# Evaluation of Data Needs, Crash Surrogates, and Analysis Methods to Address Lane Departure Research Questions Using Naturalistic Driving Study Data

**DETAILS**

0 pages | | PAPERBACK
ISBN 978-0-309-43523-9 | DOI 10.17226/22848

**AUTHORS**

The Second

STRATEGIC HIGHWAY RESEARCH PROGRAM

SHRP 2 REPORT S2-S01E-RW-1

# Evaluation of Data Needs, Crash Surrogates, and Analysis Methods to Address Lane Departure Research Questions Using Naturalistic Driving Study Data

Shauna Hallmark
Institute for Transportation, Iowa State University

Yu-Yi Hsu
Department of Statistics, Iowa State University

Linda Boyle
Seamans Center for the Engineering Arts and Sciences, University of Iowa

Alicia Carriquiry
Department of Statistics, Iowa State University

Ye Tian
Department of Statistics, Iowa State University

Abhisek Mudgal
Institute for Transportation, Iowa State University

TRANSPORTATION RESEARCH BOARD

WASHINGTON, D.C.
2011
www.TRB.org

## Subscriber Categories

Highways
Safety and Human Factors

## The Second Strategic Highway Research Program

America's highway system is critical to meeting the mobility and economic needs of local communities, regions, and the nation. Developments in research and technology—such as advanced materials, communications technology, new data collection technologies, and human factors science—offer a new opportunity to improve the safety and reliability of this important national resource. Breakthrough resolution of significant transportation problems, however, requires concentrated resources over a short time frame. Reflecting this need, the second Strategic Highway Research Program (SHRP 2) has an intense, large-scale focus, integrates multiple fields of research and technology, and is fundamentally different from the broad, mission-oriented, discipline-based research programs that have been the mainstay of the highway research industry for half a century.

The need for SHRP 2 was identified in *TRB Special Report 260: Strategic Highway Research: Saving Lives, Reducing Congestion, Improving Quality of Life,* published in 2001 and based on a study sponsored by Congress through the Transportation Equity Act for the 21st Century (TEA-21). SHRP 2, modeled after the first Strategic Highway Research Program, is a focused, time-constrained, management-driven program designed to complement existing highway research programs. SHRP 2 focuses on applied research in four areas: Safety, to prevent or reduce the severity of highway crashes by understanding driver behavior; Renewal, to address the aging infrastructure through rapid design and construction methods that cause minimal disruptions and produce lasting facilities; Reliability, to reduce congestion through incident reduction, management, response, and mitigation; and Capacity, to integrate mobility, economic, environmental, and community needs in the planning and designing of new transportation capacity.

SHRP 2 was authorized in August 2005 as part of the Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users (SAFETEA-LU). The program is managed by the Transportation Research Board (TRB) on behalf of the National Research Council (NRC). SHRP 2 is conducted under a memorandum of understanding among the American Association of State Highway and Transportation Officials (AASHTO), the Federal Highway Administration (FHWA), and the National Academy of Sciences, parent organization of TRB and NRC. The program provides for competitive, merit-based selection of research contractors; independent research project oversight; and dissemination of research results.

## SHRP 2 Report S2-S01E-RW-1

ISBN: 978-0-309-12901-5



*Note:* SHRP 2 report numbers convey the program, focus area, project number, and publication format. Report numbers ending in "w" are published as web documents only.

### Notice

The project that is the subject of this report was a part of the second Strategic Highway Research Program, conducted by the Transportation Research Board with the approval of the Governing Board of the National Research Council.

The members of the technical committee selected to monitor this project and to review this report were chosen for their special competencies and with regard for appropriate balance. The report was reviewed by the technical committee and accepted for publication according to procedures established and overseen by the Transportation Research Board and approved by the Governing Board of the National Research Council.

The opinions and conclusions expressed or implied in this report are those of the researchers who performed the research and are not necessarily those of the Transportation Research Board, the National Research Council, or the program sponsors.

The Transportation Research Board of the National Academies, the National Research Council, and the sponsors of the second Strategic Highway Research Program do not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to the object of the report.

**SHRP 2 Reports**

Available by subscription and through the TRB online bookstore:

www.TRB.org/bookstore

Contact the TRB Business Office:
202-334-3213

More information about SHRP 2:

www.TRB.org/SHRP 2

# THE NATIONAL ACADEMIES

*Advisers to the Nation on Science, Engineering, and Medicine*

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. On the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Charles M. Vest is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, on its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. Charles M. Vest are chair and vice chair, respectively, of the National Research Council.

The **Transportation Research Board** is one of six major divisions of the National Research Council. The mission of the Transportation Research Board is to provide leadership in transportation innovation and progress through research and information exchange, conducted within a setting that is objective, interdisciplinary, and multimodal. The Board's varied activities annually engage about 7,000 engineers, scientists, and other transportation researchers and practitioners from the public and private sectors and academia, all of whom contribute their expertise in the public interest. The program is supported by state transportation departments, federal agencies including the component administrations of the U.S. Department of Transportation, and other organizations and individuals interested in the development of transportation. **www.TRB.org**

**www.national-academies.org**

## SHRP 2 STAFF



## ACKNOWLEDGMENTS

This work was sponsored by Federal Highway Administration in cooperation with the American Association of State Highway and Transportation Officials. It was conducted in the second Strategic Highway Research Program (SHRP 2), which is administered by the Transportation Research Board of the National Academies. The project was managed by Charles Fay, Senior Program Officer for SHRP 2 Safety.

The research team would like to thank the Midwest Transportation Consortium (MTC), a Tier 1 university transportation center, for partial funding of three students to participate in this project. We would also like to thank the University of Michigan Transportation Research Institute (UMTRI) and Virginia Tech Transportation Institute (VTTI) for providing naturalistic driving study data and for their assistance in working with the data. In addition, the team thanks the Michigan Department of Transportation (MDOT) for providing roadway and crash data. The results and conclusions deriving from the use of this data represent the work and opinions of the Iowa State University research team and do not reflect the opinions or endorsement of UMTRI, VTTI, or MDOT.

# FOREWORD

Charles Fay, *SHRP 2 Senior Program Officer, Safety*

A large component of the safety research undertaken in the second Strategic Highway Research Program (SHRP 2) is aimed at reducing injuries and fatalities that result from highway crashes. Through a naturalistic driving study (NDS) involving more than 3,000 volunteer drivers, SHRP 2 expects to learn more about how individual driver behavior interacts with vehicle and roadway characteristics. In anticipation of the large volume of data to be collected during the SHRP 2 NDS, several projects were conducted to demonstrate that it is possible to use existing NDS data and data from other sources to further the understanding of the risk factors associated with road crashes. More specifically, the four projects conducted under the title Development of Analysis Methods Using Recent Data examined the statistical relationship between surrogate measures of collisions (conflicts, critical incidents, near collisions, or roadside encroachment) and actual collisions. This report presents the results of one of these projects, undertaken by the Institute for Transportation, Iowa State University. It documents the second phase of a two-phase project under SHRP 2 Safety Project S01E.

---

The primary objective of this work was to investigate the feasibility of using NDS data to increase our understanding of lane departure crashes. Research questions specific to lane departure were identified, and data requirements to answer these research questions were determined. Methodologies for selecting and applying crash surrogates specific to lane departure were evaluated. Finally, four analytical approaches were investigated: data mining using classification and regression tree analysis; simple odds ratio and logistic regression; logistic regression for correlated data that accounts for repeated sampling among observations; and time series analysis. The report discusses the advantages and limitations for each of these approaches. It will provide useful information for analysts of the SHRP 2 NDS data.

# CONTENTS

# Executive Summary

Lane departures are involved in a substantial number of motor vehicle crashes and account for a considerable number of fatalities. Single-vehicle, run-off-road crashes account for almost 39% of traffic fatalities. Two-vehicle head-on crashes account for 18% of noninterchange, nonintersection fatal crashes, with 75% occurring on undivided two-lane roadways. Addressing lane departure crashes is thus a major safety goal in the United States.

Lane departures represent a serious safety concern, but the relationships between factors that influence whether a vehicle departs its lane and the actions and events that determine the outcome are complex and not well understood. The focus of the second Strategic Highway Research Program (SHRP 2) safety research plan is a large field study of naturalistic driving behavior and performance using a comprehensive, state-of-the-art instrumentation package installed in the vehicles of volunteer participants. The SHRP 2 naturalistic driving study (NDS) is intended to support a comprehensive safety assessment of how driver behavior and performance interact with roadway, environmental, and vehicular factors and the influence of these factors and their interactions on collision risk, especially the risk of lane departure and intersection collisions.

SHRP 2's safety research plan will produce a database of naturalistic driving behavior data under Safety Project S06, Technical Coordination and Quality Control, and Safety Project S07, In-Vehicle Driving Behavior Field Study. Vehicle data will be available from the naturalistic driving data. Environmental data may also be extracted from the outside video views from the vehicle instrumentation system. Safety Projects S04A, Roadway Information Database Developer, Technical Coordination, and Quality Assurance for Mobile Data Collection, and S04B, Mobile Data Collection, will produce a database of roadway characteristics that can be linked to the naturalistic driving database to support safety analysis. Some roadway data will be obtained from existing data sets belonging to state and local roadway agencies in the areas where naturalistic driving study data are collected. Select roadway data elements will also be collected using mobile data units. Some roadway information may also be obtained from the vehicle instrumentation's outside video views. The resulting databases will help researchers better understand how combinations of driver behavior and roadway, environmental, and vehicle factors lead to different outcomes. Specifically, some of the data will be used in a full-scale evaluation of lane departures in SHRP 2 Safety Project S08, Analysis of the SHRP 2 Naturalistic Driving Study Data.

In preparation for the SHRP 2 NDS, the primary goals of the research discussed in this report are to identify lane departure research questions that can be answered using data collected in the field study, identify data needs to address the questions, and to demonstrate analytical methods that can be used to answer those research questions.

To accomplish these goals, the following tasks were performed:

1. Identify important research questions related to lane departures.
2. Review and extract data from existing naturalistic driving studies to assess the types of data that may be available in these types of studies.
3. Review information about the types of data that will be available from the SHRP 2 naturalistic driving study.
4. Identify which of the identified lane departure research questions (Task 1) are likely to be feasible for the full-scale study.
5. Identify types of crash surrogates that may be appropriate for use in answering lane departure research questions.
6. Use existing naturalistic driving data sets to explore methods for analyzing SHRP 2 field study data so as to answer the identified questions (Task 4).

Although lane departures can occur on any roadway type, this report addresses rural lane departures, with a focus on rural, two-lane, paved roadways.

## Identification of Lane Departure Research Questions and Necessary Factors

One of the main goals of the research was to identify a set of research questions that could be answered using the SHRP 2 NDS and the roadway characteristics databases and that would be useful in determining why drivers leave the roadway and which factors result in different outcomes. The identification of feasible research questions is presented in Chapter 2. The team also identified questions that are not likely to be feasible because of data limitations.

Research questions identified as being feasible for the SHRP 2 full-scale study of lane departures include the following:

- What environmental, roadway, driver, or vehicle factors influence whether a vehicle departs its lane?
- What environmental, roadway, driver, or vehicle factors influence lane departure outcome?
- What is the impact of lane departure countermeasures on lane departure frequency and outcome?
- What is the relationship between lane departure crash surrogates and crashes?

Research questions involving the following factors are not likely to be feasible with the data resulting from the SHRP 2 field study:

- Driver's alcohol or drug use;
- Pavement surface friction;
- Pavement edge drop-off; or
- Quantitative measures of rain, snow, or ice on the road.

## Identification of Data Necessary to Answer Research Questions

Another goal of the research was to determine what data would be necessary to answer the identified lane departure research questions. A comprehensive literature review was conducted to identify driver, roadway, environmental, and vehicle factors that have been shown to have some correlation to lane departure crashes. Factors identified include horizontal and vertical curvature, roadway cross section, driveway density, illumination, weather, presence of rumble strips, roadway delineation and signing, pavement edge drop-off, vehicle type, speeding, influence of alcohol or drugs, driver age, and distraction. The necessary accuracy and resolution for the identified factors were also determined. A summary of the data elements is provided in Chapter 4.

Once relevant factors were identified, the team reviewed existing naturalistic driving study and roadway data to determine whether it was feasible to obtain each data element identified. This exercise provided insight as to whether data elements were likely to be available in the SHRP 2 field study and how feasible it would be to extract elements that were not readily available. The team obtained a number of events from field operational tests conducted by the University of Michigan Transportation Research Institute (UMTRI) for a road departure curve warning system. The events contained instances where the drivers left their lane, as well as normal driving data on rural roadways. Raw data from their instrumentation system included vehicle variables (e.g., vehicle location, forward speed, forward acceleration, yaw, pitch, lateral acceleration) that were provided at 10 Hz and forward images that were provided at 2 Hz. Roadway and crash data were also obtained for the UMTRI study area from the Michigan Department of Transportation (MDOT).

The team also received 33 crash or near-crash lane departure events from the VTTI 100-car naturalistic driving study for rural roadways. A reduced data set rather than raw data was provided for each event for most variables. A video clip showing views outside the vehicle was also provided.

Both the UMTRI and the VTTI data were examined to determine the feasibility of extracting relevant driver, vehicle, environmental, and roadway factors. The availability of the data in the UMTRI and VTTI databases were reviewed and the limitations described.

Chapter 3 summarizes the various data sets used in the research. A description of common data terms is also provided. Appendices A and B describe the protocols, methods, and variable descriptions used to extract data from the UMTRI and VTTI naturalistic driving study data sets. Data were extracted manually, which consumed a large amount of resources. The method used to extract the data provides a framework that can be used by other researchers in working with the SHRP 2 naturalistic driving study data.

The accuracy, frequency, and resolution of data elements that were likely to be available in the SHRP 2 naturalistic driving study were also identified through a review of available documentation about the instrumented vehicle data acquisition system (Safety Project S05, Design of the In-Vehicle Driving Behavior and Crash Risk Study) and a review of preliminary roadway data elements identified in Safety Project S03 (Roadway Measurement System Evaluation), as discussed in Chapter 4. Data elements were also prioritized because resource limitations in the SHRP field study will constrain data collection.

Information about which data elements were necessary to address lane departure research questions and limitations expected in the SHRP 2 field was used to provide input to the proposed instrumented vehicle data acquisition system (Safety Project S05) and to provide feedback to identification of initial data elements in Safety Project S03. The information will also help guide identification and prioritization of roadway data collection in Safety Projects S04A and S04B.

## Methodological Approach to Selecting Lane Departure Crash Surrogates

Lane departure crashes are a key measure of road safety. However, naturalistic driving studies, even the fully deployed SHRP 2 field study, will have limited cases of lane departure crashes. The naturalistic driving study will capture crashes, near crashes, and incidents, as well as normal driving. The frequency of incidents and near-crash events is typically greater than the frequency of crashes; incidents and near-crash events may be used as crash surrogates. Using surrogates will also provide an opportunity to study what happens preceding and following a lane departure event. The most significant advantage of naturalistic driving studies is that they provide a firsthand record of the events that precede crashes and incidents.

Chapter 5 discusses potential lane departure crash surrogates that can be obtained from naturalistic driving study data. Literature on crash surrogates with an emphasis on lane departure crash surrogates is summarized, and a methodological approach for selecting and applying lane departure crash surrogates is outlined.

Existing naturalistic driving study data were also evaluated to determine starting points for setting triggers to identify lane departure events. The team reduced lane departure incidents in the UMTRI data set for data on rural, two-lane roads resulting in 22 right-side and 51 left-side lane departures. Data for which no incident had occurred, which represented normal driving data, were also extracted. The reduced data was used to assess which variables and thresholds may be the most useful in setting triggers to identify lane departure events in the full-scale field study. Data for several kinematic vehicle variables that may signal a lane departure (lateral speed, yaw rate, side acceleration, forward acceleration, roll rate, and pitch rate) were identified. Values for the kinematic variables were compared for normal driving against left- and right-side lane departures using a Wilcoxon Rank Sum Test to determine whether the normal driving data were statistically different from lane departure events.

Although the distributions for most variables were determined to be different at the 95% level of significance, a significant amount of overlap exists. This indicates the difficulty in setting thresholds low enough to include all incidents but still high enough so that a large amount of nonincident data does not have to be evaluated. Although there were not enough data to determine what threshold values should be set, initial results suggest that for left-side lane departures, roll rate, yaw rate, side acceleration, and side speed are likely to be good candidates to identify events. Results suggest that for right-side lane departures, yaw rate, side acceleration, and lateral speed are good candidates to identify events.

The VTTI and UMTRI data were evaluated separately because different data were available for each. The only kinematic variables available for the VTTI data set were forward and side acceleration. In addition, no normal driving data were available. As a result, it was difficult to assess which variables could be used to determine when lane departures had occurred using that data set.

Ways to partition normal driving data were also evaluated using the UMTRI data. Lateral offset was compared for several driving situations. Differences were noted between driving on a tangent and on left- and right-hand curves, between night and daytime driving, and between individual drivers. Some guidance on stratifying normal driving by relevant variables was also provided.

Chapter 5 only offers an approach for selecting and evaluating lane departure crash surrogates. With the data available, the team was not able to conduct an analysis to evaluate the relationship between lane departure crash surrogates and crashes. The UMTRI data set provided a large number of lane departures and normal driving data but did not include crashes. The VTTI data set contained both crashes and lane departure events; however, once the data were partitioned by roadway type, the data were insufficient to conduct an evaluation. Additionally, no normal driving data were provided for comparison.

## Exploration of Analytical Approaches to Answer Lane Departure Research Questions

Four analytical approaches were identified that can be used to evaluate the data resulting from the SHRP 2 field study and thus answer the lane departure research questions as discussed in Chapter 6:

1. Data mining using classification and regression tree analysis;
2. Simple odds ratio and logistic regression;

3. Logistic regression for correlated data that accounts for repeated sampling among observations (e.g., repeated sampling for the same driver, trip); and
4. Time series analysis.

Initial analysis of the lane departure and normal driving events extracted from the UMTRI data as discussed in Chapter 5 was conducted using the four approaches. A description of each approach is presented in Chapter 6. The data used; a description of the model, results, sample size, and implications for the full-scale study; and data segmentation methods are also presented. The focus is on rural, two-lane roadways. Because data were limited during the research, the analysis was exploratory to determine whether the approach is appropriate for the SHRP 2 field study.

Method 1 (classification and regression tree) and Method 2 (simple odds ratio and logistic regression) evaluated the likelihood of a left- or right-side lane departure. A sample-based data aggregation approach was used in the classification and regression tree analysis, and an event-based data aggregation approach was used for the logistic regression. Although available sample sizes were limited, both methods produced similar results. Both indicated that curve radius, driver age, and type of shoulder were relevant in explaining lane departures. Logistic regression also indicated that both left- and right-side lane departures were more likely to occur at night and were less likely to occur as lane width increased. The model for left-side lane departures indicated that male drivers were more likely than female drivers to be involved in a lane departure, and the model for right-side lane departures indicated that lane departures are more likely on roadway sections with a higher density of lane departure crashes and for drivers who spend more time traveling 10 mph or more over the posted speed limit.

The third method expanded on a varied logistic regression approach based on the logistic regression model just described, which may be better suited to the data from the full-scale study.

The fourth method, time series analysis, used continuous data to develop a model to predict offset as a function of several vehicle kinematic variables. The method was developed and explained in such a way that it could be adapted to the SHRP 2 field study to include various explanatory variables, including driver behavior. This approach allows information, such as driver distraction in previous time periods, to be incorporated into the model.

Each analytical approach has advantages and limitations for the full-scale study, and selecting an appropriate method depends on the specific research questions posed and the resources available to reduce the data.

The VTTI data could not be used in the analysis since only a limited number of events were available once data were extracted for only rural, two-lane roadways.

CHAPTER 1

# Background

This chapter introduces the research problem and scope and outlines the organization of the full report.

## Introduction

Lane departure crashes make up a significant number of motor vehicle crashes and account for a disproportionate number of fatalities. LeBlanc et al. (2006) estimated that road departure crashes account for 15,000 fatalities per year in the United States. Neuman et al. (2003), using Fatality Analysis Reporting System (FARS) data, estimated that almost 39% of traffic fatalities were single-vehicle, run-off-road (ROR) crashes. Moreover, Neuman et al. (2003) evaluated FARS data and estimated that 18% of noninterchange, nonintersection fatal crashes were a result of two-vehicle head-on crashes, the majority in nonpassing situations, with 75% occurring on undivided two-lane roadways.

The importance of addressing lane departure crashes has been underscored by the American Association of State Highway and Transportation Officials (AASHTO) in its Strategic Highway Safety Plan. One of the main goals of the plan is to keep vehicles on the roadway; another goal is to minimize the adverse consequences of leaving the road. Furthermore, AASHTO identified mitigating ROR crashes as one of its emphasis areas (Neuman et al., 2003).

Lane departure is a serious safety concern, yet the relationship between the factors that influence whether a vehicle departs its lane in the first place and the series of actions and events that determine the outcome are complex and not understood well. Preventing lane departure in the first place depends on understanding where in the sequence of crash events the lane departure could have been prevented and then applying appropriate countermeasures. For instance, a driver might begin to drift toward the roadway edge because of inattention or because he or she approached a curve at a greater speed than that at which the curve could be negotiated. Both could lead to a road departure, but different countermeasures would be used for each. Edgeline rumble strips may prevent the first, and in-vehicle curve warning systems may prevent the second.

Once a vehicle departs its lane, the ability to ameliorate the outcome depends on understanding the factors that influence each subsequent action or event and then applying the appropriate countermeasures to address these factors when appropriate. For instance, a vehicle may partially leave the roadway because of various factors. The ability to recover and safely return the vehicle to the travel lane depends on the roadway, environment, and in some cases, vehicle factors, along with the driver's response. To illustrate the point, in one situation a driver encounters a fully paved shoulder after leaving the roadway and is able to safely correct and return to his or her travel lane. In an otherwise similar situation, in contrast, the driver encounters gravel shoulders and then overcorrects and loses control, resulting in a rollover. In the first case, the presence of paved shoulders was a key factor in determining subsequent events and the final outcome of the road departure. In the second case, the presence of loose shoulder material and the overresponse by the driver exacerbated the situation and resulted in a more serious outcome. The ability to understand what factors influence whether a crash occurs, as well as what factors result in less serious outcomes, would greatly improve the ability to select and evaluate the effectiveness of appropriate countermeasures.

Currently, understanding why lane departures occur is limited in three ways. First, crash data are limited. Crashes are rare and random, and, as a result, safety analyses must depend on small sample sizes. In addition, crash reporting can be inconsistent, so comparison across sites is difficult. Another problem is the timeliness of crash data. Once a countermeasure is implemented, agencies prefer to evaluate the effectiveness as soon as possible before investing more resources in the treatment. However, before-and-after crash studies often cannot be completed for several years after installation of the treatment, until representative samples can be obtained and regression to the mean avoided. Crash databases also provide limited information. The sequence of events leading to a crash and the surrounding conditions coded into crash databases are provided by an officer either assessing the situation and recording surrounding conditions after the crash has occurred or questioning drivers and/or witnesses. The amount of information about the sequence of events that precedes a crash, as well as the surrounding roadway, environment, vehicle, and driver conditions that end up in the crash record depends on the reporting officer. The accuracy and usefulness of information that is recorded depend on how carefully the officer evaluates the scene, how accurately or truthfully a driver or witness recalls the sequence of events and conditions that led to the crash, and whether drivers or witnesses actually understood what occurred. For instance, a driver who leaves the roadway and encounters a pavement edge drop-off that causes him or her to lose control as he or she attempts to return to the roadway may not even realize that the reason he or she lost control was because his or her rear tire got caught on the pavement edge.

Second, the ability to fully understand lane departures is limited because crash databases only record lane departures that result in a collision. In some cases, collisions, particularly minor ones, are not even reported. Additionally, the ability to fully understand lane departures and how they can be prevented or ameliorated requires information about which conditions lead to more favorable events and outcomes. Drivers may leave the roadway edge at the same rate on paved shoulders as on unpaved shoulders but are more likely to fully recover on paved shoulders. Because no crash occurred, there is no record of positive outcomes and which factors influenced them.

Third, with current data sets, it has only been possible to study why actual lane departure crashes occur and then to attempt to develop countermeasures for these relatively rare events. Drivers are more likely to be involved in a road departure incident in which they leave the roadway and are able to avoid a crash and return to their lane than they are to be involved in an actual crash. For instance, in the Virginia Tech Transportation Institute (VTTI) 100-car study, researchers found 69 crashes, 761 near crashes, and 8,295 incidents (Klauer et al., 2006). In some cases, these incidents are indicators of near misses and provide valuable information about why crashes occur and the crash potential of a given situation. In other cases, positive outcomes indicate that some roadway, environmental, or human factor had a significant positive influence in making a safe return to the roadway possible.

Existing naturalistic driving studies, and the future on-road study scheduled under the second Strategic Highway Research Program (SHRP 2) with approximately 2,000 instrumented vehicles, will provide rich and unique databases that can be utilized to derive relationships among incidents, crashes, and human factors; roadway elements; environmental conditions; and vehicle characteristics and thus address the problems presented in the previous paragraphs (TRB, 2007). Because crashes are rare, crash surrogates have been used to better understand crash risk and overcome many of the problems with crash databases. Naturalistic driving studies obtain information on normal driving but can also capture crashes, near crashes, and incidents that may arise. The frequency of incidents and near-crash incidents is typically greater than actual crashes and can provide greater insights related to the circumstances preceding the incident, including the driver's behavior and any environmental, roadway, or traffic conditions that may have contributed to the incident. Data from incidents and near crashes can therefore be used as crash surrogate measures to further examine crash risk.

