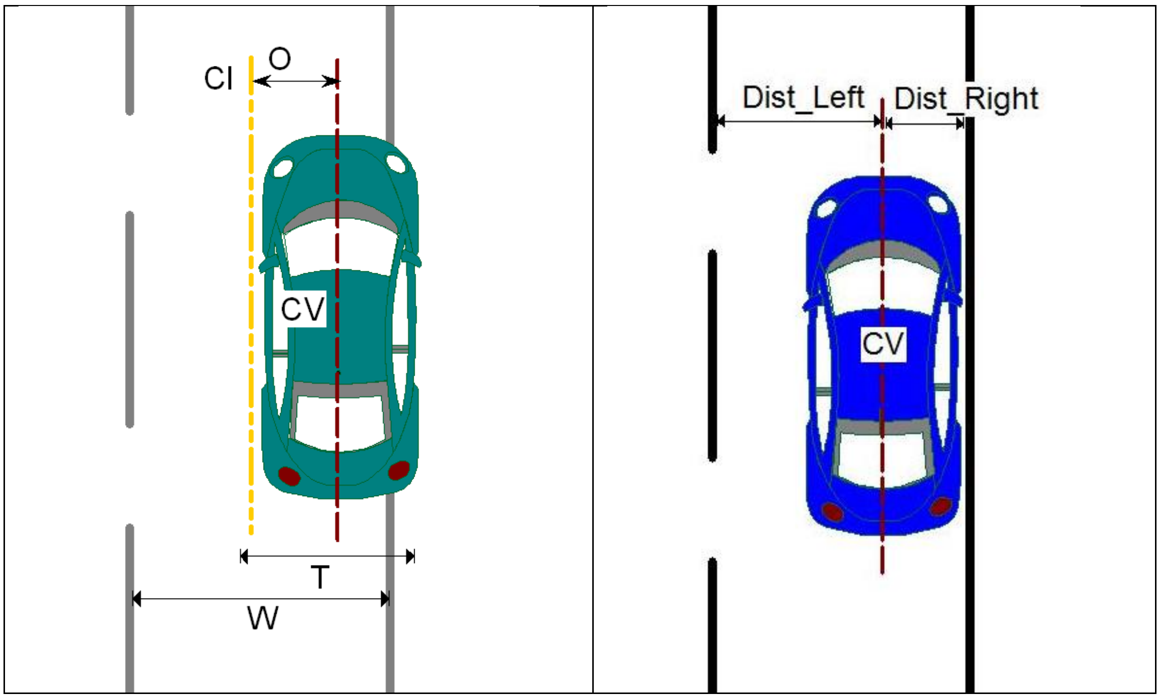


Figure A.1 Description of Variables to Calculate Lane Position

Data element: presence and distance between subject vehicle and other vehicles

Need: establish outcome from lane departure, used as a measure of level of service. Presence of other vehicles (opposing, vehicles passed) can be used to determine roadway density as an exposure method.

Source: forward video

Accuracy: ± 3 ft (0.914 m)

Resolution: collected as vehicle was approaching the curve

Comments: A subjective measure of distance will be obtained from the forward video, as shown in Figure A.1, but distance cannot be determined.

When a conflict occurs, distance to a forward or side vehicle will be determined from the forward or side radar. However, only vehicles within the radar range can be detected.

Coding

Following

0: no forward vehicle present

1: forward vehicle present but not following
2: following closely (less than 3 seconds apart)

Subject vehicle is following closely forward vehicle

Subject vehicle not considered to be following forward vehicle (Image source: UMTRI RDCW dataset)

Figure A.2 Subjective measurement of vehicle following.

Data element: lane width

Need: independent variable in the statistical analysis, also needed to establish vehicle position within its lane
Source: Mobile mapping when available; lane tracking system (varies significantly over 0.1 second intervals – could use average);
Accuracy: need to determine from mobile mapping and lane tracking.
Resolution: at curve approach, PC, apex, PT
Comments: Lane width is measured by the DAS lane tracking system and will be used when position within the lane is needed.
Coding: LaneWidth: reported in meters

Data element: shoulder width

Need: independent variable in statistical analyses. Shoulder and median width also affect potential outcomes for lane departures.
Source: mobile mapping data; may be available from roadway databases;
Accuracy: ± 0.5 ft (0.152 m)
Resolution: at curve approach, PC, apex, PT (should be checked at several points but can be reported once)
Comments: Could not be accurately measured from aerial images and is therefore not included in initial analysis as mobile mapping data not available.
Coding

Paved shoulder width
1: less than 1'
2: 1' to less than 2'
3: 2' to less than 4'
4: greater than or equal to 4'

Data element: curve length and radius

Need: independent variable in statistical analyses, may also be used to assess roll hazard
Source:
Mobile mapping
Aerial imagery
Accuracy: ± 25 ft (7.62 m) for curve length and± 10% for radius
Resolution: once per curve
Comments: Extracted for each direction and then averaged to find one value for each curve.
Coding:

Length of curve from PC to PT reported in meters (Length)
Radius of curve in meters (Radius)

Data element: curve super elevation

Need: independent variable in statistical analyses, may also be used to assess roll hazard
Potential source for data element:
Mobile mapping is likely the only feasible source.
Accuracy: Maximum super elevation for areas with no ice and snow is 12 percent; for areas with snow and ice the maximum is 8 percent. Given these ranges, ideal accuracy is 0.5 percent, but it is unknown if this accuracy can be practically measured in the field. Under normal circumstances cross slope is 1.5 percent to 2 percent. Ideally, it would be necessary to measure this variable at 0.1 percent accuracy to determine differences, but this may not be practical.
Resolution: Once per curve as reported by the mobile mapping
Comments: S04 data had both negative and positive values
Coding: Extracted once per curve for each lane.

Super-elevation in percent (Super)

Data element: driving direction

Need: independent variable in statistical analyses, also important for determining the potential outcome of a non-crash lane departure
Source aerial imagery and forward view
Accuracy: N/A
Resolution: should be indicated once per curve
Comments: none
Coding

Direction of travel (Cardinal)
0: N/S
1: E/W
2: NE/SW
3: NW/SE

Direction of curve from perspective of driver (Direction)
0: outside/left-hand
1: inside/right-hand

Data element: distance to upstream curve, distance to downstream curve from perspective of driver (meters)

Need: Drivers may negotiate curves differently if they have traveled for some distance between curves rather than having negotiated a series of curves. Also used as an independent variable in statistical analyses.
Source: aerial imagery
Accuracy: ± 25 ft (7.62 m)
Resolution: upstream and downstream per curve
Comments:
Coding:

Distance to upstream curve from perspective of driver in meters (DistUP)
Distance to downstream curve from perspective of driver in meters (DistDown)

Curve type:
0- individual curve
1- S-Curve (less than 600 feet between subsequent curves)
2- Compound curve (0' between 2 the PT and PC of subsequent curves in the same direction)

Data element: Speed limit, Curve Advisory, Chevrons and W1-6 signs

Need: independent variable in statistical analyses
Source:

- Speed limit and curve advisory speed limit from mobile mapping
- forward video/Google/forward view mobile mapping for remaining

Accuracy: The general location of the sign or an indication that the sign is present is adequate. For instance, it would be important to know the number and type of chevrons that were present on a curve, but it is not be necessary to know exactly where each sign is located. It is also assumed that all signs are compliant with National Cooperative Highway Research Program (NCHRP) 350 so that they would not need to be considered as strike able fixed objects when determining the outcome of a lane departure event. A sign located using a standard GPS with accuracy of ± 6.6 ft (2 m) would be adequate.