Using crash surrogates also provides an opportunity to study what happens before and after an incident. The most significant advantage of the naturalistic driving studies is that they provide a firsthand record of the events that precede crashes and incidents. Roadway, environmental, vehicle, and human factors can be extracted directly, rather than through secondhand information from police records and crash databases, to develop relationships among the factors that influence road departure crash risk. Improved data about actual events leading to both road departure crashes and noncrash incidents will be extremely valuable in developing a better understanding of what negative factors lead to crashes and near misses, as well as what factors result in more positive subsequent events and outcomes. Understanding the reasons why crashes do not occur yields as much useful information as evaluating why they do occur. In both cases, factors that cause a vehicle to initially leave the roadway and the relationship among road, environment, vehicle, and human factors and subsequent events and outcomes can be studied. Dingus et al. (2006) reported that the analysis of near crashes from the VTTI naturalistic driving study (NDS) has been valuable, as it demonstrates drivers successfully performing evasive maneuvers.

## Scope of Research

The main goal of the research discussed in this report was to identify relevant research questions for addressing lane departures and to determine whether these questions can be addressed using data that is expected to result from SHRP 2's full-scale NDS. For the present project, lane departure crash surrogates were also identified. The research addressed rural lane departures, with a focus on rural, two-lane, paved roadways.

To accomplish this agenda, the researchers performed the following tasks:

- Driver, roadway, environmental, and vehicle factors expected to contribute to lane departure crashes (summarized in Chapter 2) were identified through a review of available literature and through the team's expertise.
- Data from the VTTI 100-car study and the UMTRI road departure crash warning (RDCW) field operation test (FOT) NDS were evaluated to determine whether the data elements identified could be extracted. The methodology and protocol for extracting those data elements were outlined. The methodology and protocol were described so that this information could be used to extract data from the full-scale NDS (summarized in Appendices A and B).
- The accuracy, frequency, and resolution of data collection that would be necessary to address lane departure research questions were determined and summarized (described in Chapter 4). Data elements were also prioritized because resource limitations in the full-scale study will constrain data collection.
- A framework for extracting data elements from existing naturalistic studies that can be used for the full-scale study was developed. Appendices A and B describe the protocols, methods, and variable descriptions. Available documentation of the SHRP 2 Safety Projects S03 and S05 work was reviewed to determine what data sensors would be available and what data elements are expected to be available in the full-scale study. The accuracy, frequency, and resolution of data that are expected to be available to answer lane departure questions were evaluated. The team identified limitations and provided feedback to SHRP 2, as described in Chapter 4.
- Because data were limited, crash surrogates could be not be evaluated. Chapter 5 summarizes information about lane departure crash surrogates and develops a hierarchy of lane departure crash surrogates that can be used in the full-scale NDS. Existing NDS data were also evaluated to determine starting points for setting triggers to identify lane departure events.
- Several analytical approaches that can be used to answer lane departure research questions were developed. Lane departure and normal driving data were identified in the UMTRI RDCW FOT NDS database, and four approaches (data mining, calculation of odds ratio, logistic regression, and a time series analysis) were used to conduct an initial analysis of the data. A description of each approach is presented in Chapter 6. The focus was rural, two-lane, paved roadways.
- Input was also provided to researchers for SHRP 2 Safety Project S02, Integration of Analysis Methods and Development of Analysis Plan. The team collaborated regularly with the Safety Project S02 team and provided input to research questions.

The research presented in this report builds on a Phase I report (Hallmark et al., 2008), relevant background information from which is included here. Most of the information in this report, however, does not depend on the reader having reviewed the Phase I report.

## Organization of This Report

The remainder of this report is organized as follows:

- Chapter 2 provides the results of a literature review conducted to identify driver, roadway, environmental, and vehicle factors that have been shown to have some correlation to lane departure crashes. Factors identified include horizontal and vertical curvature, roadway cross section, driveway density, illumination, weather, presence of rumble strips, roadway delineation and signing, pavement edge drop-off, vehicle type, speeding, influence of alcohol or drugs, driver age, and distraction. Additionally, research questions are identified that may likely be answered using data from the full-scale study or that cannot be answered because of data limitations. The research questions addressed in the scope of this research are also identified.
- Chapter 3 summarizes the various data sets used in the research. A description of common data terms is also provided.
- Chapter 4 identifies data elements that are expected to be necessary to answer lane departure research questions based on a survey of available literature and the team's expertise regarding lane departure issues. The accuracy, frequency, and resolution of each data element are determined and described. Additionally, the availability of the data in the UMTRI and VTTI databases is reviewed and the limitations described. The chapter also reviews the available documentation of the SHRP 2 Safety Projects S03 and S05 work. The accuracy, frequency, and resolution of data that are expected to be available to answer lane departure questions in the full-scale study are evaluated. The chapter identifies limitations and provides feedback to SHRP 2, as described in Chapter 5. Data elements are also prioritized because resource limitations in the full-scale study will constrain data collection.
- Chapter 5 describes lane departure crash surrogates. Literature on crash surrogates is summarized, and a methodological approach for selecting and applying crash surrogates is outlined. Existing naturalistic driving study data are also evaluated to determine starting points for setting triggers to identify lane departure events. A discussion of ways to partition normal driving data is also evaluated using exist-

ing data. Lateral offset is compared for several driving situations. Differences are noted between driving on a tangent and on left- and right-hand curves, between night and daytime driving, and between individual drivers. This chapter also provides some guidance on stratifying normal driving by relevant variables.

- Chapter 6 describes four analytical approaches that can be used to evaluate naturalistic driving study data and answer lane departure research questions. Lane departures and normal driving cases have been identified in the UMTRI RDCW FOT NDS data, and four approaches (data mining, calculation of odds ratio, logistic regression, and a time series analysis) have been used to conduct an initial analysis of the data. A description of each approach is presented in Chapter 6. The data used, a description of the model, results, sample size, and implications for the full-scale study are also discussed. The focus is on rural, two-lane roadways. Because data were limited during the research, the analysis is exploratory, to determine whether the approach is appropriate for the full-scale study.
- Chapter 7 provides a summary of the entire project.
- Appendices A and B describe the protocols, methods, and variable descriptions used to extract data from the UMTRI and VTTI naturalistic driving study data sets. The method used to extract the data provides a framework that can be used by other researchers in working with the full-scale study. Data were extracted manually, which consumed a large amount of resources. How lane departures were identified within the UMTRI data set is also discussed in Chapter 5.

CHAPTER 2

# Identifying Final Lane Departure Research Questions and Relevant Factors

One of the goals of this project was to develop a set of research questions that could be explored using existing NDS data. The intent was to then determine which research questions could adequately be addressed given the data and the likely limitations of the SHRP 2 full-scale NDS (hereafter referred to as the full-scale NDS or full-scale study).

In order to identify lane departure research questions, the research team first identified which driver, vehicle, roadway, and environmental factors were likely to contribute to the occurrence and severity of lane departure crashes, based on an in-depth literature review described in the next section and on the team's expertise in lane departure issues. The team then reviewed data from existing NDSs, as well as information available about the full-scale NDS. Research questions that could not feasibly be answered because the necessary data would not be available or could not be extracted were identified. This chapter identifies relevant factors, Appendices A and B address the feasibility of extracting the data elements from the UMTRI and VTTI databases, and Chapter 4 comments on which data factors are expected to be available with the full-scale study. The team also identified which research questions could be addressed in the present research. Although the information to select final research questions is based on the information provided in the following chapters, the information is also provided in this chapter to simplify report organization.

## Relevant Data Elements Identified in Existing Literature

In order to formulate research questions, it was necessary to determine which data elements are the most relevant. This section addresses factors necessary or desirable to evaluate lane departure crashes. The data elements were selected through a review of currently available literature regarding which roadway, environmental, vehicle, and driver variables are correlated to road departure crash occurrence. Roadway data elements and crash data information were also selected based on the team's experience in working with road departure crashes and its understanding of what roadway variables are likely to be available or collected.

### Roadway Factors

#### *Horizontal and Vertical Curves*

Horizontal and vertical curvature, as well as grade, have been correlated with crash occurrence in a number of studies. Torbic et al. (2004) report that the crash rate for horizontal curves is approximately three times that of tangent sections. The authors also indicate that approximately 76% of curve-related fatal crashes are single-vehicle run-off-road (ROR) crashes and 11% are head-on with an oncoming vehicle. A review of the Iowa DOT crash data indicates that in Iowa (2001–2005), 12% of all fatal crashes and 15% of all major injury crashes occurred on curves; 14% of all urban fatal crashes and 11% of all urban major injury crashes occurred on curves; and 11% of all rural fatal crashes and 19% of all rural major injury crashes occurred on curves.

Miaou and Lum (1993) studied heavy truck crashes using 1985–1989 Utah Highway System Information System (HSIS) data and evaluated horizontal curvature, vertical grade, and width of paved shoulder. They found that as vertical grade increased, truck accident involvement also increased. They also found that truck crash involvement increased as horizontal curvature increased, depending on the length of the curve.

Hauer et al. (2004) used a negative multinomial model using Washington HSIS crash data to predict the nonintersection accident frequency of urban, four-lane, undivided roads (1993–1996). They found no significant correlation between crashes and vertical grade.

Lamm et al. (1988) and Council (1998) found that crash rates increased as degree of curve increased, even when traffic warning devices were used to warn drivers of the curve.

Mohamedshah et al. (1993) found a nonintuitive negative correlation between crashes and degree of curve for two-lane roadways.

Council (1998) also found that the presence of spirals on horizontal curves reduced crash probability on level terrain, but did not find the same effect for hilly or mountainous terrain.

Vogt and Bared (1998) evaluated two-lane rural road segments in Minnesota and Washington State using HSIS data and found a positive correlation between injury crashes and degree of horizontal curve.

Shankar et al. (1998) evaluated divided state highways without median barriers in Washington State and found a relationship between the number of horizontal curves per kilometer and median crossover crashes.

Zegeer et al. (1992) evaluated 10,900 horizontal curves on two-lane roads in Washington State using a weighted linear regression model. They found that crash likelihood increased as the degree and length of curve increased. Alternatively, Deng et al. (2006) evaluated head-on crashes on two-lane roads in Connecticut for 720 segments using an ordered probit model. They included geometric characteristics in the analysis, but did not find that the presence of horizontal or vertical curves was significant.

The vehicle speed reduction required for traversing a curve has an impact on the frequency and severity of crashes on curves. Abrupt changes in operating speed resulting from changes in horizontal alignment have been suggested to be a major cause of crashes on rural, two-lane roadways (Lamm et al., 1988). Higher crash rates were experienced on horizontal curves that required greater speed reductions (Anderson et al., 1999). This finding was also supported by Fink and Krammes (1995), who indicated that curves requiring no speed reduction had no significantly different mean crash rates from their preceding roadway tangents. The roadway tangent length also influences driver behavior. The effect of a long tangent preceding a curve becomes more of a factor on sharp curves. Roadway tangent lengths also impact crash rates on steep downgrade curves. Crash rates on curves with long tangent lengths are more pronounced when the curve is located on a downgrade of 5% or more, with tangent lengths of more than 200 m.

Preston (2009) found that crash rate increases as radius decreases below 2,000 ft and that around 90% of fatal and 75% of injury crashes occurred on curves with radii less than 1,500 ft.

McLaughlin et al. (2009) evaluated ROR events using the VTTI 100-car NDS data. In that study, ROR crashes and events included those where one or more tires contacted a curb or left the roadway before returning to the roadway, where the vehicle departed the road and came to a stop, where the vehicle collided with a lane delineation object (e.g., curb, construction barrels), or where the driver braked hard and swerved to avoid a crash. The authors found a total of 122 ROR events, which included 28 crashes and 94 near crashes. They reported that 30% of the ROR events occurred on curves, 56% occurred on tangent sections, and 14% occurred at intersections.

### *Roadway Cross Section*

Lane width, shoulder type, shoulder width, median type, and median width have all been associated with crash experience. A summary of some of the available literature follows. Table 2.1 also summarizes the information.

Miaou and Lum (1993) studied heavy truck crashes using 1985–1989 Utah HSIS data and evaluated horizontal curvature, vertical grade, and width of paved shoulders. The authors found that as the width of the inside paved shoulder increased, truck involvement decreased.

Mohamedshah et al. (1993) used Utah HSIS data to model truck crash involvement on two-lane rural roads. The authors found a negative relationship between two-lane truck crashes and increased shoulder width.

Vogt and Bared (1998) developed an accident model for two-lane rural segments and intersections using Minnesota and Washington State HSIS data (1985–1989). The authors found a negative correlation between lane width and shoulder width and injury crashes.

Hauer et al. (2004) used a negative multinomial model using Washington State HSIS crash data to predict the non-intersection accident frequency of urban, four-lane, undivided roads (1993–1996). The authors found no correlation between crashes and lane widths. The range of lane widths modeled was 10–12 ft. Moreover, the authors found that roadway segments with two-way left-turn lanes (TWLTL) had fewer off-road crashes.

Garber and Ehrhart (2000) considered crash factors for two-lane roadways in Virginia with speed limits of 55 mph. The authors used deterministic models to relate crash rate with mean speed, flow per lane, lane width, and shoulder width. They found that the effect of mean speed, shoulder width, and lane width was negligible.

Deng et al. (2006) used an ordered probit model to analyze head-on crashes for 720 two-lane road segments in Connecticut (1996–2001). The authors found a positive relationship between narrow roadways and narrow road segments.

Zhang and Ivan (2005) evaluated the effect of geometric characteristics on head-on crash incidents for two-lane roads in Connecticut. The authors used negative binomial generalized linear models to evaluate the effects of roadway geometric features on incidents of head-on crashes for 655 segments using 1996–2001 crash data. They found a correlation between horizontal and vertical curvature but indicated that lane and shoulder width were not significant.

Zegeer et al. (1992) evaluated the safety effects of geometric improvements for 10,900 horizontal curves on two-lane roads

*Table 2.1. Summary of Literature for Roadway Cross Section*

| Authors | Data Set | Shoulder | Lane | Other |
|---|---|---|---|---|
| Miaou and Lum, 1993 | UT HSIS data: Heavy truck crashes | Negative correlation: Inside paved shoulder width and truck crashes | | |
| Mohamedshah et al., 1993 | UT HSIS data: Heavy truck crashes | Negative correlation: Width and truck crashes | | |
| Vogt and Bared, 1998 | MN and WA HSIS data: Rural two-lane | Negative correlation: Width and crashes | Negative correlation: Width and crashes | |
| Hauer et al., 2004 | WA HSIS data: Urban four-lane undivided | | No correlation: Width and crashes | |
| Garber and Ehrhart, 2000 | VA data: Two-lane roads | No correlation: Width and crash rate | No correlation: Width and crash rate | |
| Deng et al., 2006 | CT: Head-on crashes on two-lane roads | | Positive correlation: Narrow roadway and crashes | |
| Zhang and Ivan, 2005 | CT: Head-on crashes on two-lane roads | No correlation: Width and crashes | No correlation: Lane width and crashes | |
| Zegeer et al., 1992 | WA: Horizontal curves on two-lane roads | | | Correlation: Superelevation deficiencies and crashes |
| Heimbach et al., 1974 | Two-lane highways | Lower crash rate for paved than for unpaved section | | |
| Sosslau et al., 1978 | | Negative correlation: Width and crash rate | | |
| Zegeer et al., 1981 | KY: State primary, secondary, and two-lane roads | Decrease in ROR, head-on, and opposite direction side-swipe crashes for gravel or paved shoulder width increase from 0 to 9 ft | | |
| Abboud and Bowman, 2001 | AL: 2- and 4-foot paved shoulders on two-lane roads | No correlation: Paved shoulder and crashes | | |
| Souleyrette, 2001 | IA: Rural two-lane, four-lane, expressways | Could not detect relationship | | |

in Washington State. The authors found a statistical relationship between crash occurrence for sharper curves, narrower curve widths, locations with lack of spiral transitions, and increased super-elevation deficiencies.

Heimbach et al. (1974) found that rural, two-lane highways with paved shoulders had a significantly lower crash rate than those with unstable shoulders.

Sosslau et al. (1978) found that paved shoulders exhibit safety benefits. This NCHRP report concluded that roads with paved shoulders have lower crash rates than unpaved shoulders of the same width. The report also concluded that shoulder widths, paved or unpaved, have a greater effect on crash rates than lane widths. A linear model was developed to predict crash rates for roadways with varying lane and paved shoulder widths. The model was generally able to represent predicted relationships, but there were some inconsistencies. In general, crash rates decreased as shoulder widths increased. This rule applied for sections of roadways with three degrees or less of horizontal curvature, but the opposite result was true for roadways with an average daily traffic (ADT) of fewer than 1,000 vehicles per day (VPD) or more than 5,000 VPD.

Zegeer et al. (1981) conducted a comparative analysis study of state primary, state secondary, and rural, two-lane roads in Kentucky. The sections were selected so that they did not include any major intersections. A database of 15,944 miles of road was compiled from computer tape, and eight classifications of roads based on ADT were used. Because about 70% of the total sample had no shoulder, shoulders were defined as paved or densely graded. Grass and soil were not considered shoulders because they are not suitable for driving. Zegeer et al. found that ROR, head-on, and opposite-direction side-swipe crash rates decreased as shoulder width increased from

0 to 9 ft, but the crash rates increased slightly for shoulders from 10 to 12 ft wide. Crash severity, however, did not decrease with wider shoulders. Zegeer et al.'s results indicated that it is economically beneficial for roadways with lane widths greater than 10 ft to widen the shoulders if there are at least five ROR or opposite-direction crashes in one year. For roads without shoulders, the optimal shoulder width to install was found to be 5 ft.

Not all studies have concluded that paved shoulders offer a significant benefit, however. Abboud and Bowman (2001) evaluated 2- and 4-ft paved shoulders on two-lane highways in Alabama and compared them against county statistics for the expected number of crashes on the treated segments. Crash records were not kept on specific routes with similar characteristics; therefore, total county crashes in the before and after periods were used as a control. Crash frequency by type and severity was analyzed, but no statistically significant differences were found at the 0.05 alpha confidence level.

Similarly, a study conducted by Souleyrette (2001) was unable to present significant results. Souleyrette's study focused specifically on rural, two-lane and rural, four-lane, divided, noninterstate freeways in Iowa. Only targeted crashes were considered for this study. Intersection, median, and roadway crashes were excluded because they were assumed to be nonshoulder related. Limited data availability because of conservative shoulder construction practices in Iowa prevented statistical significance from being obtained with any of the results. Trends of reduced crash rates were noted but could not be verified with confidence.

Shankar et al. (2004) used a zero-inflated negative binomial model to consider the interaction among design, traffic, and weather on roadside crashes using 318 segments. The authors found that weather and design factors play a statistically significant role in roadside crash occurrence. The authors also found that shoulder width and presence of a divided median were related to crash occurrence. In another study, Shankar et al. (1998) analyzed 275 sections of divided state highways and found that median width was a statistically significant factor in crash history.

### Driveway Density

Vogt and Bared (1998) developed an accident model for two-lane rural segments and intersections using Minnesota and Washington State HSIS data (1985–1989). The authors found a positive correlation between driveway density and injury crashes.

Deng et al. (2006) used an ordered probit model to analyze head-on crashes for 720 two-lane road segments in Connecticut (1996–2001). Among other factors, the authors found that nighttime crashes and density of access points were significantly related to more severe crashes.

### Roadway Lighting

A number of studies have demonstrated that nighttime crash rates are significantly higher than daytime crash rates and that lighting can play a positive role in reducing nighttime crashes. Deng et al. (2006) used an ordered probit model to analyze head-on crashes for 720 two-lane road segments in Connecticut (1996–2001). Among other factors, the authors found that nighttime crashes and density of access points were significantly related to more severe crashes.

A before-and-after study of lighting along a five-lane roadway in Chicago from 1952 to 1958 was reported by Lewin et al. (2003), who found a reduction of 48% in fatal night crashes. Billion and Parson (1962) compared crashes on 6 miles of unlighted and 6 miles of lighted major routes with mountable medians. The night/day crash rate per million miles was 1.5 times higher for unlighted sections than for lighted. Another study in Illinois compared the night crash rate before and after a major traffic route was lighted. A night crash reduction of 36% was recorded (Box, 1989). A New York study compared lighted and unlighted major and collector streets. The study reported that streets with little or no illumination had substantially higher night–day crash ratios (Box, 1972).

Elvik (1995) conducted a meta-analysis of 37 published studies, reported from 1948 to 1989 in 11 countries, that evaluated the safety effects of lighting. Analysis of the studies indicated roughly a 65% reduction in nighttime fatal accidents, a 30% reduction in injury accidents, and a 15% reduction in property-damage-only (PDO) accidents for both intersections and roadway segments on rural, urban, and freeway facilities when lighting was installed. The effect of installing lighting was greater at intersections than at nonintersections; similar results were found for rural, urban, and freeway environments.

A comparative study in the Netherlands reported that the ratio of night/day crashes for unlighted rural freeway routes was 28% greater than for lighted routes (International Commission on Illumination, 1992). A recent Iowa State University/Center for Transportation Research and Education (ISU/CTRE) study evaluated rural expressway safety. The researchers did not evaluate lighting per se but evaluated other safety aspects of rural expressways, such as variation in medians and older driver issues, which may be of interest in evaluating potential safety benefits of lighting (Maze and Burchett, 2004).

Two studies were found that evaluated lighting on rural primary routes. Sabey and Johnson (1973) evaluated 43 sites on trunk highways before and after lighting. The authors found a statistically significant reduction (50%) in crashes for 19 of the roads that were high-speed (70+ mph) segments. They found no statistically significant reduction for lower-speed segments. Another study by Cornel and MacKay found no statistical difference in night and serious crash frequencies before and after lighting was installed on rural highways (FHWA, 1982).

### Rumble Strips

Rumble strips have also been found to reduce lane departure crashes. Table 2.2 summarizes the results of the literature review.

Hanley et al. (2000) evaluated four crash-reduction factors currently used by the California Department of Transportation (Caltrans), including rumble strip installation, defined as any construction for which a laterally positioned rumble strip had been installed. In most cases, the researchers indicated that some shoulder widening occurred as well. They found statistically significant accident-reduction factors with rumble strip installation.

Garder and Davies (2006) evaluated the effectiveness of continuous shoulder rumble strips (CSRS) on reducing crashes on rural interstates in Maine. The authors found that the presence of CSRS reduced crashes overall by 27%, reduced sleep-related ROR crashes by about 58%, and reduced dry-road ROR crashes by about 43%. They also found that fatal crashes were reduced more than other crashes.

Smith and Ivan (2005) evaluated the crash reduction resulting from milled-in shoulder rumble strips on limited-access highways within a 3-year period before and after installation on sections of 20 freeways, including some sections without rumble strips. The authors found that shoulder rumble strips overall reduced single-vehicle, fixed-object crashes by 33%. They indicated that crashes were reduced by as much as 48.5% within interchange areas and by as little as 12.8% on sections where the speed limit was less than 65 mph. They also found that crashes increased in areas where rumble strips were not installed.

Corkle et al. (2001) summarized eight research studies on edgeline rumble strips (ERS) and found that ROR crashes were reduced by 20% to 72%.

The New York State Department of Transportation (NYSDOT) began installing continuous shoulder rumble strips in 1993. It began to include continuous shoulder rumble strips with its regular construction and as site-specific projects on existing roadways. The New York State Thruway Authority (NYSTA), which owns and operates private toll roads, installed continuous shoulder rumble strips between 1992 and 1996. The advantage of the data drawn from the NYSTA installations was uniformity, because the data were recorded by a dedicated troop of the state police force and there were a limited number of miles from which to collect data. Both agencies had a limited amount of before-and-after data, so statistical significance was not tested, but both agencies found a reduction in crashes of 65% to 70%. It should be noted, however, that some observations were made during years that included construction of a "[non-]significant percentage" of continuous shoulder rumble strips (Perrillo, 1998).

Rumble strips were installed on 80% of the Pennsylvania Turnpike between 1989 and 1994. Early results after the first five projects were completed found a 70% reduction in ROR

*Table 2.2. Summary of Literature for Rumble Strips*

| Authors | Data Set | Shoulder/Edgeline Rumble Strips | Centerline Rumble Strips |
|---|---|---|---|
| Garder and Davies, 2006 | MN | Crash reduction: Overall, 27%, sleep-related ROR, 58%; dry-road ROR, 43% | |
| Smith and Ivan, 2005 | CT: State highways | Crash reduction: SV, fixed object, 33%; ROR, 48.5% | |
| Corkle et al., 2001 | Summarized 8 studies | Crash reduction: ROR, 20 to 72% | |
| Perrillo, 1998 | NY: State highways | Crash reduction: Overall, 65% to 70% | |
| Hickey, 1997 | PA turnpike | Crash reduction: ROR, 70%; drift-off-road, 60% | |
| Miles et al., 2005 | TX: Two-lane roadway | Reduced shoulder encroachments, 46.7% | |
| Persaud et al., 2004 | 7 states: Rural two-lane | | Crash reduction: All injury, 14%; front and opposing-direction sideswipe injury crashes, 25% |
| Russell and Rys, 2005 | Summarized other studies | | Crash reduction: Injury, 15%; head-on and opposing-direction injury, 25% |
| Kohinoor and Weeks, 2009 | AZ: Arterials, minor arterials, and collectors | | Crash reduction: Fatal and serious injury crashes, 61% |
| Outcalt, 2001 | CO: Two-lane | | Crash reduction: Head-on, 34%; opposite sideswipe crashes, 36.5% |

crashes. After speculation of regression to the mean and other factors affecting the results, a follow-up study was conducted. The study included all reportable accidents from 1990 to 1995 and found a slightly more modest result of a 60% reduction in drift-off-road (DOR) crashes (Hickey, 1997). These results, however, were not tested for statistical significance.

A preliminary study (Miles et al., 2005) was conducted to determine the extent of the benefits received by ERS. The study was conducted on a two-lane road in Texas with an 11-ft travel lane in each direction separated by a 4-ft-wide center segment with centerline pavement markings. Before-and-after data were collected along this 5-mi segment of road between September 10 and 22, 2004, and between November 5 and 17, 2004, respectively (Miles et al., 2005). The geometry of the roadway limits the applicability of the findings to a typical two-lane rural road and the brief study time period limits the conclusiveness of the results, but the study still provides an interesting insight into the operational effects of ERS.

The study by Miles et al. (2005) used rumble strips that were 12 in. wide; 4 in. was on the marked edge line and 8 in. was on shoulder pavement. Pneumatic road tubes were used to collect volume, speed, and lateral position data. Video footage was also collected to classify the shoulder encroachment maneuvers and determine if the ERS caused any erratic maneuvers by drivers. A total of 2,985 shoulder encroachments were observed during the 13 days of before-installation footage and the 13 days of after-installation footage. No erratic maneuvers were observed in the video data. Statistical t-tests were performed on the data to determine significance at the 95% confidence level for any changes in driver behavior (Miles et al., 2005).

The data from Miles et al. (2005) revealed an overall reduction in shoulder encroachments during the after period of 46.7%. When broken down by encroachment type, the "other" case experienced the greatest proportional decrease in shoulder encroachments. The "other" case included "inadvertent contact with the edge line because of natural lane shifting, driver inattention or fatigue, swaying motions of trailers, or large load width." Encroachments classified as "other" are categorized as one of four types ranging from "right tires hit," where only the right tires contact the rumble strips, to "around," where both sets of tires completely cross over the rumble strips (Miles et al., 2005). While the number of encroachments decreased, the lateral position of vehicles increased in distance beyond the edge line. This was not statistically significant, however, and standard deviations were large.

Persaud et al. (2004), using empirical Bayes, analyzed about 98 treatment sites (210 mi) on rural, two-lane roadways in seven states before and after installation of centerline rumble strips. The authors found a 14% reduction for all injury crashes combined (at a 95% confidence level), a 25% reduction for front- and opposing-direction sideswipe injury crashes (at a 95% confidence level after installation), and an overall reduction in crashes of 12% (at a 95% level of significance).

Russell and Rys (2005) summarized the results of several studies and suggested that the use of centerline rumble strips reduced overall injury crashes by 15% and reduced head-on and opposing-direction crashes involving injury by 25%.

### *Roadway Delineation and Signing*

Sun et al. (2007) investigated the distribution of vehicle lateral position before and after implementation of edgeline markings on seven tangent and three curve sections of two-lane roads with less than 22-ft pavement widths in Louisiana. The authors found that after implementation of the edge lines, vehicles were more likely to move away from the pavement edge. They also found that centerline crossings increased at several sites during the daytime but decreased at night.

Donnell and Mason (2006) evaluated the operation effect of wider edge lines along curves (8 in. vs. 4 in.) in Pennsylvania. The authors compared differences in several operational metrics, including change in mean speed, lateral placement, encroachment frequency, and vehicle position in the travel lane. Results indicated that wider edge lines did not change the encroachment proportion, mean speed, or lateral position along curves.

Tsyganov et al. (2006) compared crash statistics for rural, two-lane highways in Texas with and without edge lines on roadways with 9-, 10-, and 11-ft travel lanes with shoulder widths less than 4 ft using crashes from 1998–2001. On sections with two or more accidents, highways without edge lines had an 8% higher mean accident ratio than similar sections with edge lines. The authors also found an increase in crash frequency with lane-width reductions on sections without edge lines, but not on roadways with edge lines. On curved segments, highways without edge lines had a 25.8% higher crash frequency than those with edge lines.

### *Pavement Edge Drop-off*

Evaluating fatal crashes in Georgia in 1997, Dixon (2005) randomly selected 150 two-lane rural fatal crashes on state and nonstate system roads. She estimated that in 38 of the 69 (55%) nonstate system fatal crashes, edge rutting or edge drop-off was present. The author also determined that edge drop-off appeared to be one of the crash causal factors for 21 of the 38 (55%) sites where there was drop-off. The study indicated that drop-off was from 2.5 to 5.0 in. on the rural highway edges.

The Federal Highway Administration (FHWA) estimated that approximately 160 fatalities and 11,000 injuries result from crashes related to edge drop-off each year in the United States (FHWA, 2004). Although a quantitative relationship between pavement edge drop-off and safety has not been derived, the U.S. Department of Transportation (USDOT) has suggested that a drop-off of 3 in. or more of vertical differential is considered unsafe (FHWA, 2004). AASHTO (1996) suggested that no vertical differential greater than 2 in. should occur between lanes. A study by Humphreys and Parham (1994) found that vertical drop-offs of 4 in. or more between the roadway surface and adjacent shoulder were unsafe. Zimmer and Ivey (1982) also showed that safety was related to pavement edge shape.

Hallmark et al. (2006) reviewed crash reports for Iowa and Missouri to determine whether crashes were related to pavement edge drop-off. The authors found that approximately 18% of rural ROR crashes in Iowa were potentially edge drop-off related. They also found that crashes that were potentially edge drop-off related were more likely to result in a fatal or major injury crash than other rural ROR crashes. In Missouri, the authors found that approximately 23% of rural ROR crashes were potentially pavement edge drop-off related. They also found a relationship between crashes and an edge drop-off of 2.5 in. or more.

McLaughlin et al. (2009) evaluated 122 ROR events using the VTTI 100-car NDS data. The authors reported that change in lane boundaries was involved in 22% of the events. This factor included start of median, narrowing of lane, lane drop, or unusual roadway geometry.

## Environmental Factors

Maze et al. (2006) used the 2005 FARS data to evaluate the impact of weather. They found that pavement condition is listed as rain, snow, or ice for only 12% of fatal crashes. However, as noted, this value does not represent the scope of the problem, because rain-, snow-, or ice-related incidents are only present during a small amount of driving time. For instance, in Iowa approximately 21% of crashes are winter-weather related. The amount of time snow or ice is present is significantly less than 21%. Additionally, fatal crash frequency during the winter in rural Iowa when pavement conditions are snowy or icy is about twice the fatal crash frequency when alcohol is a contributing factor.

Deng et al. (2006) used an ordered probit model to analyze head-on crashes for 720 two-lane road segments in Connecticut (1996–2001). Among other factors, the authors found a positive relationship between wet roadway surface and crashes.

Shankar et al. (2004) used a zero-inflated negative binomial model to consider the interaction between design, traffic, and weather on roadside crashes using 318 segments. The authors found that weather plays a statistically significant role in roadside crash occurrence and contributes to 19.3% of the likelihood of crash occurrence, while the weather and design interactions contribute around 6% to the likelihood of crash occurrence. Their results indicated that the presence of precipitation in the fall was positively correlated, and the presence of precipitation in the spring was negatively correlated with crash occurrence. The authors also indicated that average monthly snowfall exceeding 4 in. and the interaction between snow depth and horizontal curves had a statistically significant effect on roadside crash frequency.

McLaughlin et al. (2009) evaluated 122 ROR events using the VTTI 100-car NDS data. They reported the following among their findings:

- A ROR event is 2.5 times more likely to occur on dark unlighted roads than during daylight conditions;
- It is 1.8 times more likely on wet roads than dry;
- It is 7 times more likely on roads with snow or ice than on dry roads; and
- It is 2.5 times more likely to occur during the presence of precipitation (fog, mist, rain) than during clear conditions.

## Vehicle Variables

Vehicle type is relevant because rollover incidents may result in more serious outcomes for a ROR crash. Pickup trucks and sport-utility vehicles have a higher center of gravity, which may result in a different outcome for the same initial sequence of events during a road departure. Little information was found about which specific vehicle factors are related to ROR crashes. It is generally accepted that sport-utility vehicles and pickup trucks are more prone to rollover. However, little information was found that describes specific vehicle characteristics in relation to lane departure risk.

For naturalistic driving studies, most vehicle variables can be collected up front when the instrumentation packages are installed. It is important to provide representative distribution of vehicle types for the full-scale NDS.

## Driver Factors

### *General*

Spainhour et al. (2005) evaluated fatal crashes in Florida involving heavy trucks. The authors found that human factors were the primary contributing factor for 94% of the crashes, with the most common factors being alcohol/drug use, inattention, and decision errors.

Dissanayake (2003) used logistic regression to identify influential factors in young-driver (16 to 25 years old), single-vehicle ROR crashes. The author used crash data from 1997 and 1998 from police-reported crashes in Florida. Influence of alcohol or drugs, existence of a curve or grade, and vehicle speed significantly increased the probability of having a more severe ROR crash.

McGinnis et al. (2001) analyzed FARS and National Personal Transportation Studies data for ROR fatal crashes from

1975 to 1997. The authors evaluated how trends changed over time and found that young drivers, male drivers, drivers over the age of 70, drivers in utility vehicles, and drivers using alcohol had higher involvements in fatal ROR crashes.

Khattak and Hummer (1998) analyzed crashes on two-lane rural roadways from two counties in North Carolina. The authors indicated that consumption of alcohol, roadway surface condition, and horizontal alignment appeared to influence the occurrence of ROR crashes.

McLaughlin et al. (2009) evaluated 122 ROR events using the VTTI 100-car NDS data. The authors reported that in 40% of the events, the most common contributing factor was distraction/inattention. The most common distractor (90%) was secondary task distraction, which included use of a cell phone or dialing a cell phone, talking to or looking at passengers, or devoting attention to in-vehicle devices.

### *Younger Drivers*

Ulmer et al. (1997) examined the National Highway Traffic Safety Administration (NHTSA) General Estimate System (1993) for 16-year-old drivers and reported that 16-year-old drivers were more likely than other drivers to be involved in single-vehicle crashes and in crashes from 6:00 p.m. to 12:00 a.m. These teen drivers were also more likely to be accompanied by other teen passengers than were 17-, 18-, or 19-year-olds.

Williams et al. (1997) evaluated fatal crash involvement among 15-year-old drivers in states that required a learner's permit for 15-year-olds and found that crashes involving 15-year-old drivers were usually single-vehicle crashes, occurred late at night (between 12:00 a.m. and 6:00 a.m.), and had a number of passengers in the car. Driving factors that contributed to 15-year-old-driver fatal crashes included speeding and failure to drive in the proper lane.

A University of North Carolina study (Highway Safety Research Center, 2000) found that 80% of 16-year-old-driver nighttime crashes occurred between the hours of 9:00 p.m. and 12:00 a.m. and 73% of 17-year-old driver nighttime crashes occurred from 9:00 p.m. to 12:00 a.m. The crash risk for 16- and 17-year-old drivers was nearly three times greater between 10:00 p.m. and 12:00 a.m. than during the daylight hours. Based on the study, the risk per mile driven is even greater after midnight because most of the nighttime vehicle miles traveled (VMT) by 16- and 17-year-olds occurred before midnight. Ulmer et al. (1997) examined NHTSA's General Estimates System for 16-year-old drivers and found that 16-year-olds were more likely than other drivers to be involved in crashes from 6:00 p.m. to 12:00 a.m. Williams et al. (1997) evaluated fatal crash involvement for 15- and 16-year-olds and found that fatal crashes for 15-year-olds were more likely to occur between 12:00 a.m. and 6:00 a.m. Rice et al. (2004) evaluated how nighttime driving affected injury crash rates for young drivers in California before implementation of a graduated driver's license (GDL) in 1998 and found that crash risk increased after 10:00 p.m.

Adolescent impulsiveness is a natural behavior, but it results in poor driving judgment and participation in high-risk behaviors, such as speeding, inattention, drinking and driving, and not using a seat belt. Peer pressure also often encourages risk taking (Chein et al., 2011). According to NHTSA, risk taking among adolescents appears to be a critical factor in explaining the high number of crashes. For example, younger drivers tend to accept narrower gaps when pulling out into traffic. They also have been observed to have shorter following distances and to drive faster (Ferguson, 2003).

Williams (2001) reported on a study that indicated that for teenage drivers the presence of one passenger nearly doubles the fatal crash risk compared with driving alone. In another study, the fatal crash risk with two or more passengers was found to be five times as high as driving alone. There was excess risk for young drivers with passengers during both day and night hours (Williams, 2001). Another study indicated that the crash risk when three or more passengers were present was about four times greater than when driving alone (NHTSA, 2005b). The increased crash risk existed for both daytime and nighttime crashes, although overall crash risk was much higher at night. In one study, death rates from 10:00 p.m. to 6:00 a.m. were 1.74 times higher with passengers than without passengers. During the daytime, rates were 1.77 times higher (Williams, 2003). More teen fatal crashes occurred when passengers, usually other teenagers, were in the car than when no passengers were in the car. Two out of three teens who die as passengers are in vehicles driven by other teenagers (Williams, 2003).

## Summarized List of Factors

Table 2.3 summarizes roadway, environmental, vehicle, and driver factors that have been identified in the literature or through team expertise as being relevant to the occurrence and severity of lane departure crashes. Other factors that may be necessary to analyze lane departures using NDS data, such as factors to position a vehicle in respect to the roadway, factors to identify potential lane departures (triggers), or factors relevant to crash surrogates, are not identified.

## Research Questions

One of the goals of this project was to develop a set of research questions that could be explored using existing NDS data. The intent was to then determine which research questions could adequately be addressed given the data and the limitations of

*Table 2.3. Factors Contributing to Occurrence and Severity of Lane Departure Crashes*

| | | |
|---|---|---|
| **Roadway Factors** | | |
| Horizontal curves | Length | Radius or degree of curve |
| | Spirals | Superelevation |
| | Relationship to other curves | |
| Vertical curves | Length | Grade |
| | Relationship to other curves | |
| Roadway cross section | Lane width | Surface type |
| | Cross slope | Shoulder type and width |
| | Median type and width | |
| Signing | Presence and type | |
| Speed limit | Posted speed limit | Advisory speed limits |
| Roadway delineation | Presence and quality of pavement markings | Presence and type of overhead street lighting |
| Roadway defects | Pavement edge drop-off | Surface irregularities |
| | Surface friction | Road debris |
| Other | Driveway density | Sight distance |
| Clear zone | Type and location of objects within clear zone | Slope beyond edge of shoulder |
| | Presence of object delineators | Guardrail, barriers |
| Countermeasures not included in other roadway factors (e.g., paved shoulders) | Edgeline and center rumble strips | Speed feedback signs |
| | Additional delineation, such as channelizers, raised pavement markings | Cable median barrier |
| **Environmental Factors** | | |
| Pavement surface condition | Presence of snow, ice, rain, debris | Amount of snow, ice, rain |
| Ambient conditions | Time of day | Temperature |
| | Precipitation | Visibility (precipitation, fog, smog, dust) |
| **Vehicle Factors** | | |
| Vehicle characteristics | Size | Type (e.g., SUV, van) |
| | Width | Center of gravity |
| | Advanced technologies (e.g., lane departure warning system, OnStar) | Braking capabilities |
| Kinematic | Speed | Acceleration |
| **Driver Factors** | | |
| General | Age | Gender |
| | Driving experience | Aggressiveness |
| | Reaction time | |
| Condition | Fatigue | Medical condition |
| | Emotional state | |
| Teen-specific factors | Presence of passengers | Driver licensing requirements |
| Substance use | Alcohol | Illegal drugs |
| | Prescription drugs | |
| Distractions | Type of distraction | Duration of distraction |
| | Level of engagement in distraction | |

the SHRP 2 full-scale NDS. Three sets of questions are presented in the following sections:

1. The first set of questions includes those that were addressed in this report. These research questions reflect a need for information that would set the stage to answer research questions in the full-scale study. These questions could also be explored further in the full-scale NDS.
2. The second set of questions includes those identified through this research as being feasible for the full-scale study. Examples are also provided of more specific questions within those categories that the team feels can be realistically addressed, given what was learned during this research and what is known about the data expected from the full-scale NDS.
3. The third set of questions includes specific research questions that the team has determined cannot be realistically addressed given the data expected to be available from the full-scale study and given the review of existing NDS data.

The information that supports the research team's best estimate about what can or cannot be answered for the full-scale study is based on information provided in the following chapters. However, the research questions are placed in this section because they provide an overview of what follows in the rest of the report.

In order to identify lane departure research questions, the team first identified which driver, vehicle, roadway, and environmental factors were likely to contribute to the occurrence and severity of lane departure crashes, based on an in-depth literature review, as described in the section "Relevant Data Elements Identified in Existing Literature" (p. 10), and on the team's expertise in lane departure issues. The team then reviewed data from existing NDS from VTTI and UMTRI and evaluated the feasibility of extracting from these various vehicle, roadway, driver, and environmental factors. This information is provided in Appendices A and B. Chapter 4 summarizes data elements that are expected to be necessary to answer the research questions, reviews the roadway data elements identified by SHRP 2 Safety Project S03, Roadway Measurement System Evaluation, and reviews what is expected to be available from the instrumented vehicle study based on a review of information from Safety Project S05, Design of the In-Vehicle Driving Behavior and Crash Risk Study. The ability to extract data from existing NDS was also explored and summarized, as this ability relates to the full-scale NDS.

## Lane Departure Research Questions Addressed in Scope of Research

The first set of research questions includes those necessary to set the stage for answering research questions in the full-scale NDS. These questions were explored in this research, and the results are presented in the following sections. These questions may also be further addressed using data from the full-scale NDS.

***Research Question A-1:*** *What driver, vehicle, roadway, and environmental factors are necessary to answer a range of research questions related to lane departures using NDS and roadway data?*

Identifying data needs is an important step in determining which lane departure research questions can feasibly be answered. Roadway, driver, environmental, and vehicle factors expected to influence the occurrence and severity of lane departures was summarized using a literature review (see section "Relevant Data Elements Identified in Existing Literature," p. 10). Sources for the various data elements were identified based on the most current available information for SHRP 2 Safety Project S07, In-Vehicle Driving Behavior Field Study, and Safety Project S04B, Mobile Data Collection. Existing NDS from UMTRI and VTTI were examined, the data elements necessary to answer lane departure research questions were extracted, and the feasibility of obtaining the data was determined. This information is summarized in Chapter 4.

***Research Question A-2:*** *What kinematic variables can be used to identify lane departure incidents (e.g., lateral drift, lane departure, near crash)? For instance, a side acceleration of X ft or a roll rate of Y might define a lane departure.*

This question addresses the need to identify vehicle kinematic variables that can be used to flag lane departure incidents in the full-scale study. A significant amount of data will result, and it will be necessary to determine some method to flag potential incidents in an automated process. The team conducted an exploratory analysis of kinematic variables for normal driving, as well as for left- and right-side lane departures, using the UMTRI and VTTI data sets, as described in Chapter 5.

***Research Question A-3:*** *What environmental, roadway, driver, or vehicle factors influence whether a vehicle departs its lane?*

This research question addresses understanding environmental, roadway, driver, and vehicle variables that influence the occurrence of lane departures. Lane departures from the UMTRI data set were identified, and factors were extracted from the various corresponding data sets. Several different analyses were conducted using the UMTRI data that demonstrated approaches to answering this research question. The approaches are described in Chapter 6.

## Relevant Lane Departure Research Questions for Full-Scale NDS

The section provides general categories of lane departure questions that can be answered using the full-scale study. Examples are also provided of more specific questions within those categories that can be realistically addressed given what was learned during this research and based on what is known about the data expected to result from the full-scale NDS. This task was based on a review of information available as of September 2009, when the first draft of this report was submitted. The team is not aware of any additional information that alters its original assessment as of January 2010, which is the date for the final submission of this report.

This set of questions can be addressed by researchers for SHRP 2 Safety Project S08, Analysis of the SHRP 2 Naturalistic Driving Study Data. To address these questions, researchers will need data from instrumented vehicles (Safety Project S07), as well as data that will be gathered or collected during the roadway data collection effort (Safety Project S04A, Roadway Information Database Developer, Technical Coordination, and Quality Assurance for Mobile Data Collection, and Safety Project S04B). The determination of what can be addressed is also based on reviewing and extracting variables from the UMTRI and VTTI NDS data.

***Research Question B-1:*** *What environmental, roadway, driver, or vehicle factors influence whether a vehicle departs its lane?*

This research question addresses understanding driver, roadway, environmental, and vehicle variables that influence the occurrence of lane departures. More specific research questions that could be answered under this general topic include the following:

- How does roadway surface condition affect lane departure frequency?
- Are lane departures less likely when pavement markings are highly visible?
- Does signing have any impact on frequency of road departures (e.g., large chevrons may make a driver alert to an adverse horizontal curve)?
- How does roadway lighting affect driver scanning patterns at night and what is the impact on lane departures?
- What curve characteristics influence the likelihood of a lane departure?
- What role does distraction play in lane departure frequency?
- What is the relationship between speed and lane departures on curves?
- How does alcohol consumption influence driver response to changes in roadway geometry, and what is its impact on lane departures?
- Are drivers of sport-utility vehicles and pickup trucks more likely to engage in aggressive driving behaviors (e.g., speeding, overtaking), and what is the impact of such behaviors on likelihood of lane departures?

***Research Question B-2:*** *What environmental, roadway, driver, or vehicle factors influence lane departure outcome?*

This research question involves understanding driver, roadway, environmental, and vehicle variables that influence the outcome of a lane departure when one occurs. More specific research questions that could be answered under this general topic include the following:

- How do weather conditions affect lane departure outcome?
- Are drivers who leave the roadway more likely to recover and safely return to their lane on paved shoulders than on gravel or earth shoulders? How much of an impact does shoulder width have on outcome?
- What is the relationship between speed and lane departure outcome?
- Are drivers in vehicles without automatic braking systems (ABS) more likely to overcorrect when encountering snow/ice or loose shoulder material than drivers in vehicles with ABS?
- How does level of driver forward scanning before a lane departure influence the likelihood of recovery?
- How do drivers react when encountering various types of slope beyond the shoulder edge, and how do these reactions affect lane departure outcome?
- What factors lead to driver overcorrection, and what is its impact on lane departure outcome?

***Research Question B-3:*** *What is the impact of lane departure countermeasures on lane departure frequency and outcome?*

This research question addresses how drivers interact with countermeasures and addresses why countermeasures are or are not effective. More specific research questions that could be answered under this general topic include the following:

- Are drivers more likely to lane keep on roadways with edgeline rumble strips?
- How likely are drivers to overcorrect or counter-steer away from edgeline rumble strips and potentially encroach into an adjacent lane rather than experience a road departure?
- Are drivers who leave the roadway more likely to recover and safely return to their lane on paved shoulders than on gravel or earth shoulders? How much of an impact does shoulder width have on outcome?
- Do edgeline or centerline rumble strips have the same impact on distracted drivers as on nondistracted drivers?

- Are drivers more likely to travel at unsafe speeds during winter storm events when median cable barriers are present?
- Does additional delineation affect driver forward attention on curves? What is the impact of delineation on frequency and outcome of lane departures?

***Research Question B-4:*** *What is the relationship between lane departure crash surrogates and crashes?*

One of the main advantages of the full-scale NDS is that it will provide a unique opportunity to develop relationships between crash surrogates and crashes. Agencies frequently are unable to conduct crash analyses to compare the impact of a treatment for a number of years after the treatment is installed, or they have few sites for comparison. As a result, it is often difficult to conduct crash analyses. Understanding the relationships between lane departure crash surrogates and lane departure crashes would provide agencies with opportunities to conduct evaluations sooner.

Answering this research question can provide some information on the potential effectiveness of countermeasures such as edgeline or centerline rumble strips or treatments that reduce speeds on curves. Specific research questions developed under this category may also indicate what factors positively affect the outcome of a lane departure.

## Relevant Research Questions That Cannot Feasibly Be Addressed in the Full-Scale NDS

Several factors expected to be correlated to lane departure crash frequency and severity will not be collected in any data sets available during the full-scale NDS. Other factors may be available, but extraction may be infeasible. The team has determined that certain research questions cannot be addressed or cannot be realistically addressed given the data expected to be available during the full-scale study and based on a review of existing NDS data.

Highly relevant research questions related to the occurrence, frequency, and severity of lane departure crashes that cannot be answered include those concerning the following:

- Occurrence or level of alcohol use by the driver (i.e., what is the relationship between blood alcohol level and frequency of lane departures?): The instrumented vehicle's data acquisition system (DAS) is expected to have an alcohol sensor that will indicate the presence (but not amount) of alcohol in the vehicle and will not be able to isolate the user. Driver alcohol use may be inferred if the driver is the sole occupant.
- Occurrence or level of drug use by the driver (i.e., does the driver's illegal drug use similarly affect the frequency of lane departures?): No sensors are available that will pick up drug use or identify drugs in the driver's system.
- Pavement friction (i.e., what is the relationship between lane departures on curves and pavement surface friction?): Pavement surface friction is unlikely to be collected using the mobile mapping vans. Even if collected, surface friction will change over the course of the full-scale study, depending on such factors as wear and winter maintenance.
- Impact of pavement edge drop-off on lane departure outcome: Data on pavement edge drop-off is not likely to be collected by the mobile mapping van because drop-off can change over short periods of time. As a result, recording the presence and amount of drop-off at one time period may not reflect conditions at a future time period. For instance, there may be several inches of drop-off during the time the mobile mapping van collects data, but shoulder maintenance could occur several days later and thus change conditions drastically.
- Quantitative measure of rain, snow, and ice on road: The presence of rain, snow, or ice can be determined from the instrumented vehicle forward video or from environmental records, but the amount of precipitation on a given stretch of roadway cannot be measured.

CHAPTER 3

# Data Sets Used

The following sections describe the data sets used throughout this report for the various analyses. Appendix A provides a detailed description the variables included in the UMTRI data set, as well as a description of how a number of variables were extracted. Appendix B provides a description of variables available in the VTTI data set.

## University of Michigan Transportation Research Institute Field Operational Test In-Vehicle Data

Several field operational tests were conducted by UMTRI, including a road departure crash warning (RDCW) system. The system involved mounting instrumentation packages on 11 vehicles (of the same make and model). In each of the studies, vehicles were instrumented with a variety of sensing systems, including a forward video and a driver's face video, forward radar and side radar, and a global positioning system (GPS). The RDCW system also used a camera to record visual features that delineated the lane and road edges, and radars that monitored the lane edge. The main advantage of this data set was that the researchers were able to archive all the data so the database could be searched and specific data extracted.

The RDCW included a lane departure warning (LDW) system and a curve speed warning (CSW) system. The LDW system used a forward-looking monochrome camera to identify visual features near the lane edge. The image position of visual features and algorithms were used to calculate lane width, vehicle position within the lane, and relative motion within the lane. Other sensors, such as GPS, vehicle speed, brake position, and forward and side radars, were used to increase the accuracy and reliability of determining the lane position. The CSW system processed road geometry to estimate curvature and then, using vehicle speed and acceleration, computed a vehicle's most likely path and risk of leaving the curve (LeBlanc et al., 2006).