Resolution: as they occur

Coding:

Tangent speed limit (SpdLimit) in mph

Advisory Speed (Advisory) in mph or 999 if no advisory speed limit exists

Presence of chevrons (Chevrons)
0: not present
1: present

Presence of Curve Advisory Sign
0: not present
1: present

Presence of W1-6 Sign
0: not present
1: present

Data element: number of driveway or other access points

Need: Traffic entering and exiting the traffic stream can impact vehicle operation. This traffic would be included as an independent variable in statistical analyses.

Source: aerial imagery and forward imagery

Accuracy: N/A

Resolution: number in the upstream, curve and downstream,

Comments: 4 way intersections counted as 1 cross street

Coding: number of driveways at approach, within curve, at exit

- Cross Streets (CrossStreets) in points per section through length of curve and tangents
- Driveways (Dwys) in driveways per section through length of curve and tangents

Data element: presence of edge or centerline rumble strips

Need: independent variable in statistical analyses, also needed to establish outcome of lane departure

Source: forward video and Google Street View

Accuracy: N/A

Resolution: curve approach and in curve

Comments on extracting data from existing datasets: Only presence of RS could be extracted, not distance from road.

Coding:

Type of rumble strip (RS)
0: no rumble strip present
1: edge line rumble strips only
2: centerline rumble strips only
3: centerline and edge line rumble strips

Figure A.3 Presence of edge line only rumble strips (image source: DAS forward imagery)

Data element: roadway delineation (presence of lane lines or other on-roadway markings)

Need: critical for lane position tracking software, would be included as an independent variable in statistical analyses.
Source: Forward view
Desired accuracy: Data is a quantitative estimate of visibility of markings.
Resolution: once per mile or as situation changes
Comments: This element needs to be current to driving situation and can only be extracted from forward imagery. This information could be obtained from the UMTRI dataset but was more difficult with the VTTI dataset due to image resolution.
Coding:

Presence of Raised Pavement Markings (RPMs)
0: not present
1: present

Roadway Delineation (Delineation)
0: highly visible
1: visible
2: obscured
3: not present

Figure A.4 shows an example of a subjective measure.

Pavement markings indicated as "highly visible"

Pavement markings indicated as "visible"

Right pavement markings indicated as "obscured"

Figure A.4 Subjective measure of lane marking condition using forward imagery (Source: forward video and UMTRI RDCW dataset).

Data element: roadway furniture

<u>Need:</u> necessary to determine how roadside make up affects driving. Also how roadway furniture may be impact the severity of a lane departure crash.
<u>Source:</u> Forward view
<u>Accuracy:</u> n/a
<u>Resolution:</u> Once per curve just upstream of PC looking at curve ahead for roadway furniture rating. Once per curve at any location for presence of guardrail.
<u>Coding:</u>

Presence of Guardrail:
0: not present
1: present

Roadway furniture:
1: little to no roadway furniture
2: moderate roadway furniture
3: large amount of roadway furniture

Little to no roadway furniture

Moderate roadway furniture

Large amount of roadway furniture

Figure A.5 Subjective measurement of vehicle following (image source: DAS forward imagery)

Data element: Sight Distance

Need: the distance at which the curve is first visible will have an effect on where driver reacts to the curve as well as could play a role in lane departures
Source: Forward view and time series data
Accuracy: n/a
Resolution: Once per direction per curve
Comments: This was calculated once per curve using the best forward video available. At times night was the only condition to assess sight distance of the curve. Timestamp at which curve could first be seen was recorded and then used to find corresponding distance upstream in time series data
Coding: distance in meters to PC

A.2 Environmental factors

The following section summarizes environmental factors necessary to address lane departure research questions, indicates potential sources in the existing datasets, suggests accuracy and frequency needs, and includes comments about the accuracy and availability in the existing datasets.

Data element: roadway surface condition (presence of roadway irregularities such as pot holes)

Need: independent variable in statistical analyses, may also impact potential outcome of lane departure

Source: forward or other outward facing video, status and frequency of wiper blades, outside temperature if available, roadway weather information system (RWIS) data if archived

Accuracy: measure is subjective and therefore inapplicable

Resolution: at curve approach, in curve

Comments:

Coding:

Roadway surface condition (PaveCnd)
0: normal surface condition, no obvious damage present
1: moderate damage
2: severe damage, presence of potholes

Pavement condition indicated as "normal"

Pavement condition indicated as "moderate"

Figure A.6 Subjective measure of roadway pavement surface condition using forward imagery (image source: DAS forward imagery)

Data element: environmental conditions such as raining, snowing, cloudy, clear, etc. (may not correspond to roadway surface condition)

Need: independent variable in statistical analyses, may affect sight distance and is related to visibility

Source: forward imagery or archived weather information, ambient temperature probe

Accuracy: subjective measure

Resolution: once per vehicle trace

Comments: A general assessment of environmental conditions can be obtained from the forward video. Even with wiper position, it is difficult to tell how heavy rainfall is. Archived weather information could provide general information for an area but cannot tell the exact environmental conditions for the location where the subject vehicle is located.

Coding:

Roadway surface condition (Surface)
0: dry pavement surface
1: pavement wet but not currently raining
2: wet and light rain
3: wet and heavy rain
4: snow present but road is bare
5: snow along road edge and/or centerline
6: light snow on roadway surface
7: roadway surface covered

Pavement surface condition (snow present but roadway bare)

Pavement surface condition (wet but amount of water cannot be determined)

Surface irregularities

Figure A.7 Pavement surface condition from forward imagery. (Source: UMTRI RDCW dataset)

Data element: ambient lighting

Need: independent variable in statistical analyses
Source: derived from sun angle, twilight, and forward view
Accuracy: subjective measures
Resolution: once per trace or as conditions change
Comments: A relative estimate of ambient lighting can be obtained in most cases from the forward imagery. The limitations are that it was difficult during high cloud cover or low visibility to subjectively estimate ambient lighting.
Coding

Ambient lighting (Lighting) time of day and lighting
0: daytime
1: dawn/dusk
2: nighttime, no lighting
3: nighttime, lighting present

Data element: visibility

Need: independent variable in statistical analyses, serves as a measure of sight distance and can also indicate surface conditions
Source: Forward view is the only reasonable data source
Accuracy: subjective variable
Resolution: once per trace
Comments: This element is available from forward imagery. In some cases it may be difficult to tell whether visibility or image resolution causes securement as shown in Figure A.8. The source of decreased visibility could not be determined. Low visibility is shown in Figure A.9, but it is unknown if the source is fog, smoke, or dust.
Coding:

Visibility
0: clear
1: reduced visibility
2: low visibility

Figure A.8 Image shows some reduced visibility but may be due to sun angle or image resolution. (image source: DAS forward imagery)

Figure A.9 Low visibility appears due to fog. (image source: DAS forward imagery)

A.3 Exposure factors

The following section summarizes exposure factors necessary to address lane departure research questions, indicates potential sources in the existing datasets, suggests accuracy and frequency needs, and includes comments about the accuracy and availability in the existing datasets.

Data element: density

Need: exposure measure
Source: forward video
Accuracy: N/A
Resolution: Number of vehicles on approach, within curve, at exit
Comments: The number of oncoming vehicles, vehicles passed by the subject vehicle, or vehicles that the subject vehicle passes can be counted using the forward and side imagery. Density can be calculated knowing the number of vehicles encountered over a specific distance. Density is a good measure of roadway level of service. However, counting vehicles in the forward or side imagery is time-consuming.
Coding:
Number of vehicles passing subject vehicle during period (Density) in vehicles per meter, calculated through curve