The RDCW system alerted the drivers when they drifted from their lanes or went too fast to safely negotiate a curve. The system was tested over 10 months with 78 drivers who were evenly split by gender and age. Data were collected for a one-week period prior to activation of the system for each driver. During the first week of driving, the RDCW was functional and recording data just as if the system were operational, but alerts were not provided to the driver. As a result, the first week of data collection reflected naïve driving with no in-vehicle warning system alerts.

The RDCW system included six levels of alerts that would have indicated to the driver that he or she was about to leave his or her lane or was traveling too fast on a curve. The alerts included right- and left-lane departure cautionary alerts, right- and left-lane departure imminent alerts, and cautionary CSW and imminent CSW alerts. A seventh designation was used to indicate that a vehicle was negotiating a curve, but this did not include any alert.

The research team requested data from the road departure crash warning (RDCW) field operation test (FOT) for the one-week period prior to activation of the RDCW system for all instances when one of the six alerts was recorded. Data for all rural roadways were requested. Rural data are defined in the UMTRI FOT as data from a location with a population less than 50,000 persons.

Data for instances when alerts were recorded were used as starting points to identify potential lane departures. This information was used as described in Chapter 5 to develop thresholds for crash surrogate events. The team also requested data on regular driving that did not involve any alert.

UMTRI provided data in the form of a database, as well as of forward imagery. Data were provided for 44 different drivers and were divided by alert type. The research team received over 2,000 alerts (1,506,525 rows of data). Each alert included approximately 600 rows of data at 10 Hz (one row represents 0.1 s). Data were provided for 30 s before the alert was recorded and approximately 30 s after. Instances of a vehicle negotiating

Image source: Esri. © 2010 i-cubed. Data source: UMTRI RDCW data set.

***Figure 3.1. Example of vehicle trace.***

a curve were also provided as samples of regular driving and included approximately 60 s of data. The database contained a number of data fields (columns) with data from the instrumentation system, such as lateral acceleration and forward speed. The data fields included are described in Appendix A. The data corresponding to each alert are referred to in the following sections as "vehicle trace," as shown in Figure 3.1.

Forward imagery was provided for each vehicle trace. Images were provided at 2 Hz (2 per s or 1 image per 5 rows of vehicle trace data) during times when an alert was not recorded. Data were provided at 10 Hz (10 per s or 1 image per row of vehicle trace data) for the 4 s before and 4 s after an alert was recorded. Imagery was provided as compressed JPEG images. Images had to be lined up with the corresponding rows of data. An example forward view is shown in Figure 3.2.

Hereafter, this data set will be referred to as the "UMTRI naturalistic driving study data set" or "UMTRI data set."

Data included several roadway types, including rural freeway, rural freeway ramp, rural expressway, rural four-lane, rural two-lane paved, and rural two-lane gravel. Since a large amount of data was provided, the team focused on two-lane roadways to meet project goals and deadlines. The lane tracking system did not perform well on unpaved roadways, so the study further focused on two-lane, paved roadways.

Source: UMTRI RDCW data set.

***Figure 3.2. Example of forward video view.***

The data set included a database file with the following variables:

- Driver number;
- Trip number;
- Alert time;
- Time;
- Alert ID;
- Alert type;
- Age;
- Gender;
- Curve (present or not);
- Right and left boundary types (type of lane marking);
- Latitude and longitude (used to establish vehicle position);
- Heading;
- Available maneuvering room, right and left (distance to left and right measured by radar);
- Brake (brake engaged or not);
- Engaged (cruise control engaged or not);
- Lane offset (vehicle offset from lane center);
- Lane offset confidence;
- Lane width;
- Track width (width of vehicle; consistent, since same vehicle model was used);
- Speed;
- Lateral speed;
- Side and forward acceleration;
- Roll, roll rate, pitch rate, yaw rate;
- Solar angle;
- Wiper (e.g., off, low);
- Headlamp (off, on, parking, high);
- Road class (roadway type: unknown, limited use, major surface, minor surface, local);
- Curve advisory speed, if present;
- Posted speed limit;
- Curve radius;
- Distance to curve point of intersection;
- Annual average daily traffic (AADT); and
- Number of lanes.

A variable was also included for the widths of the left and right shoulders. However, all shoulders were indicated in the data as being 5 m, so it was assumed that this variable was incorrect. Shoulder width was then measured using the forward imagery instead.

## Michigan Geographic Framework and Sufficiency Report

The Michigan Geographic Framework (MGF) is the digital base map for the state of Michigan. Public roads are one of the many features maintained as part of this framework. Roadway attributes include linear referencing descriptors, road name, address ranges, functional class, and legal ownership. The location of most roadway-based data is described using the MGF linear references.

The Michigan Department of Transportation (MDOT) sufficiency log is an annual report created by the Office of Transportation Planning for the trunk line (state-maintained) highways. The sufficiency report includes a broad range of roadway attributes. Sample attributes include MGF-based linear references, road type, surface and shoulder type, surface and shoulder width, number of through and turn lanes, traffic volume, and various sufficiency-based ratings.

The databases were obtained as part of another project at CTRE. MDOT confirmed through e-mail that the researchers are allowed to use the data set for the research described in this report.

## Transportation Crash Master

The Transportation Crash Master is an extract of MDOT's crash report information system (CRIS) database. It contains general information about crashes on all public roads, including attributes for up to three vehicle units. Crash location is provided through both MGF-based linear references and derived geographic coordinates.

The database was obtained as part of another project at CTRE. MDOT confirmed through e-mail that the researchers can use the data set for the research described in this report. One stipulation of using the data is that crashes cannot be shown in a map in any document or presentation. Data were available for 2000 to 2006.

## Aerial Imagery

Aerial imagery from Esri Corporation was also used. The team had access to a program in Esri's geographic information system (GIS), ArcMap, which interactively brings up aerial imagery for a location where other spatial data are already available. The imagery is hosted by Esri and available online (see http://resources.esri.com/arcgisonlineservices). The data come from a variety of sources, including U.S. Geological Survey (USGS) Digital Orthophoto Quarter Quadrangles (DOQQs). Figures 3.3 and 3.4 show vehicle traces overlaid with aerial imagery in ArcMap. Google Earth was also used as a source for aerial imagery.

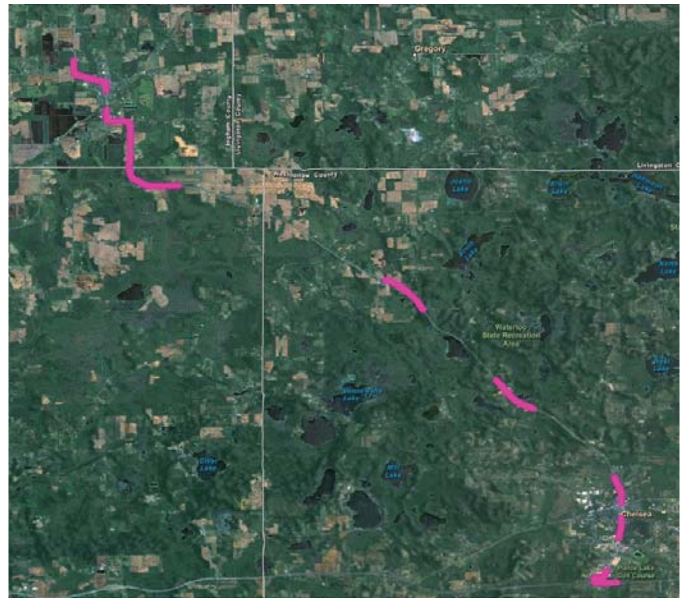


Image source: Esri. © 2010 i-cubed. Vehicle trace source: UMTRI RDCW data set.

***Figure 3.3. Vehicle traces overlaid with aerial imagery in ArcMap.***

Image source: Esri. © 2010 i-cubed. Vehicle trace source: UMTRI RDCW data set.

***Figure 3.4. Single-vehicle trace overlaid with aerial imagery in ArcMap.***

## VTTI Naturalistic Driving Study from Data Request

The team requested all rural lane departure crashes, near crashes, and incidents from the VTTI 100-car naturalistic driving study that occurred in open country or on the interstate. A sample of nonevent data was also requested. The team requested reduced data from events; raw data from the instrumented vehicle's data acquisition system (DAS); and forward, side, and back video. The team also requested samples from nonevent driving. The team received a total of 33 crashes or near crashes on roadways classified as open country or interstate run-off-road. No incident or nonevent data were received. It is unknown if the data represented all run-off-road crashes and near crashes.

A number of variables were requested for the raw data. The raw data received included the following variables:

- Trip time;
- Trip time elapsed;
- Event start and end time;
- Side and forward acceleration;
- Speed;
- Throttle (throttle position);
- Available maneuvering room forward and rear (included range, rate, and azimuth);
- Brake (on or off);
- Turn signal state (off, right, left);
- Wiper state;
- Light condition;
- Urban/rural setting; and
- Weather condition.

A variable with latitude and longitude was provided, but its data fields were invalid and, as a result, vehicle data could not be spatially located. Other data from the DAS that the team believed were available and requested but were not received included the following:

- Vehicle position within lane and lane width from lane tracking;
- Curve advisory speed and radius;
- Volume;
- Posted speed limit;
- Shoulder width; and
- Shoulder type.

A video clip was provided for each event that included approximately 45 s of data. Images showing a forward, rear, and right-side outside view were provided. It is assumed that the forward view provides as much coverage (distance outside the front of the vehicle viewed in the image) as was available in the original data set. It is also assumed that the image resolution is the same as was available in the original VTTI data set and was not degraded.

## VTTI Naturalistic Driving Study from the Internet

VTTI recently made some data from their 100-car study publicly available on a data distribution website (http://forums.vtti.vt.edu/index.php?/files/category/3-100-car-data/). The following data were obtained:

- Continuous data: Contains data around each crash or near crash at 10 Hz.
- Event eyeglance data: Has recording of participant eyeglance location for each crash or near crash.
- Event video reduction data set: Contains reduced data from the video for crashes and near crashes and has information such as event nature, event type, and precipitating event.
- Event narrative: Has narrative for each crash/near crash.

This data set had data for each of the 33 crashes/near crashes provided by VTTI. Some events were also available for eight additional crashes/near crashes that may be used in the analysis of lane departures. The data set provided some additional information, such as throttle position, lane markings, and distance to objects left and right.

## SHRP 2 Full-Scale Instrumented Vehicle Study

SHRP 2 is in the process of implementing a large field study of instrumented vehicles driven by naïve drivers. The naturalistic driving study will instrument approximately 2,000 vehicles in six states (Indiana, Pennsylvania, Florida, New York, North Carolina, and Washington) (Campbell, 2009). The instrumentation will be left in place for 12 or 24 months.

Roadway data will also be obtained through several sources. Roadway databases from study states will be obtained. SHRP 2 also plans to collect a limited amount of high-resolution roadway data using deployed mobile mapping units.

A description of the data that are expected to be collected and that are relevant to answering lane departure research questions is provided in Chapter 4.

## Summary of Terms Used to Describe Data

- **Continuous data:** This type of data is described more in Chapter 6. Continuous data are naturalistic driving study data reported at the resolution at which they were collected. For instance, each row represents one observation of vehicle driving for 0.1 s (10 Hz). This is also referred to in some studies as "time series data."
- **Dynamic:** Dynamic variables are those variables that change in the short term. These include vehicle kinematic variables such as speed, acceleration, or position. In reality, most roadway variables would not change over the course of the study, but since the roadway a driver is currently traversing will change in the short term, roadway characteristics are considered to be dynamic variables. All environmental variables are also considered to be dynamic.
- **Event:** This is described more in Chapter 8. An event is an interval of time centering on a situation of interest, such as a lane departure. For instance, an event may consist of 30 s of data before and after a lane departure occurs.

- **Frame:** A frame is one row of data from the continuous data. At 10 Hz this represents a 0.1-s interval.
- **Incident:** An incident is an occurrence of a situation of interest, such as a lane departure.
- **Static:** Static is a way to describe variables that do not change over the course of the study, such as driver gender, and results from preinstrumentation driver surveys. Age is likely to change during the course of the study, but for all intents and purposes age can be considered a static variable. Vehicle static variables include vehicle type, track width, and center of gravity.
- **Vehicle trace:** This is used to describe the intervals of data provided in the UMTRI data set. Each vehicle trace consisted of approximately 60 s of data at 10 Hz (around 600 rows of data), which could be spatially located.

CHAPTER 4

# Roadway, Driver, Environmental, and Vehicle Data Needs and Limitations to Address Lane Departures Using Naturalistic Driving Study Data

## Background

This chapter discusses the roadway, driver, environmental, and vehicle factors that must be included to address the research questions outlined in Chapter 2 (“Research Questions,” p. 17). Factors that are expected to be relevant to lane departure crashes were identified through a review of available literature, as well as through the team’s expertise in lane departure issues.

Many factors deal with roadway features and the correlation between roadway countermeasures and lane departures or lane departure crashes. The intent of including research questions related to roadway features is to provide roadway agencies with information about the roadway factors that positively or negatively influence the likelihood of a lane departure so that agencies can better address safety in roadway design and assess the benefits of various countermeasures, such as rumble strips, flattening or better delineating curves, and mandating paved shoulders on reconstruction and rehabilitation projects.

Several environmental factors were also identified as contributing to lane departures and lane departure outcome. A number of driver factors are also relevant, such as driver characteristics (e.g., age, gender, driving experience), distraction, and the driver’s emotional or physical state (e.g., medical condition, alcohol or drug use). Vehicle variables are those that may cause a vehicle to be more or less likely to depart its lane and affect the subsequent outcome, such as vehicle braking system, center of gravity, and vehicle size. Finally, some additional variables that are not used to evaluate the likelihood of a lane departure or outcome are nonetheless necessary to identify vehicle state (e.g., yaw rate, vehicle position). These variables also need to be included in data streams resulting from the full-scale NDS. Hence, they are included in subsequent discussions about necessary data variables.

In order to answer these questions and address the relationship between lane departures and roadway, driver, environmental, and vehicle factors in the full-scale study, data needs should be identified and limitations in data quality, availability, or accuracy should be addressed. This chapter identifies factors that are expected to be relevant, identifies sources and limitations from a review of existing naturalistic driving study data (UMTRI and VTTI), and identifies data limitations that are expected to be relevant to the full-scale study.

Documentation describing the data sources in the full-scale in-vehicle naturalistic driving study was also reviewed. The expected data elements from mobile mapping (SHRP 2 Safety Project S03, Roadway Measurement System Evaluation) and the in-vehicle instrumentation (Safety Project S05, Design of the In-Vehicle Driving Behavior and Crash Risk Study) relevant to lane departures were identified. The availability of the data in the full-scale driving study is also commented on, and limitations are identified.

The data sets used to evaluate the feasibility of extracting necessary data elements to answer the stated lane departure research questions are described in Chapter 3. A number of variables were reduced by the team from the data sets described in this chapter. A detailed description of how variables were extracted is provided in Appendices A and B.

The next section, “Review of Roadway, Environmental, and Vehicle Data Elements Available in Existing Naturalistic Driving Study Data” (p. 28), summarizes the review of existing data sources to determine whether the necessary data elements could be extracted. This section describes the minimum roadway, driver, environmental, and vehicle data elements necessary to answer the research questions. The expected accuracy and resolution requirements are also discussed. Additionally, the availability of the data in the existing data sets and the limitations in extracting these data are discussed. An indication of the accuracy and resolution that would be desirable is also provided.

The team first reviewed the various data sets that are currently available, as described in the previous sections, and commented on their adequacy for answering the lane departure research questions.

Another section below, “Review of Planned Data Collection for Full In-Vehicle Naturalistic Driving Study” (p. 39), discusses the review of SHRP 2 Safety Projects S03 and S05 documents that describe the data collection systems for the proposed full-scale naturalistic driving study. The research team's understanding of the relevant data collection sensors and techniques and the expected accuracy and frequency of data collection are summarized. The data elements were compared with the requirements set out in the next section, "Review of Roadway, Environmental, and Vehicle Data Elements Available in Existing Naturalistic Driving Study Data." The adequacy and limitations of the methods, data accuracy, and data collection frequency for answering lane departure research questions are discussed. The following summarizes information for roadway, driver, environmental, and vehicle data elements.

It should be noted that the review of the full-scale data collection methods was based on the Safety Project S03 and S05 documents that were available to the research team as of September 2009. The review was also based on the team's understanding of the different sensors/methods. The data review was completed before the draft version of this report was provided to SHRP 2 in September. The team has reviewed any information that has become available during the review period for this report, and as of January 2010, the team has had no additional information that changes the findings presented here.

The naturalistic driving study data from UMTRI and VTTI were used to evaluate the variables that may be the most useful in setting triggers to identify lane departure events and to assess what thresholds may be used. Data were reduced as described in Appendices A and B.

The UMTRI data resulted in a number of encroachments, but no conflicts or crashes. Only data for rural, paved, two-lane roadways were included. The VTTI data included near crashes and crashes, but no encroachments. Additionally, variables were not consistent between the two data sets. As a result, the two data sets were evaluated separately, as discussed in the following sections.

## Review of Roadway, Environmental, and Vehicle Data Elements Available in Existing Naturalistic Driving Study Data

The ability to answer the research questions depends on obtaining the appropriate data about driver, roadway, environmental, and vehicle factors that will probably affect the likelihood of a lane departure, as well as on obtaining data that are needed to determine crash surrogate thresholds and vehicle position. Data at the appropriate resolution are also necessary to develop measures of exposure.

Data from the existing data sets described in Chapter 3, including naturalistic driving study data from UMTRI and VTTI, were reviewed to determine whether necessary data elements could be extracted. This section describes the minimum roadway, driver, environmental, and vehicle data elements necessary to answer the research questions. The accuracy and resolution requirements needed are also discussed. The availability of the data in the existing data sets and the limitations in extracting these data are discussed. An indication of the accuracy and resolution that would be desirable is also provided. Exposure factors are also included.

Each data variable has a list of possible sources. For instance, some types of roadway data could be obtained from aerial imagery, mobile mapping, state databases, or even the forward video from the naturalistic driving study data acquisition system. In the course of this research, the team often compared information from one source to another as a check. The team encourages researchers who will use the data from the full-scale naturalistic study and other data sets to do the same.

### Identification of Necessary Variables

At a minimum, factors that should be included in causal relationships are those that have already been identified in other studies. A comprehensive literature review was conducted, and a list of potential variables that affect the likelihood and severity of lane departure crashes is reported in Chapter 2 ("Relevant Data Elements Identified in Existing Literature," p. 10).

These factors were reviewed with the research questions in mind, and a list of driver, roadway, environmental, and vehicle factors that were determined to be important in addressing lane departures were summarized based on this information and the expertise of the research team.

In addition to factors that are expected to positively or negatively affect the likelihood of a lane departure crash, some other information will also be necessary, such as vehicle position and vehicle kinematics. Vehicle kinematics is necessary to identify triggers that can be used to flag lane departure events in the full-scale study.

Resolution and accuracy were determined on the basis of the team's experience, common accuracies for the metric, or expert opinions. For instance, superelevation is typically from 2% to 12%. As a result, an accuracy of at least ±0.5% seems logical. The desired accuracy of the lane tracking system was specified as 0.1 m. The lane position tracking system is critical for addressing lane departure questions. Several experts were questioned about the level of risk of different lane departure events. The experts unanimously agreed that even one tire leaving the paved roadway surface onto a grass, gravel, or mixed-surface shoulder constitutes a highly dangerous situation. As a result, the lane position tracking system should be accurate and reliable enough to determine when one or more tires have departed the roadway surface. The average tire width is 6 in. (0.15 m), so a desirable accuracy of ±0.1 to 0.15 m was specified.

## Vehicle Factors Needed to Answer Lane Departure Research Questions

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

### *Data Element: Vehicle Spatial Position (Latitude, Longitude)*

- **Need:** Establishes vehicle position so that data can be linked among spatial databases and spatial relationship between subject vehicle and features established.
- **Potential source for data element:** GPS.
- **Desired accuracy:** Standard GPS accuracy is approximately 2 to 5 m. Accuracy of 5 m is sufficient to locate vehicle data to a roadway, but higher accuracy is necessary to determine where a vehicle is at a particular point in time relative to roadway features collected from other sources.
- **Desired frequency:** 10 Hz.
- **Comments on extracting data from existing data sets:** UMTRI was uncertain about the accuracy of the GPS system in their RDCW system but estimated the accuracy to be between 2 and 3 m (from e-mail correspondence). When vehicle traces were overlaid with aerial imagery and correlated with forward video, the spatial location of the vehicle appeared quite accurate. GPS data were not provided with the VTTI data set.

### *Data Element: Distance Between Vehicle and Roadside Objects*

Outside objects are generally those that a vehicle may strike, such as a utility pole or other vehicle, once they leave their original lane of travel.

- **Need:** Establishes vehicle position relative to objects so that time to collision can be calculated and level of risk assessed.
- **Potential source for data element:** (1) Distance can be determined using spatial location of vehicle and object, or (2) distance to objects can be determined using forward or side radar if the object is within the range of the forward or side radar.
- **Desired accuracy:** ±3.0 ft (0.914 m). If a vehicle were traveling at 60 mph (80.67 ft/s) and the nearest strikable fixed object it may hit were within 3 ft, the error in calculating total transfer capability would be 3 ft ÷ 80.67 ft/s = 0.0372 s. For a vehicle traveling 35 mph, the error would be 0.058 s.
- **Resolution:** NA.
- **Comments on extracting data from existing data sets:** Even if objects can be located with a high level of accuracy, vehicle position from GPS accuracy is only likely to be at best ±6.56 ft (2 m), so the ability to correctly measure distance is constrained by limitations in GPS accuracy. With use of radar, accuracy is determined by that of the forward or side radar units. The drawback to this method is that the radar can only indicate that an object is within the range of the radar. It will be necessary to identify the object using forward or side video.

### *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 the departure angle from the roadway and the 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.
- **Desired accuracy:** It is not specifically stated, but it appeared that the accuracy of the lane tracking software in the UMTRI data was 0.328 ft (0.1 m). Since this is less than the width of an average tire (around 6 in.), it is expected that this accuracy is sufficient.
- **Resolution:** Collection of vehicle position at 10 Hz is adequate to establish angle of departure and offset.
- **Comments on extracting data from existing data sets:** Lane position was not provided with the VTTI data set. Lane position in the UMTRI data set was given in terms of a measure of lane width and offset from the lane center at a given point in time. Vehicle width was known and constant among all vehicles. Using these three variables, a vehicle's position within its lane could be determined as shown in Figures 4.1 and 4.2. The data provided in the UMTRI data set were determined to be adequate to extract data necessary to answer the research questions. Lack of some type of lane positioning information would seriously affect the ability to determine when crash surrogate events occurred.

### *Data Element: Longitudinal Acceleration ($a_x$) and Speed ($v_x$)*

- **Need:** Magnitude of acceleration (positive or negative) can indicate an evasive action and can be used as a measure to determine thresholds between levels of crash surrogates. Acceleration rates can also be used as indicators of aggressive driving.
- **Potential source for data element:** Longitudinal acceleration and speed are measured from an accelerometer or are output from an on-board system. These data were provided with both the UMTRI and VTTI data sets. Brake engagement data were also provided, which can be used

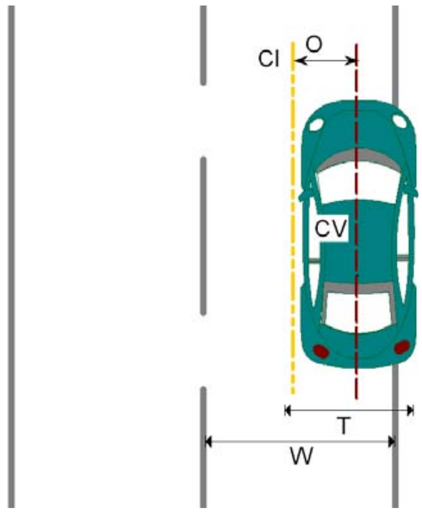



*Figure 4.1. Determination of lane departure and amount of shoulder encroachment.*

to indicate that a vehicle is braking (decelerating) but not to provide the magnitude of braking.

- **Desired accuracy:** 0.1 ft/s$^2$ and 0.1 ft/s (0.03 m/s$^2$ and 0.03 m/s). Acceleration is also frequently expressed in "*g*'s."
- **Resolution:** Data collected at 10 Hz should be sufficient.
- **Comments on extracting data from existing data sets:** None.

### *Data Element: Lateral Acceleration ($a_y$) and Lateral Speed ($v_y$)*

- **Need:** Indicate side movement, which can be used to determine when a lane departure has occurred and the severity of the lane departure. Lateral acceleration and speed are also used to determine roll hazard.
- **Potential source for data element:** Lateral acceleration and speed, usually measured from an accelerometer, were available in both the VTTI and UMTRI data sets.
- **Desired accuracy:** 0.1 ft/s$^2$ and 0.1 ft/s (0.03 m/s$^2$ and 0.03 m/s). Acceleration is also frequently expressed in "*g*'s."
- **Resolution:** Data collected at 10 Hz should be sufficient.
- **Comments on extracting data from existing data sets:** None.

### *Data Element: Pitch, Roll, Yaw*

- **Need:** Define vehicle rotation around several axes and are used to define levels between crash surrogate thresholds and assess roll hazard.
- **Potential source for data element:** Usually measured from an accelerometer.
- **Desired accuracy:** Unknown.
- **Resolution:** Data collected at 10 Hz intervals should be sufficient.
- **Comments on extracting data from existing data sets:** This data element was available in the UMTRI data set. No limitations were noted.

### *Data Element: Presence and Distance Between Subject Vehicle and Other Vehicles*

- **Need:** Establish outcome from lane departure and are 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.
- **Potential source for data element:** Forward or side video, forward or side radar.
- **Desired accuracy:** ±3 ft (0.914 m).
- **Resolution:** Collected as they occur.
- **Comments on extracting data from existing data sets:** Oncoming vehicles and vehicles that were passed or that passed the subject vehicle could be determined from the forward video from both UMTRI and VTTI. However, only a subjective measure of distance could be obtained from the forward video, as shown in Figure 4.3 (following closely, following, forward vehicle present but not following, no

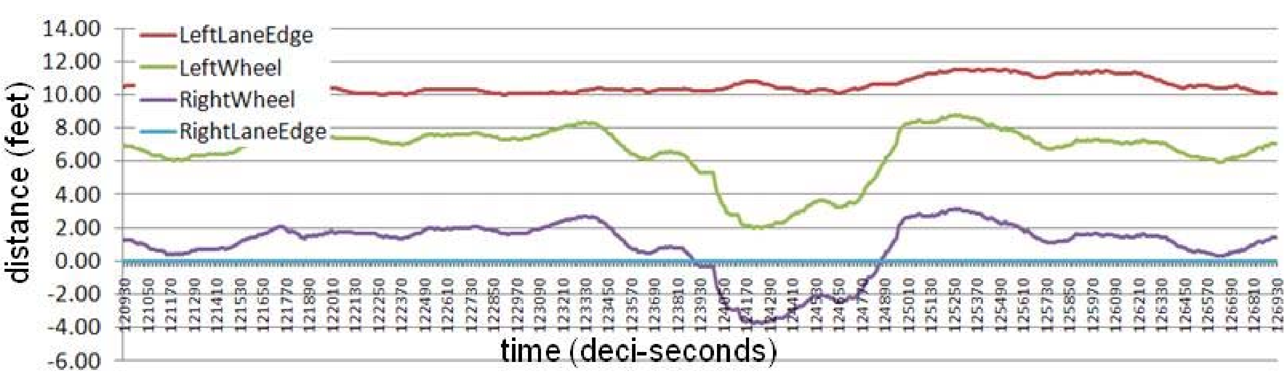



*Figure 4.2. Vehicle tire path calculated from UMTRI data (shows location and amount of encroachment onto shoulder).*

(a)

(b)

Image source: UMTRI RDCW data set.

***Figure 4.3. Subjective measurement of vehicle following: (a) subject vehicle is following forward vehicle and one oncoming vehicle is passing subject vehicle; and (b) subject vehicle not considered to be following forward vehicle.***

forward vehicle); distance could not be determined. Distance to a forward or side vehicle could be determined from the forward or side radar. However, only vehicles within the radar range could be detected.

## Roadway Factors Needed to Answer Lane Departure Research Questions

The following section summarizes roadway factors necessary to address lane departure research questions, indicates potential sources in the existing data sets, suggests accuracy and frequency needs, and includes comments about the accuracy and availability in the existing data sets. Several factors that are highly relevant but that will not be available in any of the data sets include the presence, type, and amount of pavement edge drop-off and pavement surface friction.

A number of studies have suggested that drop-off can contribute to outcome when a driver leaves the edge of the roadway and that pavement edge drop-off–related crashes tend to be more severe than other types of run-off-road crashes (Hallmark et al., 2006). The team provided this as a suggestion during discussions about the in-vehicle instrumentation package. However, the only method to collect these data with the in-vehicle instrumentation is to use a camera pointed at the roadway. This would capture presence and type of drop-off, but not its amount. Pavement edge drop-off could be collected with the mobile mapping units, but the amount and presence of drop-off can vary significantly over time. As a result, recording drop-off with the mobile mapping units would only be relevant for that point in time.

Surface friction is also an important factor in run-off-road crashes, particularly on curves. However, none of the vehicle instrumentation packages can capture this variable. It can and may be collected by the mobile mapping units. However, as with pavement edge drop-off, surface friction can vary significantly over time, especially in climates where winter weather maintenance is frequent. As a result, collection of this variable may only be representative of a particular point in time.

### *Data Element: Lane Width*

- **Need:** Independent variable in the statistical analysis. Also needed to establish vehicle position within its lane.
- **Potential source for data element:** When using lane width as an independent variable, data can be obtained from existing roadway data sets or from mobile mapping. Lane width is expected to be collected using the lane position tracking system. Lane width was available with the UMTRI data from the forward lane position tracking system.
- **Desired accuracy:** An accuracy of 0.328 ft (0.1 m) is likely the best that can be achieved with the forward lane position track system.
- **Resolution:** Data collected at 10 Hz should be sufficient.
- **Comments on extracting data from existing data sets:** The ability to measure lane width using the lane position tracking system is critical for establishing vehicle position within the lane and determining when and by how much a vehicle departs its lane. Lane width was not provided with the VTTI data set and could not be measured using any of the data provided. Data from the UMTRI system were adequate for research needs.

### *Data Element: Roadway and Shoulder Surface Type*

- **Need:** Independent variable in statistical analyses. The type of shoulder will also affect potential outcomes for lane departures.
- **Potential source for data element:** Existing roadway data sets or mobile mapping. Roadway and shoulder type could also be determined from forward video.
- **Desired accuracy:** Categorical data (should include asphalt, concrete, gravel, earth).
- **Resolution:** Several times per mile or when characteristics change.
- **Comments on extracting data from existing data sets:** While identification of features was possible with forward imagery from the VTTI and UMTRI data sets, color imagery would enhance the ability to distinguish features.

### *Data Element: Shoulder and Median Width*

- **Need:** Independent variable in statistical analyses. Shoulder and median width also affect potential outcomes for lane departures.

- **Potential source for data element:** Existing roadway data sets or from mobile mapping. Roadway and median width could also be measured using forward video from UMTRI when distances were calibrated.
- **Desired accuracy:** ±0.5 ft (0.152 m).
- **Resolution:** Several times per mile or when characteristics change.
- **Comments on extracting data from existing data sets:** Shoulder width was calculated in the UMTRI data set using side radar. The shoulder width measurement, however, was inaccurate because it was only measuring whether an object was located within the radar range. Shoulder and median width could be calculated when a distance was calibrated in the forward imagery.

### Data Element: Number of Lanes, Access Control, and Presence and Type of Median

- **Need:** Establish roadway type. An independent variable in statistical analyses.
- **Potential source for data element:** Mobile mapping, forward imagery, aerial imagery, roadway databases.
- **Desired accuracy:** NA.
- **Resolution:** Once per mile or when characteristics change.
- **Comments on extracting data from existing data sets:** Roadway type, number of lanes, and type of access control were provided with both UMTRI and VTTI data. Whether or not the roadway was divided was indicated, but no information about presence and type of median was included. Roadway characteristics could easily be determined from the aerial imagery as well.

### Data Element: Curve Length and Radius

- **Need:** Independent variable in statistical analyses. May also be used to assess roll hazard.
- **Potential source for data element:** Mobile mapping or aerial imagery.
- **Desired accuracy:** ±25 ft (7.62 m).
- **Resolution:** Once per curve.
- **Comments on extracting data from existing data sets:** Radius was measured from the lane position tracking system in the UMTRI data, but the data received were inaccurate. Radius and curve length were measured from aerial imagery. The accuracy of this method is not known.

### Data Element: Curve Superelevation, Lane Cross Slope

- **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.
- **Desired accuracy:** Maximum superelevation for areas with no ice and snow is 12%; for areas with snow and ice the maximum is 8%. Given these ranges, ideal accuracy is 0.5%, but it is unknown if this accuracy can be practically measured in the field. Under normal circumstances cross slope is 1.5% to 2%. Ideally, it would be necessary to measure this variable at 0.1% accuracy to determine differences, but this may not be practical.
- **Resolution:** Superelevation would need to be measured as it changes along a curve. Cross slope could be collected several times per mile or when characteristics change.
- **Comments on extracting data from existing data sets:** Superelevation and lane cross slope were not available in any data sets used and could not be extracted from other sources.

### Data Element: Curve Direction from Perspective of Driver (Curve Left or Right)

- **Need:** Independent variable in statistical analyses. Also important for determining the potential outcome of a non-crash lane departure.
- **Potential source for data element:** Needs to be determined for direction of travel. Potential sources are aerial imagery or mobile mapping. A forward video image can also be used to determine direction.
- **Desired accuracy:** NA.
- **Resolution:** Should be indicated once per curve.
- **Comments on extracting data from existing data sets:** None.

### Data Element: Distance Between Curves

- **Need:** Drivers may negotiate curves differently if they travel for some distance between curves rather than negotiate a series of curves. Also used as an independent variable in statistical analyses.
- **Potential source for data element:** Aerial imagery or mobile mapping. Distance could also be calculated from the vehicle instrumentation.
- **Desired accuracy:** ±25 ft (7.62 m).
- **Resolution:** Once per curve.
- **Comments on extracting data from existing data sets:** Could be determined from either aerial imagery or the UMTRI vehicle traces. Both were adequate.

### Data Element: Type and Characteristics of Curve Spirals

- **Need:** Independent variable in statistical analyses.
- **Potential source for data element:** Mobile mapping is the only reasonable method that can be used to determine the presence of spirals, along with their characteristics, such as radius and length.
- **Desired accuracy:** ±25 ft (7.62 m).

- **Resolution:** Once for each spiral.
- **Comments on extracting data from existing data sets:** The team reviewed aerial imagery but could not detect or measure spirals. No information on spirals was provided with any of the data sets, and this information could not be extracted.

### *Data Element: Amount of Grade (Percent), Length of Grade (ft or m), and Location and Characteristics of the Crown and Crest Vertical Curve*

- **Need:** Independent variable in statistical analyses. Also affects lane departure outcome.
- **Potential source for data element:** Existing roadway databases and mobile mapping are the best sources. Presence of vertical grade and where a vehicle is relative to a vertical curve can be determined from forward video, as shown in Figure 4.4. An estimate could be determined from topographic maps, but this would be time-consuming.
- **Desired accuracy:** 0.5% for grade and ±25 ft (7.62 m) for length.
- **Resolution:** Could be measured each time grade changes and at beginning and ending points on vertical curves.
- **Comments on extracting data from existing data sets:** Grade was not provided in any of the data sets reviewed. As indicated in Figure 4.4, grade could subjectively be determined in the UMTRI data set by viewing the forward imagery. The amount or length of grade could not be determined.

### *Data Element: Signing (Would Include Features Such as Overhead Beacons, Signals, and Other Traffic Control Signs)*

As a minimum, signs collected should include all traffic control signs (stop and yield), warning signs (e.g., overhead flashing beacons, curve warning, curve advisory speed, change in alignment warnings, as shown in Figure 4.5), railroad crossing signs and markings, and school crossing signs and markings.

- **Need:** Independent variable in statistical analyses.
- **Potential source for data element:** Mobile mapping, existing sign inventories, forward video.
- **Desired 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 along a curve, but it would not be necessary to know exactly where each 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 strikable 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.
- **Comments on extracting data from existing data sets:** The main limitation in the data sets reviewed was that they did not include most of the signing data. The UMTRI data set provided the speed limit and advisory speed when known, but no other sign information was available. A sign's presence could be detected in most cases in the forward imagery for the UMTRI data set. However, because forward imagery was only provided at 5 Hz (two images per second), in some cases depending on where the sign was relative to the vehicle's position when the image was taken, the lettering on the sign could not be distinguished (especially at night). In most cases, the vehicle was not close enough in one frame, and in the next frame the vehicle had passed the sign.

Source: UMTRI RDCW data set.

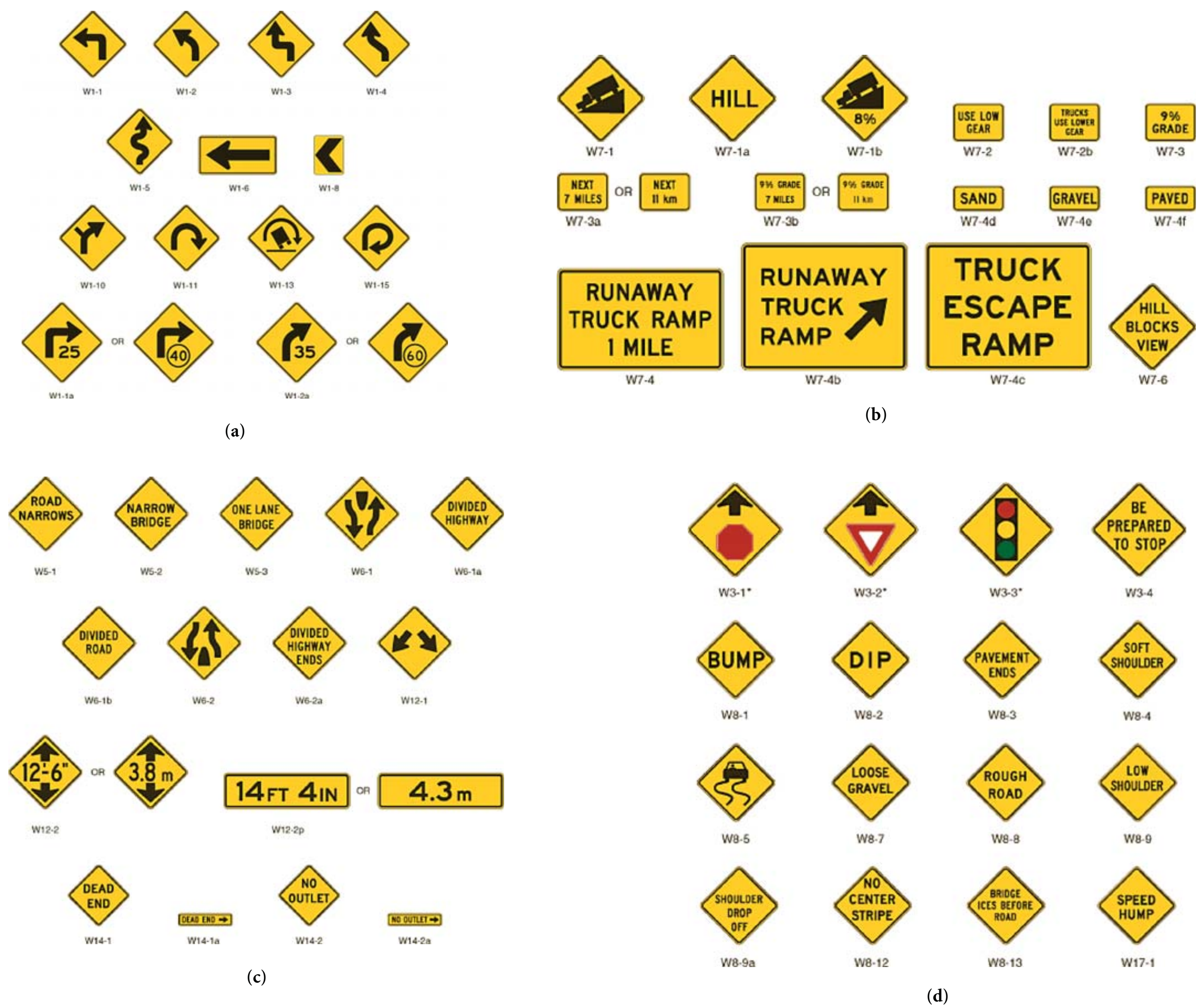
***Figure 4.4. Forward imagery indicating that vehicle is currently on an upgrade.***

### *Data Element: Number of Driveway or Other Access Points (Driveways/Mile, Access Points/Mile)*

- **Need:** Traffic entering and exiting the traffic stream can impact vehicle operation. This traffic would be included as an independent variable in statistical analyses.
- **Potential source for data element:** Mobile mapping, aerial imagery, forward imagery.
- **Desired accuracy:** NA.
- **Resolution:** NA.
- **Comments on extracting data from existing data sets:** Driveway and other access point densities for each vehicle trace in the UMTRI data set were determined by overlaying the traces with aerial imagery. Driveways and access points could be identified from the forward imagery in most cases, but this would be an extremely time-consuming method to extract the data.

### *Data Element: Presence and Type of Edge and Centerline Rumble Strips*

- **Need:** Independent variable in statistical analyses. Also needed to establish outcome of lane departure.
- **Potential source for data element:** Roadway databases, mobile mapping, and forward video.
- **Desired accuracy:** NA.
- **Resolution:** Once per mile or when rumble strip starts or ends.
- **Comments on extracting data from existing data sets:** This element would best be determined from mobile mapping,




Source: FHWA 2007.

***Figure 4.5. Types of signing to be included: (a) horizontal alignment warning signs; (b) vertical grade warning signs; (c) miscellaneous warning signs; and (d) roadway condition and advance traffic control warning signs.***

but it also can be determined from the forward video, as shown in Figure 4.6, particularly with color. However, characteristics such as width, depth, or skip distance would be difficult or impossible to extract from images.

### *Data Element: Roadway Delineation (Presence of Lane Lines or Other On-Roadway Markings)*

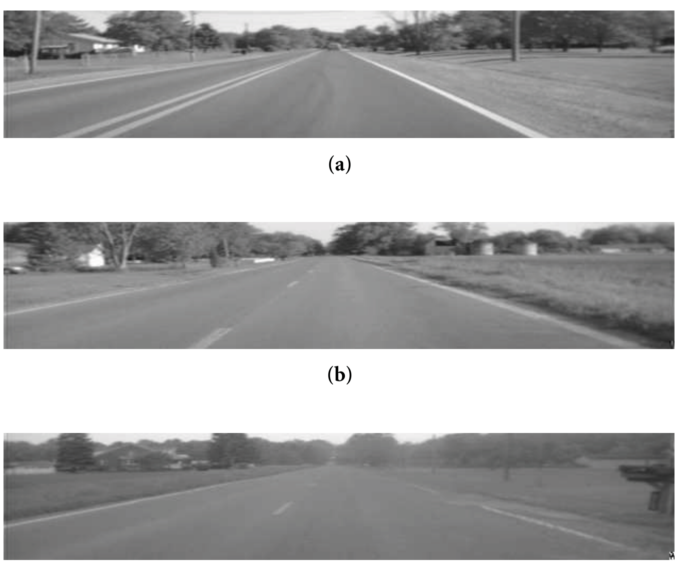
- **Need:** Critical for lane position tracking software. Would be included as an independent variable in statistical analyses.
- **Potential source for data element:** An initial estimate could be obtained from mobile mapping. Some states, such as Iowa, have good pavement marking inventories. These sources could be used as references. However, pavement markings can wear fairly quickly under adverse conditions, so a method to determine pavement marking condition current to each driving situation would be necessary. This information could only come from forward imagery and would need to be a qualitative assessment (i.e., highly visible, visible, obscured, not present). Figure 4.7 shows an example of a subjective measure.
- **Desired accuracy:** Data would include a quantitative estimate of visibility of markings.
- **Resolution:** Once per mile or as the situation changes.

*Figure 4.6. Image from DriveCam showing presence and type of rumble strips.*

- **Comments on extracting data from existing data sets:** This element needs to be current to the driving situation and can only be extracted from forward imagery. This information could be obtained from the UMTRI data set, but was more difficult with the VTTI data set because of image resolution.

### Data Element: Location and Type of Roadside Objects

- **Need:** Necessary to determine potential outcome of lane departures. May be included as an independent variable in statistical analyses. (Features such as guardrail along the edge of the roadway may impact driver behavior.)
- **Potential source for data element:** Mobile mapping would be the primary source. Some data can be obtained from aerial imagery. The presence of fixed objects can be identified in the forward imagery. Forward and side radar readings can be used to determine presence and distance of objects within range.
- **Desired accuracy:** Since location of fixed objects will be used to determine time to collision or potential outcome of a lane departure, a high level of accuracy is desirable. It is expected that ±3 ft (0.914 m) is sufficient.
- **Resolution:** Roadside objects should be collected when they appear; data should be collected for objects within the clear zone.
- **Comments on extracting data from existing data sets:** Only a limited number of fixed objects, such as trees, could be determined from the aerial imagery (image resolution was approximately 1 to 3 m depending on the area). Presence of fixed objects was also identified in the forward imagery from UMTRI. A rough estimate of distance from the edge of the roadway could also be made, but this was not accurate enough to assess the outcome of lane departures. Forward and side radars can indicate the presence but not the type of objects that are within the radar range. Additionally, only objects within the range of the radars can be identified.



(a)

(b)

(c)

Source: UMTRI RDCW data set.

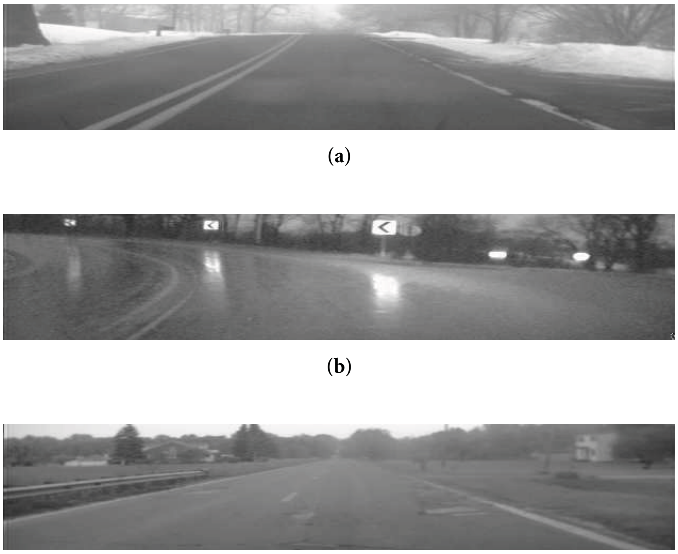
*Figure 4.7. Subjective measure of lane marking condition using forward imagery: (a) pavement markings indicated as "highly visible"; (b) pavement markings indicated as "visible"; and (c) right pavement markings indicated as "obscured."*

## Environmental Factors Needed to Answer Lane Departure Research Questions

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

### Data Element: Roadway Surface Condition (Weather Related, as Well as Presence of Roadway Irregularities Such as Potholes)

- **Need:** Independent variable in statistical analyses. May also impact potential outcome of lane departure.
- **Potential source for data element:** Forward- or other outward-facing video, status and frequency of wiper blades, outside temperature if available, roadway weather information system (RWIS) data if archived.
- **Desired accuracy:** Measure is subjective and therefore inapplicable.
- **Resolution:** Collected at 10-min intervals or as conditions change.



(a)

(b)

(c)

Source: UMTRI RDCW data set.

***Figure 4.8. Pavement surface condition from forward imagery: (a) snow present but roadway bare; (b) wet but amount of water cannot be determined; and (c) surface irregularities.***

- **Comments on extracting data from existing data sets:** A subjective measure of roadway surface condition and roadway irregularities could be obtained from both the VTTI and UMTRI forward video. Measures such as presence of water on the roadway can be determined as shown in Figure 4.8. Amount of water on the roadway or presence of ice and vertical elevation differences between lanes and shoulder (i.e., pavement edge drop-off) cannot be determined with any available data sources.

### *Data Element: Environmental Conditions Such as Raining, Snowing, Cloudy, and Clear (May Not Correspond to Roadway Surface Condition)*

- **Need:** Independent variable in statistical analyses. May affect sight distance and is related to visibility.
- **Potential source for data element:** Forward imagery or archived weather information, ambient temperature probe.
- **Desired accuracy:** Will be a subjective measure.
- **Resolution:** Collected at 10-min intervals or when conditions change significantly.
- **Comments on extracting data from existing data sets:** A general assessment of environmental conditions could be obtained from the forward video provided in the UMTRI and VTTI data sets. Even with wiper position known, it was difficult to tell how heavy the rainfall was. Archived weather information can provide general information for an area but cannot tell the exact environmental conditions in the location of the subject vehicle.

### *Data Element: Ambient Lighting, Includes Presence of Street Lighting*

- **Need:** Independent variable in statistical analyses.
- **Potential source for data element:** Sun angle, dawn, dusk, day, and night indicator can be obtained from time stamp data and U.S. Naval Observatory astronomical data. They can also be obtained from light meter and headlamp use.
- **Desired accuracy:** Only subjective measures will be used.
- **Resolution:** Can be recorded as the situation changes (day to dusk) or when significant changes occur during the day because of clouds. The measure can be somewhat generic (dark, dark with continuous lighting, dawn, dusk, daytime clear, or daytime with need for headlamps).
- **Comments on extracting data from existing data sets:** A relative estimate of ambient lighting could be obtained in most cases from the UMTRI and VTTI forward imagery. The limitations are that it was difficult during high cloud cover or low visibility to subjectively estimate ambient lighting.

### *Data Element: Visibility*

- **Need:** Independent variable in statistical analyses. Serves as a measure of sight distance and can also indicate surface conditions.
- **Potential source for data element:** Forward or other outside imagery is the only reasonable data source for visibility.
- **Desired accuracy:** Subjective variable, so accuracy is irrelevant.
- **Resolution:** Sampling at 10-min intervals would be sufficient.
- **Comments on extracting data from existing data sets:** This element was available from forward imagery in the VTTI and UMTRI data sets; however, in some cases it was difficult to tell whether visibility or image resolution was the problem, as shown in Figure 4.9. The cause of decreased visibility could not be determined. Low visibility is shown in Figure 4.10, but it is unknown if the cause is fog, smoke, or dust.

## Exposure Factors Needed to Answer Lane Departure Research Questions

The following section summarizes exposure factors necessary to address lane departure research questions, indicates poten-

Source: UMTRI RDCW data set.

***Figure 4.9. Image shows some reduced visibility, but it may be the result of sun angle or image resolution.***

Source: UMTRI RDCW data set.

***Figure 4.10. Low visibility appears to be caused by fog.***

tial sources in the existing data sets, suggests accuracy and frequency needs, and includes comments about the accuracy and availability in the existing data sets.

### *Data Element: Annual Average Daily Traffic*

- **Need:** Exposure measure.
- **Potential source for data element:** Roadway databases; most states have archived in some form.
- **Desired accuracy:** Most current year available.
- **Resolution:** NA.
- **Comments on extracting data from existing data sets:** This element could be determined from the crash data when no other source was available.

### *Data Element: Time Driving Into Trip*

- **Need:** Exposure measure.
- **Potential source for data element:** Vehicle data stream.
- **Desired accuracy:** ±1 s.
- **Resolution:** Is expected to be available in at least 1-s intervals.
- **Comments on extracting data from existing data sets:** NA.

### *Data Element: Amount of Driving on Different Roadway Types Under Different Environmental Conditions (Roadway Type, Rural vs. Urban, Dry vs. Snow, Snow vs. Ice, Ice vs. Wet)*

- **Need:** Exposure measure.
- **Potential source for data element:** Vehicle data stream.
- **Desired accuracy:** ±1 s.
- **Resolution:** Is expected to be available at least 1-s intervals.
- **Comments on extracting data from existing data sets:** NA.