A.4 Driver Video Reduction

Table A.1 Eye Glance Coding

LOCATION OF EYE GLANCE	CODING RULE
Forward	Gazes to the center, left or right that involve little or no head movement and appear to be mostly directed to the left or right portions of the windshield should be coded as 'Forward'.
Center Console	Eyes move slightly down and to the right. There is little or no head movement (e.g., HVAC, radio).
Steering Wheel	Eyes move down slightly. There is little or no head movement (e.g., speedometer, fuel gauge, cruise control).
Down	Draw an imaginary horizontal line in the middle of the steering wheel. If a gaze is directed above the line it should be coded as 'Steering wheel' or 'Center console'. If it is below that line, it should be coded 'Down'. There is some head movement associated with a 'Down' glance (e.g., looking at something in lap or floor)
Up	Eye movement to the upper-left or upper- central portion of the windshield it should be coded as 'Up'. This glance is rare and is usually associated with the visor or sun-roof, if present.
Left	Any gazes to the left of the A-pillar should be coded as 'Left' whether the driver is looking at the left mirror or out the driver's side window.
Right	Any gazes that involve both eye and head move to the right should be coded as 'Right' whether the driver is looking at the right mirror, glove box, front-seated passenger, or out the passenger's side window.
Rear-view mirror	Eye movements up and to the right with a slight head movement should be coded as 'Rear-view mirror'. These include scanning the roadway behind the vehicle as well as glances to the rear-seated passengers.
Over the shoulder	Any glance over the left or right shoulder of the driver. This will require the driver's eyes to pass the B-pillar.
Other	Blinks, squints, or closed eyes that last more than 10 frames. Any blinks, squints or closed eyes less than that should be disregarded.
Missing	Code as 'Missing' if: • the eyes are obscured or obstructed for more than 10 frames • the video freezes or video signal is dropped, or • the locus of gaze cannot be inferred due to glare, excessive head movement or camera location.

Table A.2 Potential Distractions associated with eye glances

Distraction	**Probable Glance Locations**	**Situation**
Passenger	Right (front-seated passenger), Rear-view mirror or Over the shoulder (rear-seated passenger)	A glance associated with a front or rear-seated passenger with indication of a conversation or other distracting activity. The glance location depends on the seating position of the passenger.
Route planning (locating, viewing, or operating)	Steering wheel, Down, Center console	A glance associated with the actions performed during the use of a paper map or in-vehicle navigation system. The glance location depends on where the driver holds the instrument while looking at it.
Moving or dropped object in vehicle	Down	A glance associated with the driver reaching for something in the vehicle. The glance location depends on the location of the object.
Animal/insect in vehicle	All locations are possible	A glance associated with the driver being preoccupied by the presence of an animal/insect and taking action to remedy the distraction. The mere presence is not to be coded as a distraction. The glance location depends on where the animal/insect is located in the vehicle.
Cell phone (locating, viewing, operating)	Steering wheel, Down, Center console	A glance associated with the actions performed during cell phone use. The glance location depends on where the driver holds the phone while looking at it.
IPod/MP3 (locating, viewing, operating)	Steering wheel, Down, Center console	A glance associated with the actions performed during the use of an in-vehicle entertainment system. The glance location depends on the location of the device.
In-vehicle controls	Center console, Steering wheel, Down	A glance associated with the actions performed using the in-vehicle controls (e.g., HVAC, radio, cd player, wipers, windows, door locks). The glance location depends on the control being activated.
Drinking/Eating	Steering wheel, Down	A glance associated with locating/adjusting food item or drink container. The glance location depends on where the driver is holding the food/drink.
Smoking	Steering wheel, Down, Center Console, Left	A glance associated with locating, lighting, smoking or disposing of ashes. The glance location depends on where the driver holds the cigarette and where they discard the ashes.
Personal Hygiene	Up, Rear-view mirror, Steering wheel, Down	A glance associated with the driver performing an action related to personal hygiene (e.g., fixing hair, applying makeup, blowing nose etc.). The glance location depends on the activity the driver is performing.
Other task	Any are possible	A glance not fitting another category (make a note if used)