### *Data Element: Density*

- **Need:** Exposure measure.
- **Potential source for data element:** Forward video.
- **Desired accuracy:** NA.
- **Resolution:** NA.
- **Comments on extracting data from existing data sets:** The number of oncoming vehicles, of vehicles that passed by the subject vehicle, or of vehicles that the subject vehicle passes can be counted using the forward and side imagery. Density can be calculated from 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.

### *Data Element: Lane Departure Crash Rate*

- **Need:** Exposure measure.
- **Potential source for data element:** State or local crash databases.
- **Desired accuracy:** NA.
- **Resolution:** NA.
- **Comments on extracting data from existing data sets:** Whether or not spatially located crash databases are available is the only limitation. Lane departure crashes along each vehicle trace were extracted from the Michigan Department of Transportation crash database. Crash density was calculated as crashes per mile. Crash data were unavailable for Virginia, and the vehicle traces were not spatially located.

## Driver Factors Needed to Answer Lane Departure Research Questions

### *Data Element: Age, Gender*

- **Need:** Needed for sampling. Included as an independent variable in statistical analyses.
- **Potential source for data element:** Driver questionnaire at beginning of study.
- **Desired accuracy:** NA.
- **Resolution:** NA.
- **Comments on extracting data from existing data sets:** Age and gender information are available in the UMTRI roadway data sets, as well as in VTTI video reduction data sets.
- **Limitations:** NA.

### *Data Element: Measures of Driver Riskiness*

- **Need:** Included as an independent variable in statistical analyses.
- **Potential source for data element:** The general riskiness of each participant can be obtained from the survey at the beginning of the study. The amount of time in hard acceleration or the amount of time exceeding the speed limit can be calculated from vehicle data.
- **Desired accuracy:** NA.
- **Resolution:** One riskiness level according to the Dula Dangerous Driving Index (DDDI) questionnaire for each trip/event.
- **Comments and limitations on extracting data from existing data sets:** The subject aggression levels were collected by DDDI questionnaires at the beginning of the VTTI 100-car

study, but the detailed results for each driver were unavailable. Acceleration and braking data can be obtained from the data sets, but speed limit information is available neither in the VTTI nor the UMTRI databases.

### *Data Element: Driver Distraction (e.g., Passengers, Cell Phone Usage, Eating)*

- **Need:** Included as an independent variable in statistical analyses.
- **Potential source for data element:** Driver distractions during the trip or event can be obtained directly from the driver's face video or from video reduction data.
- **Desired accuracy:** Image resolution (640 × 640 pixels) for the imagery data.
- **Resolution:** 15 Hz as the minimum and 30 Hz as preferred.
- **Comments on extracting data from existing data sets:** Driver distractions can be extracted from UMTRI driver's face video. The driver distraction information has already been extracted by VTTI from the driver's face video. The eye location data that can indicate type and duration of driver distraction, as well as narrative distraction information at the time of a near crash or crash, is provided by VTTI.
- **Limitations:** There are some missing values in the eye location data set of the VTTI data, which means eye location during some time intervals is unknown. This makes it hard to assess whether or not the driver was distracted.

### *Data Element: Kinematic Measures Before and After Incident, Such as Acceleration, Braking, and Steering*

- **Need:** Included as an independent variable in statistical analyses.
- **Potential source for data element:** Steering angle may be obtained from vehicle instrumentation; acceleration and braking information can be obtained from the time series data sets.
- **Desired accuracy:** Image resolution (640 × 640 pixels) for the imagery data.
- **Resolution:** 15 Hz as the minimum and 30 Hz as preferred.
- **Comments on extracting data from existing data sets:** The steering, acceleration, and braking information during near-crash and crash events was extracted from VTTI video data and provided in the video reduction data set by VTTI. The acceleration and braking information can be extracted from UMTRI foot/brake pedal video. Braking information is also available in both the VTTI and UMTRI roadway data sets.
- **Limitations:** Because there is no steering wheel sensor, information on steering is from video and is highly dependent on video view and quality for both VTTI and UMTRI data. Steering information for VTTI data is only available at the time of the near-crash and crash events but not at other times when nonincident lane departure events took place. Such information is hard to obtain because the face videos from VTTI that can provide steering information are inaccessible to us as a result of institutional review board (IRB) constraints.

### *Data Element: Alcohol and Drug Usage*

- **Need:** Included as an independent variable in statistical analyses; however, this only refers to drug or alcohol use if present in the driver video.
- **Potential source for data element:** Alcohol and drug use information can be obtained from video reduction data set.
- **Desired accuracy:** Image resolution (640 × 640 pixels) for the imagery data.
- **Resolution:** 15 Hz as the minimum and 30 Hz as preferred.
- **Comments on extracting data from existing data sets:** Alcohol and drug usage information at the time of near-crash and crash events was extracted and provided in the VTTI video reduction data set.
- **Limitations:** Alcohol and drug usage information for UMTRI data is inaccessible.

### *Data Element: Driver Fatigue*

- **Need:** Included as an independent variable in statistical analyses. It should be noted that it is not simple to determine what constitutes driver fatigue and how fatigue should be measured. Some researchers have suggested that the variable that should be measured is drowsiness. This variable is only mentioned here because it has been included in other naturalistic driving studies.
- **Potential source for data element:** Indicators of driver fatigue can be obtained from driver's face video or the video reduction data set.
- **Desired accuracy:** Image resolution (640 × 640 pixels) for the imagery data.
- **Resolution:** 15 Hz as the minimum and 30 Hz as preferred.
- **Comments on extracting data from existing data sets:** Indicators of driver fatigue information at the time of near-crash and crash events were extracted and provided in the VTTI video reduction data set. Eye location data from the VTTI database also provides information about the time and duration of drivers' eye closures, which can indicate fatigue during the trip. Driver fatigue indicators can also be extracted from UMTRI driver's face video.
- **Limitations:** NA.

### *Data Element: Lane Departure Intention*

- **Need:** Drivers may intentionally leave their lane for a number of reasons that may result in a conflict or crash (e.g.,

driver goes around stalled vehicle on shoulder or passes the vehicle).
- **Potential source for data element:** Data can only be obtained from lane position tracking algorithms and associated data streams such as forward video.
- **Desired accuracy:** NA.
- **Resolution:** NA.

### *Other Observations Regarding Data Elements*

The VTTI data sets provide gender and age information of the primary drivers but not for secondary drivers who also used the subject cars. Because there is no driver's face video provided by VTTI, demographic information of the secondary drivers could hardly be obtained. This is not a problem for UMTRI data, which present all needed demographic information in both the data sets and report appendix.

### Summary of Vehicle, Roadway, Environmental, and Exposure Factors

The factors discussed in the preceding sections are summarized in Tables 4.1 to 4.5. An indication of the priority for the data element is also provided.

## Review of Planned Data Collection for Full In-Vehicle Naturalistic Driving Study

The team first reviewed the various data sets that are currently available, as described in the previous sections, and commented on their adequacy for answering the lane departure research questions. The next step was to review any relevant information from SHRP 2 Safety Projects S03 (Roadway Measurement System Evaluation), S04A (Roadway Information Database Developer, Technical Coordination, and Quality Assurance for Mobile Data Collection), and S05 (Design of the In-Vehicle Driving Behavior and Crash Risk Study) that was available. Existing documentation that describes the instrumentation package, roadway data collection protocol, and description of other data sources for the full in-vehicle naturalistic driving study were reviewed, including the following:

- *Design of the In-Vehicle Driving Behavior and Crash Risk Study. Task 4: Sample Design Interim Report.* SHRP 2 Safety Project S05. Dingus et al. Virginia Tech Transportation Institute. November 2007.
- "Gaps Identified in the SHRP 2 Safety Program, Relative to Project S04." White paper. SHRP 2 Safety Project S03. John E. Hunt and Anita P. Vandervalk. Applied Research Associates, Inc., and Cambridge Systematics, Inc. Received February 2009. Includes appendices.
- "Sampling Thoughts for S05 Following the July 2007 SHRP Safety Research Workshop." White paper. Jim Hedlund. July 28, 2007.
- *SHRP 2 Safety Program Data Processing Steps.* Draft. November 4, 2008.
- *Design of the In-Vehicle Driving Behavior and Crash Risk Study. Task 9: Data System Interim Report (Task 6: Driver Face and Other Video Recording and Processing).* SHRP 2 Safety Project S05. Dingus et al. Virginia Tech Transportation Institute. September 3, 2008.
- *Design of the In-Vehicle Driving Behavior and Crash Risk Study. Task 9: Data System Interim Report (Task 7: Data Items and Instrumentation Package Specifications).* SHRP 2 Safety

***Table 4.1. Vehicle Factors***

| Data Element | Data Stream | Accuracy | Frequency | Priority |
|---|---|---|---|---|
| Vehicle position (latitude, longitude) | GPS | Best possible (±6.6 ft [2 m]) | 10 Hz | High |
| Distance between vehicle and strikable objects | Spatial location of vehicle/objects or radar | ±3.0 ft (0.914 m) | NA | High |
| Lane position, lane offset | Measured by lane position tracking system using forward- or other outward-facing video, GPS, and other data streams | ±0.1 ft (0.305 m) | 10 Hz | High |
| $a_x$ and $v_x$ | Accelerometer or On-Board Diagnostics (OBD) | ±0.1 ft/s$^2$ and 0.1 ft/s (0.0305 m/s) | 10 Hz | High |
| $a_y$ and $v_y$ | Accelerometer or OBD | ±0.1 ft/s$^2$ and 0.1 ft/s (0.0305 m/s) | 10 Hz | High |
| Pitch, roll, yaw | Accelerometer | | 10 Hz | High |
| Distance between vehicles | Imagery, radar | ±3.0 ft (0.914 m) | NA | High |

*Table 4.2. Roadway Factors*

| Data Element | Data Stream | Accuracy | Frequency | Priority |
|---|---|---|---|---|
| Lane width | Mobile mapping, forward video | ±0.328 ft (0.1 m) | 10 Hz | High |
| Roadway and shoulder surface type | Roadway data sets, mobile mapping, forward video | NA | NA | High |
| Shoulder and median width | Roadway data sets, mobile mapping, forward video | ±0.5 ft (0.15 m) | 10 Hz | High |
| Number of lanes, access control, presence and type of median | Mobile mapping, aerial imagery | NA | Once per mile or when characteristics change | Medium |
| Curve length and radius | Mobile mapping or aerial imagery | ±25 ft (7.62 m) | Once per curve | High |
| Superelevation, lane cross slope | Mobile mapping van | ±0.5%, 0.1% | Several times per mile | Medium |
| Curve direction | Forward imagery, aerial imagery | NA | Collected for each curve | High |
| Distance between successive curves | Mobile mapping data, aerial imagery | ±25 ft (7.62 m) | Once per curve | Medium |
| Type and characteristics of curve spirals | Roadway data sets, mobile mapping, forward imagery | ±25 ft (7.62 m) | Once per curve | Medium |
| Amount of grade (percent), length of grade, and location and characteristics of the crown and crest vertical curve | Roadway data sets, mobile mapping | 0.5% for grade and ±25 ft (7.62 m) | Begin and end points of grade change | Medium |
| Signing | Existing sign inventories, mobile mapping, forward imagery | ±6.6 ft (2.0 m) | Once per sign | High |
| Number of driveway or other access points | Mobile mapping, aerial imagery, forward imagery | NA | As needed | Medium |
| Presence and type of edge and centerline rumble strips | Roadway data sets, mobile mapping, forward imagery | NA | Start and end of rumble strip | High |
| Roadway delineation | Forward imagery | NA | Once per mile or as situation changes | Medium |
| Location and type of roadside objects | Mobile mapping data, aerial imagery | ±3.0 ft (0.914 m) | As they occur | Medium |

*Table 4.3. Environmental Factors*

| Data Element | Data Stream | Accuracy | Frequency | Priority |
|---|---|---|---|---|
| Roadway surface condition | Archived RWIS data, forward- or other outward-facing imagery, status and frequency of wiper blades, outside temperature | Will be qualitative measure | 10-min intervals or if conditions change | High |
| Ambient condition | Archived weather information, forward imagery | Will be qualitative measure | 10-min intervals or if conditions change | Low, if surface condition is collected |
| Ambient lighting including street lighting | Sun angle, dawn, dusk, day, night indicator can be obtained from time stamp data and U.S. Naval Observatory astronomical data, subjective measure from forward imagery | Will be qualitative measure | Once per mile or as conditions change | Medium |
| Visibility | Forward- or other outward-facing imagery | Will be qualitative measure | Once per mile | High |

**Table 4.4. Exposure Factors**

| Data Element | Data Stream | Accuracy | Frequency | Priority |
|---|---|---|---|---|
| AADT | Roadway data sets | Most current year available | NA | Medium |
| Time into trip | Vehicle data stream | NA | 10 Hz | Medium |
| Amount of time driving on various roadway types under different conditions | Vehicle data stream | NA | 10 Hz | High |
| Density | Forward/side imagery | NA | NA | High |
| Lane departure crash data | State or local crash databases | NA | NA | Medium |

Project S05. Dingus et al. Virginia Tech Transportation Institute. September 3, 2008.

- *Design of the In-Vehicle Driving Behavior and Crash Risk Study. Task 9: Data System Interim Report (Task 7: Data Items and Instrumentation Package Specifications—Appendices A and B).* SHRP 2 Safety Project S05. Dingus et al. Virginia Tech Transportation Institute. September 3, 2008.

The following sections describe the research team's understanding of relevant data collection sensors/techniques and the expected accuracy and frequency of data collection. The data elements were compared with the requirements set out in the section "Review of Roadway, Environmental, and Vehicle Data Elements Available in Existing Naturalistic Driving Study Data" (p. 28). The adequacy and limitations of the methods, accuracy, and data collection frequency for answering lane departure research questions are discussed. The following summarizes information for roadway, environmental, and vehicle data elements. This review is based on the information available to the research team as of January 2010.

## Review of Planned Data Elements from Mobile Mapping for Full-Scale Naturalistic Driving Study

The team reviewed a white paper developed for Safety Project S03 entitled "Gaps Identified in the SHRP 2 Safety Program, Relative to Project S04" by John E. Hunt, Applied Research Associates, Inc., and Anita P. Vandervalk, Cambridge Systematics, Inc. The white paper contains an appendix that lists items to be included in the data collection demonstration that potential vendors for Safety Project S04B were asked to collect. The team reviewed the data elements along with the expected accuracy and frequency, and the following section provides its comments about how well the data would answer the lane departure research questions discussed in previous sections.

The data elements for the data collection demonstration (rodeo) as indicated by Hunt and Vandervalk are provided in Tables 4.6 to 4.10. A description of the data features and data elements, along with the expected frequency and/or accuracy,

**Table 4.5. Driver Factors**

| Data Element | Data Stream | Accuracy | Frequency | Priority |
|---|---|---|---|---|
| Age and gender | Driver questionnaire | NA[a] | NA | High |
| Measures of riskiness | Questionnaire, roadway data sets | NA[a] | Once per trip/event | Medium |
| Driver distraction | Face imagery, video reduction data sets | Image resolution ($640 \times 640$ pixels) | 15 (minimum), 30 (preferred) Hz | High |
| Driver action before and after incident | Face imagery, roadway data sets, video reduction data | Image resolution ($640 \times 640$ pixels) | 15 (minimum), 30 (preferred) Hz | Medium |
| Alcohol and drug usage | Video reduction data | Image resolution ($640 \times 640$ pixels) | 15 (minimum), 30 (preferred) Hz | Medium |
| Driver fatigue | Face imagery, video reduction data | Image resolution ($640 \times 640$ pixels) | 15 (minimum), 30 (preferred) Hz | Medium |

[a]Not applicable because majority of measures are qualitative.

*Table 4.6. Final Rodeo Asset Data Elements*

| Feature | Data Elements | Frequency | Accuracy | Adequate for Lane Departure Research Questions |
|---|---|---|---|---|
| Barrier (presumably this includes median barriers and guardrail) | Barrier type, post type, end treatment type and location (roadside or median) | 100% | NA | Yes |
| | Begin and end location | | ±3 ft (0.914 m) | Yes |
| | Barrier height | | ±1 in. (0.025 m) | Yes |
| | Presence of offset bracket or rub rail | 100% | | Yes |
| On-street parking | Begin, end of parking | | ±3 ft (0.914 m) | Yes |
| | Location (right, left, both) | 100% | | Yes |
| Pavement markings | Begin and end point, type, centerline marking type | 100% | | Yes |
| | Marking offset | | ±1 in. (0.025 m) | Yes |
| | Retroreflectivity | | ±0.1 m cd/m$^2$ | Yes |
| | Location of special pavement marking | 100% | | Yes |
| | Description of special pavement marking | | ±3 ft (0.914 m) | Yes |
| | Presence and location of raised pavement markers | 100% | | Yes |
| Roadside obstacles | Type | 100% | | Yes |
| | Offset | | ±0.25 ft (0.076 m) | Yes |
| | Location | | ±3 ft (0.914 m) | Yes |
| Rumble strips | Location | 100% | | Yes |
| | Begin and end | | ±3 ft (0.914 m) | Yes |
| | Offset from edge of lane | | ±1 in. (0.025 m) | Yes |
| Sidewalk | Begin and end | | ±3 ft (0.914 m) | Yes |
| | Separated from road | 100% | | Yes |
| Signs | Support type, multiple signs, and sign type | 100% | | Yes |
| | Support location | | ±3 ft (0.914 m) | Yes |
| Street lighting | Location | | ±3 ft (0.914 m) | Yes |

is provided. Data elements not necessary to answer lane departure research questions are not included.

The positional accuracy for various data elements is listed as ±3 ft (0.914 m). Position of the vehicle and objects is most important in determining total transfer capability or distance to collision (DTC). If a vehicle were traveling 60 mph (80.67 ft/s) and the nearest strikable fixed object were located within ±3 ft, the error in calculating total transfer capability would be 3 ft ÷ 80.67 ft/s = 0.0372 s. For a vehicle traveling at 35 mph, the error would be 0.058 s. An error of 0.1 s would be acceptable for calculating time to collision, so the stated accuracy of the data collection is within that range. The distance error would be ±3 ft (0.914 m). However, the error from the vehicle position is not considered at this point.

With the exception of lane width, all data elements met or exceeded the desired accuracy that was determined to be necessary to answer lane departure research questions as defined in the section "Review of Roadway, Environmental, and Vehicle Data Elements Available in Existing Naturalistic Driving Study Data" (p. 28). The accuracy of lane width is stated as ±0.5 ft (0.152 m). Lane width is critical in determining whether or not a vehicle that leaves the lane edge has crossed onto an unpaved shoulder; an accuracy of ±0.328 ft (0.1 m) would be preferable. The lane position tracking software that is part of the instrumentation package will measure lane width as well and use this information to determine vehicle position within its lane. The lane tracking software is expected to be less accurate than the mobile mapping data collection method. It will

### Table 4.7. Final Rodeo Geometric Data Elements

| Feature | Data Elements | Frequency | Accuracy | Adequate for Lane Departure Research Questions |
|---|---|---|---|---|
| Grade | Direction and percent | | ±0.5% | Yes |
| | Location | | ±3 ft (0.914 m) | Yes |
| Cross slope | Location | | ±3 ft (0.914 m) | Yes |
| | Roadway cross slope | | ±0.01% | Yes |
| | Clear zone cross slope | | ±0.25% | Yes |
| | Clear zone width | | ±0.5 ft (0.152 m) | Yes |
| Curvature | Horizontal PC (point of curvature) and PT (point of tangency); vertical PC and PT | | ±3 ft (0.914 m) | Yes |
| | Horizontal curve, vertical curve, and transition curve length | | ±2 ft (0.61 m) | Yes |
| | Horizontal and vertical curve radius | | ±25 ft (7.62 m) | Yes |
| | Horizontal curve super elevation | | ±0.05% | Yes |
| | Vertical curve type, presence of transition curve | 100% | | Yes |
| | Stopping sight distance | | ±10 ft (3.05 m) | Yes |

### Table 4.8. Final Rodeo Intersection Data Elements

| Feature | Data Elements | Frequency | Accuracy | Adequate for Lane Departure Research Questions |
|---|---|---|---|---|
| Intersection configuration and dimensions | Type, number of approaches, number of through lanes, presence of channelization, number of left-turn lanes, number of right-turn lanes, presence of crosswalks, presence of illumination | 100% | | Yes |
| | Location | | ±3 ft (0.914 m) | Yes |
| | Skew | | ±0.5° | Yes |
| | Length of left- or right-turn lanes | | ±2 ft (0.61 m) | Yes |
| Traffic control | Type | 100% | | Yes |
| Signal | Type, pedestrian signal head present | 100% | | Yes |
| | Location | | ±3 ft (0.914 m) | Yes |
| Stop control | Type, presence of flashing beacon | | 100% | Yes |

*Table 4.9. Final Rodeo Pavement Condition Data Elements*

| Feature | Data Elements | Frequency | Accuracy | Adequate for Lane Departure Research Questions |
|---|---|---|---|---|
| Pavement edge | Amount of pavement edge drop-off | | ±0.5 in. (0.013 m) | Yes |
| | Location | | ±3 ft (0.914 m) | Yes |
| Pavement profile | Roughness measures | | ±10 in./mile | Yes |
| | Critical pavement failure | | 100% | Yes |
| Skid | Macrotexture | Reported at 0.1 mile intervals | | Yes |

*Table 4.10. Final Rodeo Roadway Data Elements*

| Feature | Data Elements | Frequency | Accuracy | Adequate for Lane Departure Research Questions |
|---|---|---|---|---|
| Bridges | Begin and end | | ±3 ft (0.914 m) | Yes |
| | Presence of approach slab or bridge rail | 100% | | Yes |
| | Offset | | ±2 ft (0.61 m) | Yes |
| Driveway | Location | | ±3 ft (0.914 m) | Yes |
| | Type | 100% | | Yes |
| Lanes | Number or special lane function type | 100% | | Yes |
| | Lane width | | ±0.5 ft (0.152 m) | No |
| | Location, lane add point, lane drop point | | ±3 ft (0.914 m) | Yes |
| Median | Type | 100% | | Yes |
| | Location | | ±3 ft (0.914 m) | Yes |
| | Width | | ±0.5 ft (0.152 m) | Yes |
| Rail crossings | Location | | ±3 ft (0.914 m) | Yes |
| | Number of tracks, control type, crossing number | 100% | | Yes |
| | Grade of approach or leave side | | ±0.5% | Yes |
| Ramps | Location | | ±3 ft (0.914 m) | Yes |
| | Type of terminal, type of section | | | Yes |
| Shoulder | Type | 100% | ±0.5 ft (0.152 m) | Yes |
| | Paved width, shoulder total width | 100% | ±3 ft (0.914 m) | Yes |
| | Location | | | Yes |

be important to compare actual lane width to what is collected by the lane position tracking system to serve as a check. Hence, accurate measurement of lane width is important.

It should be noted that the upcoming project SHRP 2 Safety Project S04A will make additional decisions about what data will be collected by the mobile mapping vehicles in Safety Project S04B. This will change the scope of what is being collected.

## Review of Planned In-Vehicle Instrumentation Package and Available Data Elements

The following sections summarize a review of the instrumentation package that is planned for the full-scale naturalistic in-field driving study. The sections provide a summary of information that was available as of January 2010.

It should be noted that the final specifications for the data acquisition system (DAS) are not yet available, and some differences will be present between what has been reviewed here and what is available with the final DAS.

The in-vehicle instrumentation package is expected to consist of the following sensors/elements (Dingus et al., 2008a, Task 6; Dingus et al., 2008b, Task 7):

- Two forward-looking cameras;
- Three rear-looking cameras;
- GPS;
- Incident button;
- Microphone;
- Speaker;
- Alcohol sensor;
- Light sensor;
- Bluetooth radio to communicate with the forward radar;
- Acceleration and orientation sensor;
- On-Board Diagnostics (OBD) II;
- Forward radar; and
- Machine vision capabilities, including lane position and edge sensing, eyes-forward monitor, and traffic signal state.

Each sensor or element of the DAS that will provide relevant information for answering lane departure research questions is discussed below. The following information is provided for each sensor/element:

- Description;
- List of data elements that are best collected from that sensor/element or are only available from that source, as well as potential data elements that could be collected when they cannot be obtained from other sources;
- Expected accuracy;
- Expected resolution of data collection; and
- Limitations in obtaining the type, amount, or quality of data necessary.

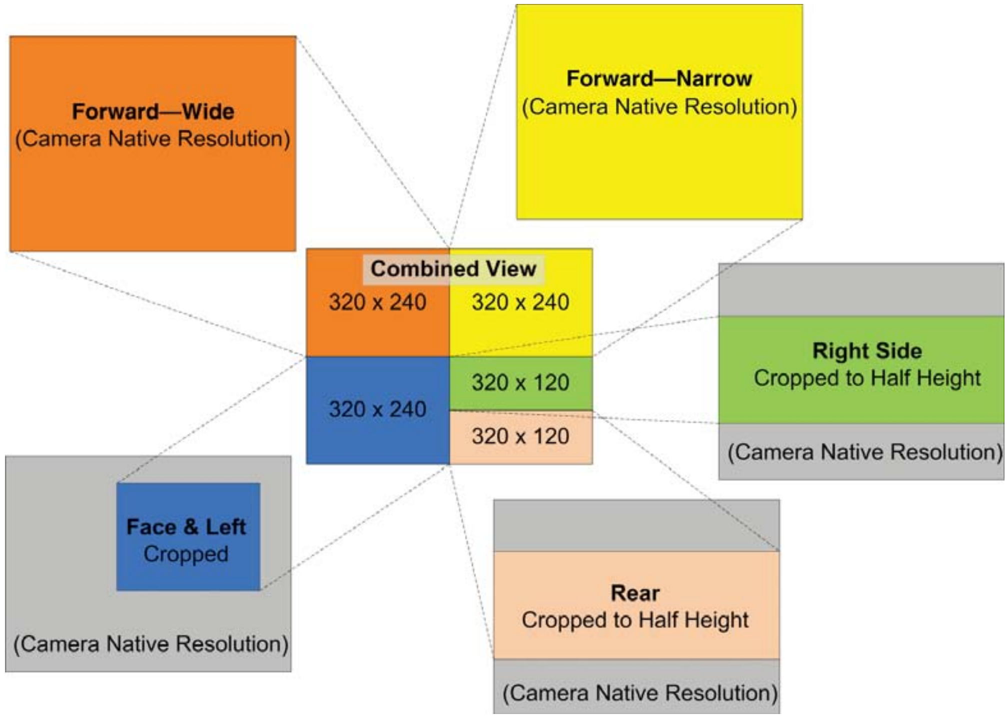
## Sensor/Elements Related to Roadway, Environmental, and Vehicle Factors

### *Sensor/Element: GPS (Dingus et al., 2008b, Task 7, Appendices A and B)*

- **Necessary data elements from sensor/element:** Position (latitude, longitude, altitude), heading.
- **Data from sensor/element as secondary source:** Forward and side speed and acceleration (best from On-Board Diagnostics [OBD]).
- **Accuracy:** Not stated.
- **Resolution:** 10 Hz. (It has recently been brought to the research team's attention that GPS data may be collected at 1 Hz.)
- **Comments and limitations:** Accuracy was not stated but is highly relevant in determining vehicle position. Accuracy is necessary to link the vehicle to the appropriate roadway and extract corresponding roadway data elements. It will also be necessary to determine vehicle position when a lane departure event occurs. The best source of distance data between the vehicle and other objects will be radar, but the vehicle position from the GPS will also be necessary.

### *Sensor/Element: Forward Video (Dingus et al., 2008a, Task 6)*

- **Description:** Two forward videos are expected for the full-scale study, which will provide a forward view of the roadway from the perspective of the driver's field of view. One video will show a wide forward view of approximately 60° (primary forward view), and the other will be a narrow view of approximately 25° (secondary forward view). The secondary forward view will be zoomed/tilted to provide support for assessing traffic signal state. Both videos will be taken with color cameras.
- **Necessary data elements from sensor/element:** The following data elements are those where the forward camera is the best or only available source of information:
  - Sequence of events;
  - Ambient conditions (e.g., clear, raining, snowing);
  - Road surface condition (e.g., dry, wet, snow covered, surface irregularities);
  - Oncoming traffic density;
  - Identification of surrounding objects when vehicle engages in lane departure;
  - Identification of pedestrians;
  - Visibility;
  - Lane position tracking;
  - Verification of lane departure;
  - Curve direction;
  - Status of lane markings (e.g., highly visible, obscure); and
  - Signal state.




Source: Dingus et al., 2008a, Task 6.

***Figure 4.11. Combined video views from VTTI.***

It may be necessary to determine sign location (signs may be placed or removed after a roadway is scanned using the mobile mapping system); it will also be possible to provide a qualitative measure of sign condition and retroreflectivity.

- **Data from sensor/element as secondary source:** While it may not be practical to obtain roadway data from the forward video, it can serve as both a check and a source when the data cannot be obtained from other sources. Data elements that can be obtained include the following:
  – Indication of whether the subject vehicle is following another vehicle and how closely (subjective measure only);
  – Roadway data element type (e.g., shoulder type, presence and type of guardrail, type of signing, driveways, type of intersection control, presence and type of rumble strips);
  – Although not ideal, measurement of certain roadway elements is possible when the forward image is calibrated (e.g., shoulder width, distance to nearby fixed objects); and
  – Qualitative measure of ambient lighting.
- **Accuracy:** It was stated that the forward view video resolution would be at least 320 × 240 pixels.
- **Resolution:** 10 Hz. Video will be stored in quad-format storage at one-fourth of the resolution, as shown in Figure 4.11. An example of actual images in the stored format is shown in Figure 4.12.
- **Comments and limitations:** Use of color cameras for the forward views will provide enhanced ability to distinguish roadway features. Color video will be particularly useful in identifying roadway surface condition and critical for determining traffic signal state. It will also be useful in determining condition of lane lines and signs.


***Figure 4.12. Example of video view.***

The wide-view forward video appears to have a fish-eye view that distorts objects and perspective. Although distance and identification of objects may come from other data sources, the fish-eye view could affect the ability to detect nearness of objects, identify objects, and identify sequence of events. The distortion appears worst at the edges of the image. This distortion could affect a data reductionist's ability to determine run-off-road (ROR) activities.

It will also be important to know if the original imagery data will be retained and will be available. It was stated that the forward video would be collected with a minimum resolution of $320 \times 240$ pixels. The images will be cropped and compressed into a view as shown in Figure 4.12. If the forward views are compressed and/or cropped, it may be advantageous for some applications to have the ability to view the forward image in its original form.

### Sensor/Element: Rear-View Video (Dingus et al., 2008a, Task 6; Dingus et al., 2008b, Task 7, Appendices A and B)

- **Description:** The rear view provides an image of the roadway from the rear window of the vehicle, generally reflecting the driver's rear-view perspective. The view should include the rear driving environment and may provide some ability to identify presence of back seat passengers.
- **Necessary data elements from sensor/element:** Presence of other vehicles, ambient conditions if they cannot be identified with the forward video.
- **Data from sensor/element as secondary source:** NA.
- **Accuracy:** Minimum resolution is $320 \times 120$ pixels.
- **Resolution:** 10 Hz.
- **Comments and limitations:** Given the image resolution provided in Figure 4.12, it may be difficult to distinguish vehicles that are not closely following the subject vehicle. This information may not be relevant for lane departures, however. The data collection resolution of 10 Hz is adequate.

### Sensor/Element: External Right-Side View (Dingus et al., 2008a, Task 6)

- **Description:** The right-side view provides an image of the roadway from the right of the driver, reflecting the driver's right-side view.
- **Necessary data elements from sensor/element:** Presence of vehicles to the subject vehicle's right rear; presence of objects to right rear.
- **Data from sensor/element as secondary source:** NA.
- **Accuracy:** Minimum resolution is $320 \times 120$ pixels.
- **Resolution:** 10 Hz.
- **Comments and limitations:** The data collection resolution of 10 Hz is adequate. The image should be able to provide information on the presence of another vehicle during a right-lane departure. However, it may be difficult to distinguish relative position given the resolution indicated in Figure 4.12.

### Sensor/Element: Instrument Panel View (Dingus et al., 2008a, Task 6)

- **Description:** Provides an over-the-shoulder view of the steering column and instrument panel.
- **Necessary data elements from sensor/element:** Relevant for driver factors.
- **Data from sensor/element as secondary source:** May provide indication of change in steering angle.
- **Accuracy:** Minimum resolution is $320 \times 240$ pixels.
- **Resolution:** 10 Hz.
- **Comments and limitations:** Should be adequate.

### Sensor/Element: Machine Vision Lane Tracking (Dingus et al., 2008a, Task 6; Dingus et al., 2008b, Task 7, Appendices A and B)

- **Description:** VTTI Road Scout (VRS) uses machine vision to determine the presence and type of lane lines and will calculate a vehicle's position within its traffic lane. The system has the capability to determine when a vehicle is in the lane, crosses a solid marking, crosses a dashed line during lane changes, or when a lane change is aborted. The system is self-calibrating. It was tested on a four-lane divided road with a grass median, a four-lane divided road with left curb (no lane line), a two-lane road with pavement markings, a two-lane road with no pavement markings, and a two-lane gravel road. It is expected to work in the following environments:
  - Interstates;
  - Lined highways;
  - Tangent and curved sections;
  - Under nighttime driving with or without overhead illumination;
  - In inclement weather when lane lines are visible;
  - On blacktop and concrete; and
  - Situations where only the centerline is present or visible.

  The system could not determine vehicle position on gravel or on two-lane, unmarked suburban segments. It was not clear if the system had been tested on two-lane rural segments. It can detect an upcoming curve and detect differences in lane marking type (none, double line [solid or dashed], single line [solid or dashed], road gutter, road edge, and raised pavement markings).

System output includes the following:

– Lane offset, which is the position of the center line of the vehicle with respect to the centerline of the roadway;
– Lane width, determined by estimating the width of the marked lane in inches;
– Line distance, which is the distance from the center of the vehicle to the right- or left-lane line; and
– Probability percentages, which is a measure of the likelihood that the pavement markings exist (serves as a key indicator for the overall reliability of the system).

The system is also expected to be able to estimate horizontal curve radius. However, the accuracy is unknown.

- **Necessary data elements from sensor/element:** Lane position, lane location, lane changes, type of lane lines, lane width, offset from center of lane, distance of wheel from right- or left-lane boundary, angle of departure (calculated), lane departure, road departure.
- **Data from sensor/element as secondary source:** Radius.
- **Accuracy:** (Units reported in meters.) The accuracy of lane-position offset was stated as ±0.656 ft (0.2 m) 95% of the time when lane tracking confidence is high. The system will store distance from right tire to right-lane boundary, distance of left tire to left-lane boundary, and lane width with the same accuracy. The accuracy of the horizontal curvature is expected to be ±10% of actual roadway radius when lane-tracking confidence is high.
- **Resolution:** Minimum 10 Hz.
- **Comments and limitations:** The expected accuracy of lane-position tracking for the proposed full-scale study is expected to be ±0.656 ft (0.2 m), which is lower than the accuracy for the UMTRI data. The average tire width is around 6.5 to 9 in. (0.165 to 0.229 m). Hence, the error is within a normal tire width. While an accuracy of 0.328 ft (0.1 m) is preferable, the proposed accuracy is expected to be sufficient. Additional collection of vehicle position at 10 Hz is adequate to establish the angle of departure and offset.

  The radius calculated by the UMTRI lane-tracking system appeared to be inaccurate. It will be important to verify the accuracy of the VRS system in calculating curve radius.

### *Sensor/Element: Forward Sensor/Radar (Dingus et al., 2008b, Task 7, Appendices A and B)*

- **Description:** The system will have a forward radar capable of tracking and storing information on the five objects closest to the vehicle. Objects can be identified and tracked for up to 200 m in front of the vehicle within ±0.324 radians (18°) within the horizontal field of view centered on the test vehicle heading. The system can identify vehicle type (car, motorcycle, truck, pedestrian/bicyclist) and will indicate what it detects to be the lead vehicle, defined as the closest vehicle occupying the same lane. It was not clear if the system can identify roadside objects other than that they are present.
- **Necessary data elements from sensor/element:** Distance to and location of the nearest strikable object (including other vehicle); vehicle spacing as indicator of aggressive driving.
- **Data from sensor/element as secondary source:** The type of object will likely be determined from forward video but can be confirmed with radar.
- **Accuracy:** Distance (in meters), accuracy will be ±1.64 ft (0.5 m); vehicle target range (in meters per second), accuracy is ±1.64 ft/s (0.5 m/s); relationship to target object (stored in a Polar or Cartesian coordinate system), accuracy is ±0.052 radians (3°) for Polar; lateral and longitudinal offset (in meters), accuracy ±3.28 ft (1.0 m).
- **Resolution:** Will store at a minimum of 40 Hz for each track.
- **Comments and limitations:** The stated accuracy of ±1.64 ft (0.5 m) is adequate to measure time and distance to collision.

### *Sensor/Element: Automated Collision Identification and Notification (Dingus et al., 2008b, Task 7, Appendices A and B)*

- **Description:** System that will continuously monitor vehicle sensors to determine when a potential collision has occurred. The parameters include (1) a longitudinal acceleration of at least 3.5 *g* for at least 500 m, (2) lateral acceleration of at least 3.5 *g* for at least 500 m, or (3) air bag deployment status.
- **Necessary data elements from sensor/element:** Indication of collision.
- **Data from sensor/element as secondary source:** NA.
- **Accuracy:** NA.
- **Resolution:** NA.
- **Comments and limitations:** NA.

### *Sensor/Element: Light Sensor (Dingus et al., 2008b, Task 7, Appendices A and B)*

- **Description:** Senses amount of light.
- **Necessary data elements from sensor/element:** Amount of daytime lighting, presence and amount of nighttime lighting.
- **Data from sensor/element as secondary source:** NA.
- **Accuracy:** Illumination (in lux from 3 to 80,000), accuracy is ±3%.
- **Resolution:** 10 Hz.
- **Comments and limitations:** Appears to be adequate for project objectives.

### *Sensor/Element: Sensor to Record Internal Ambient Temperature (Dingus et al., 2008b, Task 7, Appendices A and B)*

- **Description:** Record of vehicle cabin temperature.
- **Necessary data elements from sensor/element:** Outside temperature would be useful in making some estimates about roadway surface condition.
- **Data from sensor/element as secondary source:** NA.
- **Accuracy:** ±1°C.
- **Resolution:** Sampled as changes occur of at the rate of once per 5 min.
- **Comments and limitations:** It is not clear why internal ambient temperature is being recorded and not outside ambient temperature.

### *Sensor/Element: Acceleration and Orientation Sensor (Dingus et al., 2008b, Task 7, Appendices A and B)*

- **Description:** Records lateral acceleration, longitudinal acceleration, vertical acceleration, and yaw rate.
- **Necessary data elements from sensor/element:** Lateral and forward acceleration, pitch, pitch rate, yaw, yaw rate.
- **Data from sensor/element as secondary source:** NA.
- **Accuracy:** Lateral acceleration (±0.01 $m/s^2$), longitudinal acceleration (±0.01 $m/s^2$), vertical acceleration (±0.01 $m/s^2$), and yaw rate (radians/second).
- **Resolution:** Stored at 10 Hz during normal operation and higher during high rate of acceleration.
- **Comments and limitations:** The stated accuracy appears adequate to answer the research questions.

### *Sensor/Element: OBD (Dingus et al., 2008b, Task 7, Appendices A and B)*

- **Description:** OBD monitors parts of the chassis, body, and accessory devices and the diagnostic control network of the car.
- **Necessary data elements from sensor/element:** The following data elements will be provided and are relevant for answering the research questions:
  - Forward speed (m/s);
  - Accelerator pedal position (percent);
  - Brake state (on/off);
  - Steering wheel position (radians of rotation);
  - Brake pedal force ($lb/in^2$);
  - Horn status (on/off);
  - Gear;
  - Headlight status (on/off, parking);
  - High beam (on/off);
  - Wiper status (intermittent, slow, high, manually activate);
  - Cruise control (on/off);
  - Seat belt status (on/off);
  - Front-seat passenger status (present/not present);
  - Wheel speed (m/s);
  - Automatic braking system (ABS) activation;
  - Air bag deployment;
  - Electronic stability control (indication of when active, if present);
  - Traction control (indication of when active, if present);
  - Lane departure warning system (indication of when active, if present);
  - Forward collision warning system (indication of when active, if present);
  - Distance driven during trip (in kilometers);
  - Turn signal status (off, left, right, hazard); and
  - Driver-initiated event (indication of when driver presses event button).
- **Data from sensor/element as secondary source:** NA.
- **Accuracy:** Accuracy was not stated for any of the relevant items. The accuracy of the OBD system is assumed to be sufficient.
- **Resolution:** 10 Hz.
- **Comments and limitations:** Accuracy and resolution are expected to be sufficient. No additional necessary items from the OBD were determined.

## Sensor/Elements Related to Driver Factors

### *Sensor/Element: Driver-Face Video (Dingus et al., 2008a, Task 6)*

- **Description:** The face video provides an image of the driver's face, which can indicate driver distraction and eye location.
- **Necessary data elements from sensor/element:** Presence of driver impairments (distraction, fatigue, emotion) over time, especially before and at the time of a lane departure. A driver's glance direction may indicate driving-related behaviors, preincident awareness of conflict, and postincident behavior.
- **Data from sensor/element as secondary source:** Driver identification.
- **Accuracy:** Minimum resolution is 640 × 640 pixels.
- **Resolution:** 15 Hz as the minimum and 30 Hz as preferred.
- **Comments and limitations:** The data collection resolution of 15 to 30 Hz should be adequate. The image should be able to provide information on driver fatigue, emotion, and any secondary tasks other than driving, such as talking on a cell phone or with passengers and reaching for objects in the vehicle. However, some secondary tasks, such as text messaging, might not be captured clearly, since the image might not be large enough. This kind of information needs to be provided by other videos.

### *Sensor/Element: Instrument Panel View (Dingus et al., 2008a, Task 6)*

- **Description:** Provides an over-the-shoulder view of the steering column and instrument panel.
- **Necessary data elements from sensor/element:** Hands on steering wheel; secondary tasks in which the driver engages.
- **Data from sensor/element as secondary source:** Steering behavior at the time of incident.
- **Accuracy:** Minimum resolution is 320 × 240 pixels.
- **Resolution:** 15 Hz as the minimum and 30 Hz as preferred.
- **Comments and limitations:** Should be adequate.

*Note:* The instrument panel view has been mentioned before, but not in detail.

### *Sensor/Element: External Passenger-Side (Right-Side) View (Dingus et al., 2008a, Task 6)*

- **Description:** This view offers information regarding the traffic around the subject vehicle and can also provide information about the passenger(s) in the front and rear seats.
- **Necessary data elements from sensor/element:** Presence of vehicles to subject vehicle's right rear, presence of objects to right rear.
- **Data from sensor/element as secondary source:** Can provide some information about driver distraction related to passengers.
- **Accuracy:** Minimum resolution is 320 × 120 pixels.
- **Resolution:** 10 Hz.
- **Comments and limitations:** Should be adequate.

*Note:* This view has been mentioned before, but not in respect to the driver.

### *Sensor/Element: Passive Alcohol Sensor (Dingus et al., 2008a, Task 6)*

- **Description:** Provides information on the presence of alcohol.
- **Necessary data elements from sensor/element:** Alcohol concentration level.
- **Data from sensor/element as secondary source:** NA.
- **Accuracy:** ±0.01 at 0.1 level.
- **Resolution:** Once per hour.
- **Comments and limitations:** The sensor is only able to detect presence of alcohol in the vehicle. The sensor may be affected by air circulation within the vehicle. Judgments about whether the driver has been drinking will have to be made using other information (e.g., the presence of alcohol and the fact that the driver is the only occupant may suggest that the driver has been drinking). The system cannot indicate blood alcohol level.

## Summary

The following summarizes information about review of data elements necessary and review of what is expected to be available in the full-scale study.

### General Comments about Extracting Data from Existing Data Sources

In general, most of the roadway, environmental, and vehicle data elements desired could be extracted from the UMTRI naturalistic driving study data and related aerial imagery, crash databases, and roadway databases. The naturalistic driving data indicated when the vehicle was traveling on a curve. When vehicle traces were overlaid with aerial images and compared, the identified curve locations were quite accurate. Two data items that were not accurate were shoulder width and curve radius. This is based on a review of the UMTRI data, as described in Appendix A.

The forward imagery was adequate for all the applications for which it was used. One advantage of the UMTRI imagery over the proposed forward imagery for the full-scale data collection is the width of the forward view that was available. It was easy to see a large portion of the forward roadway (including all of the shoulders), identify objects to the edge of the roadway, identify ambient conditions, and confirm that a vehicle was departing its lane from the forward view. While the forward view with the proposed full-scale in-vehicle instrumentation is in color, which offers additional advantages, the forward view does not offer the same wide view. As a result, it may be difficult to identify roadside features. The distortion of the image (in fish-eye view) in the proposed instrumentation package is also particularly problematic. The lateral portion of the proposed imagery does, however, offer a better view of overhead features. The two forward views are compared in Figure 4.13.

The lane position tracking was very useful in the UMTRI data. It was relatively simple to determine when a vehicle had left the roadway. It was also relatively easy to tell when the lane position data were "bad." The ability to determine a vehicle's position within its lane is critical for identifying lane departure events and answering the lane departure–related research questions.

Sufficient data were not provided with the VTTI data set to make the same determinations as for the UMTRI data. In general, the image resolution was too low that it was diffi-

(a)

(b)

Source: (a) UMTRI RDCW data set; (b) Dingus et al., 2008a, Task 6.

***Figure 4.13. Comparison of forward imagery: (a) forward view from UMTRI data set; and (b) potential forward view for full-scale data collection.***

cult to make out many features in the forward video. It was even difficult to tell from the forward view that the vehicle was leaving the roadway or which object was in the vehicle's path. It is not known if the image data were reduced from their original format. The database provided with the vehicle trace data was adequate to extract information such as lateral acceleration.

## Summary Comments and Concerns about Proposed Full-Scale Data Collection Methods

The following is a summary of comments or concerns that arose during the review of the instrumentation packages that will be available for the full-scale data collection effort (SHRP 2 Safety Projects S03 and S05).

- **Mobile mapping vans (SHRP 2 Safety Project S03)**
  - The accuracy and resolution of data collection for all data elements (except for lane width) met or exceeded the level of accuracy that was determined to be necessary to answer lane departure research questions.
  - The proposed accuracy of lane width to be collected with mobile mapping vans is ±0.5 ft (0.152 m). Lane width will be calculated with the vehicle instrumentation package lane position tracking system in order to determine vehicle position. The lane tracking system is expected to be less accurate than the mobile mapping vans, so it will be important to have an accurate measurement from the mobile mapping system as a check. A more accurate measurement of lane width would be recommended, if possible. An accuracy of ±0.25 ft (0.076 m) would be preferable.
  - Recent information from various SHRP 2 safety meetings during the summer of 2009 have indicated that the final mobile mapping data collection may be somewhat different from what was reviewed in this document.
- **In-vehicle instrumentation (SHRP 2 Safety Project S05)**
  - Accuracy of the differential GPS was not stated but is highly relevant for determining vehicle position relative to roadway features.
  - The use of color imagery for the forward video is a welcome addition and will allow objects to be distinguished more easily. Use of color cameras for the forward views will provide an enhanced ability to distinguish roadway features. It will be particularly useful for identifying roadway surface conditions and critical for determining traffic signal state. It will also be useful for determining the condition of lane lines, and signs and will allow identification of traffic signal state. While not as relevant to lane departures, this ability will be critical for answering intersection questions.
  - The image resolution of $320 \times 240$ pixels for the forward view appeared adequate.
  - The fish-eye view for the forward view distorts objects and perspective. This is highly problematic for identifying objects and gauging distances. Although distance and identification of objects may come from other data sources, the fish-eye view could affect the ability to detect the nearness of objects, identify objects, and identify the sequence of events. The distortion appears worst at the edges of the image. This distortion could affect the ability to measure distances to roadside objects. The distortion could also affect the ability to identify objects.
  - It was indicated that all raw data (both video and sensors) would be continuously recorded and preserved (Dingus et al., 2008b, Task 7). It may be useful to view the raw forward, back, or side video because some information may be obtained that cannot be obtained with the compressed-resolution images. It is important to clarify whether the data will be stored in a format that is linked to the other data and whether the data can be accessed.
  - The lane position tracking system is critical for addressing lane departure questions. Several experts were questioned about the level of risk of different lane departure events. They unanimously agreed that even one tire leaving the paved roadway surface onto a grass, gravel, or mixed-surface shoulder constitutes a highly dangerous situation. As a result, the lane position tracking system should be accurate and reliable

enough to determine when one or more tires have departed the roadway surface. The planned lane position tracking system for the full-scale study has a stated accuracy of ±0.656 ft (0.2 m). While this is still within the range of a normal tire width and is likely to be adequate, any improvements to the planned accuracy would be beneficial. A lower level of accuracy may be acceptable where the surface beyond the lane marker is hard and level.

– It is important that the lane tracking system be verified both in terms of accuracy and in situations where it may not perform well.
– It was stated that a combination of in-vehicle sensors could be used to determine curve radius. It is important that the accuracy be verified. The radius measurements received with the UMTRI data did not appear to be accurate.
– It was not stated whether the forward radar has the ability to identify roadside objects.
– It is unclear why internal vehicle cabin temperature is measured and recorded but not the outside ambient temperature.
– The method to determine the accuracy of each stated sensor/data collection element should be stated. For example, if radius is determined using a combination of in-vehicle sensors, the test method and results to determine accuracy should be stated.
– Other discussions about the in-vehicle data collection system have suggested that a head-pose tracker may be available.

CHAPTER 5

# Defining and Evaluating Lane Departure Crash Surrogate Thresholds Using Naturalistic Driving Study Data

Lane departure crashes are the best measure of safety. Naturalistic driving studies, however, even the fully deployed SHRP 2 field driving study, will have limited cases of lane departure crashes. The naturalistic driving studies will capture crashes, near crashes, and incidents, as well as normal driving. The frequency of incidents and near-crash events is typically greater than the frequency of crashes; incidents and near-crash events may be used as crash surrogates.

Using surrogates will also provide an opportunity to study what happens preceding and following an incident or event. The most significant advantage of naturalistic driving studies is that they provide a firsthand record of the events that precede crashes and incidents. Roadway, environmental, vehicle, and human factors can be extracted directly rather than from secondhand information from police records and crash databases to identify relationships among factors that influence lane departure crash risk. This firsthand information can also be used to determine the factors that lead to a positive outcome. For instance, if a similar number of lane departures occur on roadway sections with and without paved shoulders and the paved shoulders have a higher proportion of safe outcomes (vehicles can return safely to the road), the incidents can be used to evaluate the effectiveness of paved shoulders.

This chapter discusses potential lane departure surrogates that can be obtained from naturalistic driving study data. Several data sets were used to evaluate thresholds for lane departure crash surrogates. The data sets are described fully in Chapter 3. For simplicity, naturalistic driving study data from UMTRI's road departure crash warning (RDCW) field operation test (FOT) is referred to in this chapter as "the UMTRI data set," and the naturalistic driving study data from VTTI's 100-car study is referred to as the "VTTI data set."

The naturalistic driving study data from UMTRI and VTTI were used to evaluate which variables may be the most useful in setting triggers to identify lane departure events and to assess what thresholds may be used. Data were reduced as described in Appendices A and B.

The UMTRI data resulted in a number of encroachments but no conflicts or crashes. Only data for rural, paved, two-lane roadways were included. The VTTI data provided near crashes and crashes but no encroachments. Additionally, variables were not consistent between the two data sets. As a result, the two data sets were evaluated separately, as discussed in the following sections.

## Introduction

Frequency and severity of crash data are commonly used to assess whether driver, road, traffic, or environmental factors influence safety and to evaluate whether a countermeasure is effective. However, crash-based safety analyses are plagued by several problems (Songchitruksa and Tarko, 2006). Crashes are rare, and events surrounding a crash are oftentimes random. As a result, safety analyses often depend on small sample sizes. Additionally, crash reporting can be inconsistent, which makes comparisons across sites difficult. Another problem is the timeliness of crash data. Once a countermeasure is implemented, agencies like to evaluate the immediate effectiveness to assess whether more resources should be invested. However, before-and-after crash studies often cannot be completed until several years after treatment installation because a representative sample is not available immediately to assess significant differences with sufficient power.

Some researchers have addressed limitations in crash data by using surrogates to measure crash risk. Surrogates may take two forms: safety surrogates and crash surrogates. The difference between the two is related to whether their underlying relationship to safety has been established. The types of surrogates described in the following two paragraphs are examples of safety surrogates. In most cases, the underlying relationship between crashes and the safety surrogate is assumed. Selected safety surrogates are believed to have some relationship to safety, although a demonstrated relationship rarely exists. Additionally, when safety surrogates are used in studies, there is no attempt to define the relationship between the safety surrogate and crashes.

FHWA (2009) suggests that a reduction in violations is a viable safety surrogate to evaluate the effectiveness of red-light running countermeasures (e.g., camera enforcement). Other common surrogates are traffic conflicts, traffic violations, road user behavior, and speed (Forbes et al., 2003). Change in speed is frequently used to assess the effectiveness of treatments such as traffic calming. It is assumed that if speeds are reduced, crashes will also be reduced. Lane deviation has also been used as a safety surrogate measure for assessing the likelihood of run-off-road (ROR) crashes (LeBlanc et al., 2006) and the likelihood of crashes resulting from distraction (Donmez et al., 2006).

Retting et al. (2007) and Bonneson et al. (2004) used reduction in red-light running violations as a safety surrogate to assess the effectiveness of red-light-running cameras. Garber et al. (2005) evaluated reductions in red-light running citations to evaluate the effectiveness of red-light-running cameras. The effectiveness of rumble strips has been evaluated on the basis of lateral placement and speed (Porter et al., 2004), drivers' lane position with respect to a forced rumble strip encounter (Noyce and Elango, 2004), and vehicle's lateral position and change in vehicle separation (Pratt et al., 2006). Taylor et al. (2005) observed vehicle placement relative to the edge line using single versus double paint lines to delineate presence of shoulder rumble strips.

The second type of surrogate is the crash surrogate. This type of surrogate is expected to have some statistically measurable relationship to crashes. Ideally, the relationship between crashes and the surrogate measure is evaluated or known. If so, the crash surrogate can be used as a measure of effectiveness, and the reduction in crashes can be predicted. Shankar et al. (2008) defines a crash surrogate as a marker that is correlated with a crash, usually based on time, so that as time increases the crash likelihood also increases. The authors also define a crash surrogate as a measure that is as responsive to the same interventions as the related crashes. For instance, edgeline rumble strips are expected to decrease the number of right-side lane departure incidents, as well as the cases of ROR crashes.

The main difference between safety surrogates and crash surrogates is that safety surrogates are used by researchers and agencies as a stand-in variable for crash data. They are widely used, but there is no proven relationship between crashes and the variable used. It is assumed that if the safety surrogate changes (e.g., reduction in speeds), crash severity or frequency will improve. Additionally, studies in which safety surrogates are used do not attempt to derive a relationship.

## Background on Crash Surrogates

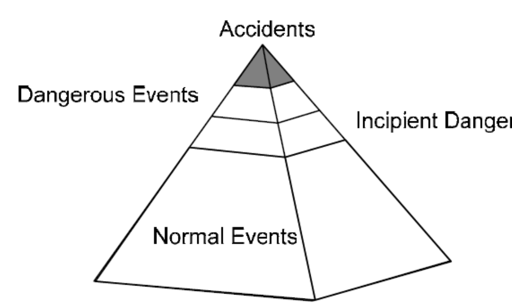
This section discusses the crash surrogates that have been used in other studies for different types of crashes.

Songchitruksa and Tarko (2006) used extreme value theory to model crash risk and frequency for right-angle crashes at intersections. Degree of separation was used as the surrogate variable, and the authors ordered traffic events from safest to most dangerous to assess risk. They defined the boundary between crash and noncrash events using the concept of crash proximity. They evaluated their methodology and concluded that there was a promising relationship between safety estimates and historical crash data.

Archer (2001) used the traffic conflict technique, which registers the occurrence of near accidents in real-time traffic to represent accident frequency and outcome. The surrogate measure proposed by Archer (2001) is defined as "time to accident." Gettman and Head (2003) derived surrogate crash measures from simulation models. The authors indicated that the best surrogate measures for crash risk were time to collision, post encroachment time, deceleration rate, maximum speed, and speed differential. Mounce (1981) evaluated the correlation between stop sign violations and crash rates. Salman and Al-Maita (1995) evaluated traffic conflicts and crashes at 18 T-intersections in Jordan and developed a statistically significant linear regression model that related annual number of crashes to mean hourly conflicts. Chin et al. (1992) used a time-to-collision method to model freeway merging conflicts.

Other researchers have treated crash surrogates as a continuum, with regular traffic incidents at one end and crashes at the other. Thresholds are used to partition incidents. Chin and Quek (1997) describe this as a probability distribution of incidents. Songchitruksa and Tarko (2006) use the notion of ordering traffic incidents from the safest to the most dangerous (Figure 5.1).

Different researchers have applied different crash surrogates to define the boundary between incidents. A common




Source: Songchitruksa and Tarko, 2004.

***Figure 5.1. Continuum of traffic incidents.***

measure that has been used is time to collision (TTC), where at TTC = 0 the subject vehicle and another vehicle/object collide, resulting in a crash. Songchitruksa and Tarko (2006) use the concept of degree of separation. When there is considerable separation between vehicles on the same path that are passing a conflict point, the passage is considered safe. As separation between vehicles decreases, risk increases until the two vehicles collide. This continuum of separation of time is called post-encroachment time (PET). Amount of separation can be partitioned into different levels of crash incidents that correspond to different levels of risk. In Songchitruksa and Tarko (2006), the threshold between the crash and crash-free boundaries is called crash proximity measure. Others refer to this threshold as a crash prevention boundary. Burgett and Gunderson (2001) define a crash prevention boundary as an analytically derived deterministic expression that separates driver performance into successful crash avoidance and unsuccessful crash avoidance. For any given set of conditions, there is a subset of driver brake response and level of deceleration that will result in crash avoidance and a subset of values that will result in a crash.

Hayward (1972) suggests the use of TTC when modeling situations where both vehicles continue in the same path without changing their speed. TTC is a good crash surrogate for two-vehicle crashes because the distance or the time separating two vehicles can be clearly identified.

Crash prevention boundaries have also been used as a crash surrogate. Burgett and Miller (2001) used velocity, separation distance, deceleration, and braking to develop crash prevention boundaries for rear-end crashes. Szabo and Wilson (2004) used the amount of acceleration necessary to avoid a crash as a function of the timing or location of a warning to define a crash prevention boundary.

Other crash surrogates used include proportion of remaining stopping distance (Allen et al., 1978), and deceleration rate (Songchitruksa and Tarko, 2004). Finally, Campbell et al. (2003) suggested that it is preferable to use physical measures of vehicle kinematic motions as the crash margin measure because collisions can be explicitly identified.

## Summary of Crash Surrogates Used for Lane Departures

Several measures have been used as crash surrogates for lane departures, including lane keeping, TTC, and crash prevention boundaries. However, little information was available that describes developing statistical relationships between lane departure crashes and lane departure crash surrogates. As a result, it is assumed that these types of relationships will need to be derived from research projects developed as part of the SHRP 2 Safety Project S08, Analysis of the SHRP 2 Naturalistic Driving Study Data. Crash surrogates that have been used for lane departure crashes are discussed in the following sections.

### Lateral Drift or Lane Keeping

Lateral drift or lane keeping is one measure that has been used by several researchers to evaluate lane departures. Several studies conducted at UMTRI, VTTI, and the University of Iowa (UI) have provided insights based on measures of lateral drift. UMTRI uses lane keeping to identify when a vehicle leaves the roadway. UMTRI researchers also define lane offset as the distance between the centerline of the vehicle and the centerline of the lane. Lane position and relative motion within the lane are determined by analyzing the forward-looking monochrome camera data. On tangent sections, the UMTRI lane departure warning system (LDWS, part of the RDCW) shows a "lower cautionary alert" when the vehicle is close to a dashed lane line, which indicates potential movement into an adjacent travel lane with no other evidence of an imminent risk of sideswipe collision. Additionally, UMTRI uses the term "lane intrusion," which suggests that a lane departure is imminent or likely. On curves, the UMTRI definition of a likely or imminent lane departure for the curve speed system (also part of the RDCW) was based on an estimate of most likely path, given vehicle speed, driver braking, turn signal use, assumptions about the unaware driver's response time, likely deceleration rate, and a threshold lateral acceleration of 0.25 *g*, which assumes no superelevation exists on the curve (LeBlanc et al., 2006).

Oxley et al. (2004) summarized work by Steyer et al. (2000) and noted that one of the important safety-related features in curve negotiation is vehicle lateral placement. Steyer et al. (2000) argued that the driving path should be considered when investigating crashes on curves. The authors make the distinction between right and left curves. Drivers who make left-side encroachments (across the centerline) may be doing so intentionally as they "cut the corner" or "straighten on the curve" if they cannot detect any opposing traffic. Steyer et al. (2000) also indicated that there was lateral placement related to curve radius, curve length, grade, and available sight distance.

### Time to Lane Departure or Time to Collision

Several crash surrogates use time to some critical event as the measure of crash risk. Time to lane departure (TTLD) reflects the time remaining before a vehicle crosses the lane line if the vehicle maintains its current trajectory. Several researchers have used TTLD or TTC as a crash surrogate measure. Szabo and Wilson (2004) used TTC in their lane departure warning system to determine the point where a vehicle is about to leave the roadway and the driver should be alerted of the danger. Pomerleau et al. (1999) used TTLD to determine when a lane departure warning system should provide a driver alert.

TTLD was defined as the time until an outer tire edge crosses the lane line. Mammar et al. (2006) developed a method to calculate time to line crossing (TTLC), considering both straight and curved vehicle paths. The authors estimated that a lane departure because of driver drowsiness leads to a slower rate of TTLC than in situations such as loss of vehicle control. They also indicated that real-time computation of TTLC is difficult because of limitations in determining vehicle kinematic variables, vehicle trajectory prediction, and lane geometry. As a result, approximate formulas such as the ratio of lateral distance to lateral speed have been used. The authors did develop a linear dynamic mode to predict future vehicle position on the basis of lateral displacement, vehicle position, steering angle, relative yaw angle, and roadway geometry.

## Distance Intruded

Distance intruded measures the distance a vehicle crosses into an adjacent lane or shoulder. VTTI uses distance intruded to determine when a lane departure occurs. Its lane tracking system sets a trigger to define a "lane bust" or "lane abort" incident. A lane bust occurs when a vehicle crosses a solid lane line. An incident is triggered when the vehicle moves a minimum of 3 ft outside a lane boundary without completing a lane change while traveling at a speed of 45 mph or higher (Dingus et al., 2006).

## Crash Prevention Boundaries

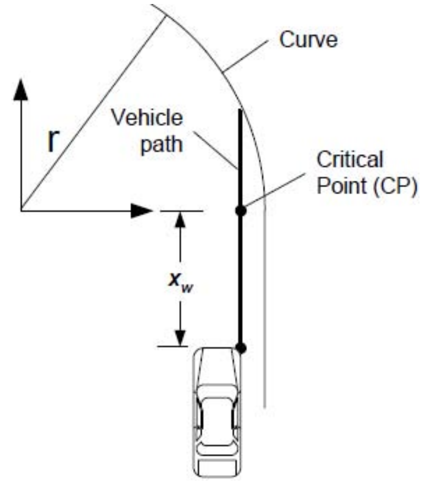
Szabo and Wilson (2004) used the concept of crash prevention boundaries to assess the effectiveness of RDCWs. Two metrics were used, one for curves and one for tangent sections. For curve negotiations, there is a critical point at which drivers can receive a road departure warning and respond appropriately by decelerating as needed, as shown in Figure 5.2. If the alert is provided after this point, the driver will not be able to safely negotiate the curve.

This critical point location is given by

$$d_{req} = \frac{v_o^2 - v_s^2}{2(x_w - t_r v_o)} \tag{5.1}$$

where

$v_o$ = initial forward speed,
$v_s$ = safe speed for the curve,
$t_r$ = driver reaction time,
$x_w$ = distance between the warning location and critical point (CP), and
$d_{req}$ = deceleration necessary to achieve safe speed at CP.




Source: Szabo and Wilson, 2004.

***Figure 5.2. Relationship between curve geometry and warning point.***

The lateral acceleration limit to determine safe speed entering the curve is

$$v_s = (a_s r)^{0.5} \tag{5.2}$$

where

$r$ = curve radius and
$a_s$ = lateral acceleration limit.

The authors used a similar concept for tangent sections based on the geometry of a lateral drift into a jersey barrier, as shown in Figure 5.3.

The equation to determine the warning location and necessary lateral acceleration to avoid a lateral lane departure is

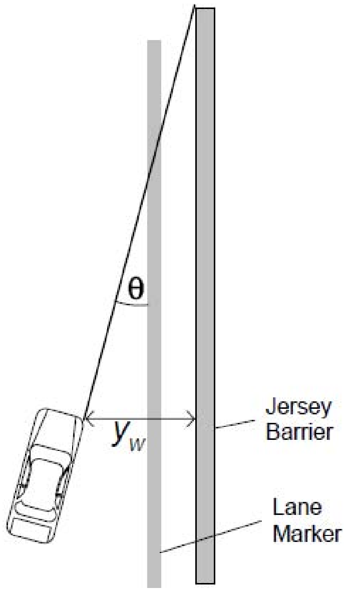
$$a_{lat} = \frac{(v_o \theta)^2}{2(y_w - t_r v_o \theta)} \tag{5.3}$$

where

$a_{lat}$ = lateral acceleration to avoid departure,
$\theta$ = departure angle, and
$y_w$ = distance from warning location and road boundary.

Movement beyond the crash prevention boundary represents a situation in which the vehicle is not likely to recover.

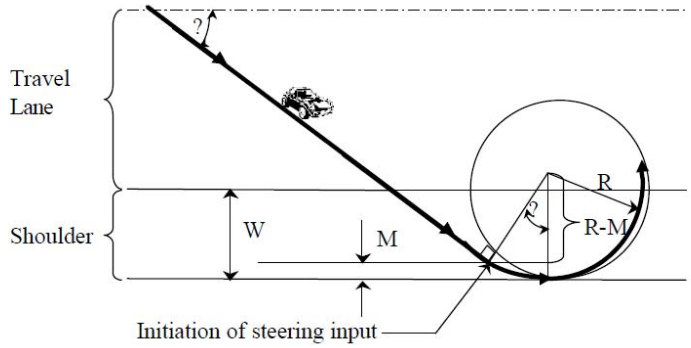
Burgett and Gunderson (2001) also discussed the concept of a crash prevention boundary for road departure crashes, which is a function of driver brake response time and the level of deceleration needed to avoid a crash. The authors discussed the concept in relation to a driver traveling at a constant speed




Source: Szabo and Wilson, 2004.

***Figure 5.3. Relationship between lane geometry and warning point.***

on a tangent roadway section. The point at which the vehicle crosses the lane edge is defined as $t = 0$, and the crash prevention is defined by the driver's steering maneuver and level of lateral acceleration created by the steering maneuver, as depicted in Figure 5.4. The authors use the geometric relationship between speed, side acceleration, departure angle, steering angle, reaction time, and radius of curve for curve sections to develop crash prevention boundaries. An example is shown in Figure 5.5 for a 1,000-ft radius curve with a vehicle speed of 50 mph, a shoulder width of 10 ft, and an initial vehicle offset of 2 ft from the edge of the road.




Source: Burgett and Gunderson, 2001.

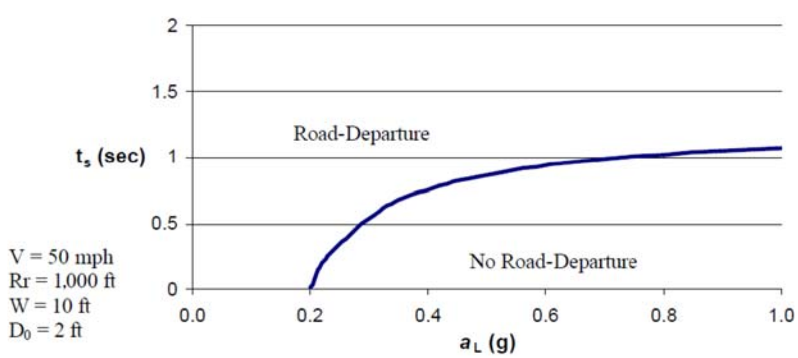
***Figure 5.4. Relationship of driver steering maneuver and prevention of a road departure.***

## Selection of Lane Departure Surrogates

One of the major benefits of naturalistic driving studies is that they can capture all levels of incidents, including those related to lane departures. One of the research questions addressed in this project involved assessing existing naturalistic driving study data and determining the most appropriate crash surrogates to use in lane departure analyses. Potential crash surrogates were also evaluated to determine what vehicle kinematic triggers could be used to flag potential lane departures in the full-scale study. Since a large amount of data will result from the full-scale NDS, it will be necessary to have some automated method to flag events of interest.

The following section organizes information related to lane departure crash surrogates and outlines a process that could be used to develop lane departure crash surrogates in the full-scale study.

Lane departure crashes provide a more complex situation for developing crash surrogates than other crash types. Most




Source: Burgett and Gunderson, 2001.

***Figure 5.5. Crash prevention boundary for a curve road departure.***

crash types, such as broadside or rear-end, can be defined by a time or distance metric (time or distance collision) because the hazard, collision with another vehicle, is clear. In a lane departure crash, the main hazard is not always clearly identified, and in some cases multiple hazards may be present. For instance, when a vehicle departs the edge of the roadway, multiple hazards may be present and multiple outcomes (sequences of events) may be possible. Potential hazards include encountering pavement edge drop-off, which could lead to loss of control; encountering differential friction between the roadway surface and shoulder, which could also lead to loss of control; having the vehicle rollover; or striking a fixed object.

The events accompanying each hazard can also result in a number of different outcomes, each of which can lead to the vehicle encountering different hazards. An initial sequence of events could result in several different outcomes based on different hazards, driver responses, and roadway conditions. Figure 5.6 shows three possible outcomes for the same initial ROR event. Each outcome may have a different type of crash surrogate to describe it. The first sequence of events for all three scenarios includes the vehicle running off the roadway to the right, encountering loose shoulder material, overcorrecting, and then crossing the centerline. In the first scenario, the vehicle then runs off the road to the left and strikes a tree. In the second scenario, the vehicle runs off the road to the left and overturns before striking the tree. In the third scenario, the vehicle crosses the centerline and strikes another vehicle head-on. The initial sequence of events was the same, but three different outcomes were possible with three different hazards. Each of the first events (e.g., ROR, encountering loose shoulder material) could have led to a different subsequent series of events. For instance, the vehicle could have left the roadway to the right, encountered loose shoulder material, and rolled over.

Each stage of a lane departure can result in a number of outcomes, and each outcome may need to be described by different crash surrogates. As a result, lane departures in the present study were divided into categories where hazards would be consistent. Five categories were selected, which include normal driving, lateral drift within the travel lane, right- or left-side lane departure where the vehicle stays within the traveled way (lane encroachment), right- or left-side lane departure where the vehicle leaves the traveled way (shoulder encroachment), and lane departure crash.

A crash surrogate or surrogates were identified for each category, except for normal driving because normal driving by definition is absence of conflict. A crash surrogate is the metric used to set boundaries between events and assess crash risk (e.g., time/distance to encroachment on the lane edge line). The crash surrogates and thresholds for each category were determined after reviewing the available literature on crash surrogate measures, evaluating existing naturalistic driving study data, and assessing what is likely to be available with the full-scale SHRP 2 naturalistic driving study. The crash surrogates and threshold parameters for each category are described below, along with the rationale for the selection. Table 5.1 summarizes the lane departure categories and associated crash surrogates.

## Normal Driving

- **Description:** This category represents the range of behavior remaining when crash surrogates and crash activity are removed.

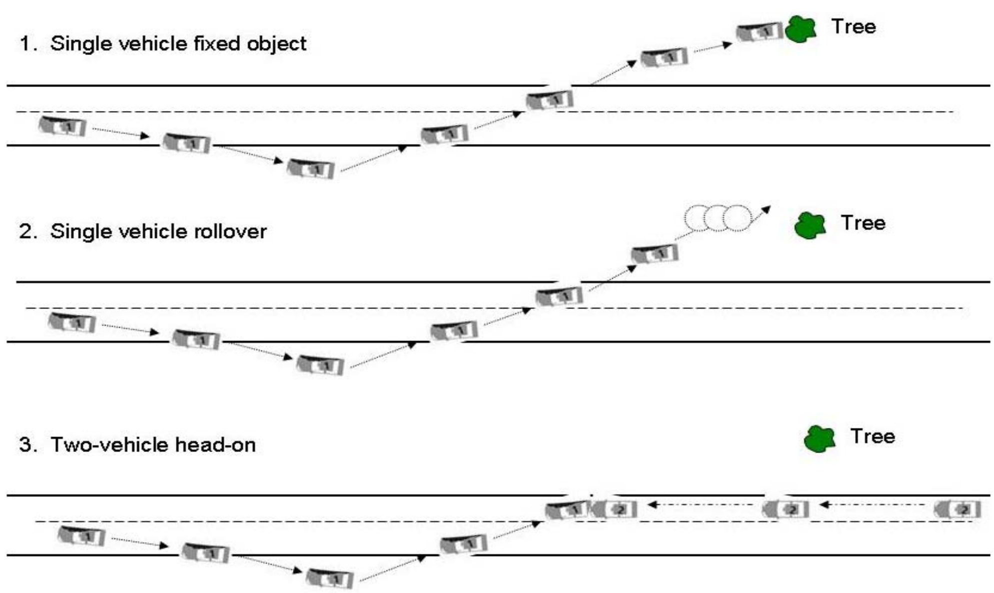



*Figure 5.6. Three possible roadway departure outcomes from same initial sequence of events.*

*Table 5.1. Summary of Lane Departure Categories*

| Category | Hazard | Surrogate | Metric |
|---|---|---|---|
| Normal driving | None | Lane deviation | ft |
| Lateral drift | Crossing lane line | Distance to lane departure (DTLD) or time to lane departure (TTLD) | s or ft |
| Lane encroachment | Sideswipe with adjacent or opposing vehicle | Time to collision (TTC) or distance to collision (DTC) | s or ft |
| | Head-on collision | TTC or DTC | s or ft |
| | Crossing lane line onto shoulder (left or right side) | Same as for shoulder encroachment<br>Time to lane edge (TTLE) or distance to lane edge (DTLE) | s or ft |
| Shoulder encroachment | Shoulder (loss of control) | Change in steering angle or yaw rate | Degree/s |
| | Rollover | Rollover potential | Lateral acceleration (*g*) |
| | Fixed object collision | TTC or DTC | s or ft |
| | Crossing shoulder edge | Time to shoulder edge (TTSE) or distance to shoulder edge (DTSE) | s or ft |
| Lane departure crash | Vehicle, rollover, fixed object | NA | NA |

- **Crash surrogates:** No crash surrogates are used for normal driving. However, amount of lane deviation is the metric used to distinguish between normal driving and a lateral drift.
- **Lower boundary:** This category has no lower boundary.
- **Upper boundary:** The boundary or threshold level between normal driving and lateral drift will need to be determined in the full-scale study. Lane keeping for individual drivers will vary, and drivers will maintain lane position differently based on a variety of factors, such as different roadway conditions (e.g., two-lane versus four-lane, presence and type of curve), weather conditions, time of day, and length of time driving. In order to set this threshold, it will be necessary to develop a range of normal vehicle activity under different situations and then determine what constitutes normal driving for a given scenario.

Depending on the resources available, normal driving can be established for individual drivers for situations of interest (daytime versus nighttime driving) or can be determined for a cohort of drivers. Some evaluation of what might define normal driving was conducted using the UMTRI data set and is discussed in the section "Identifying Lane Departure Incidents Using Existing Data Sets" (p. 65).

## Lateral Drift

- **Description:** This category includes incidents in which a vehicle's deviation within its lane or its direction of travel to the right or left will result in a lane departure unless the driver changes course.
- **Crash surrogates:** A single crash surrogate can be used to measure lateral drift because the only hazard is leaving the vehicle lane. The crash surrogate selected to assess lateral drift is distance to lane departure (DTLD), as shown in Figure 5.7. Distance is used rather than time because the lane tracking system used in the full-scale naturalistic study is expected to provide distance measurements.

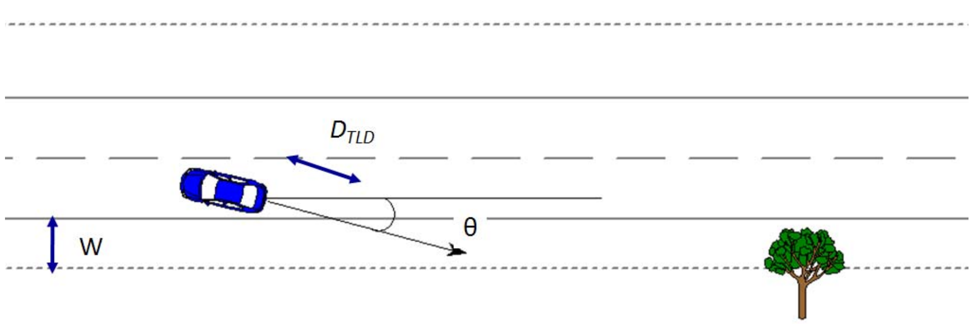



*Figure 5.7. Distance to lane departure.*

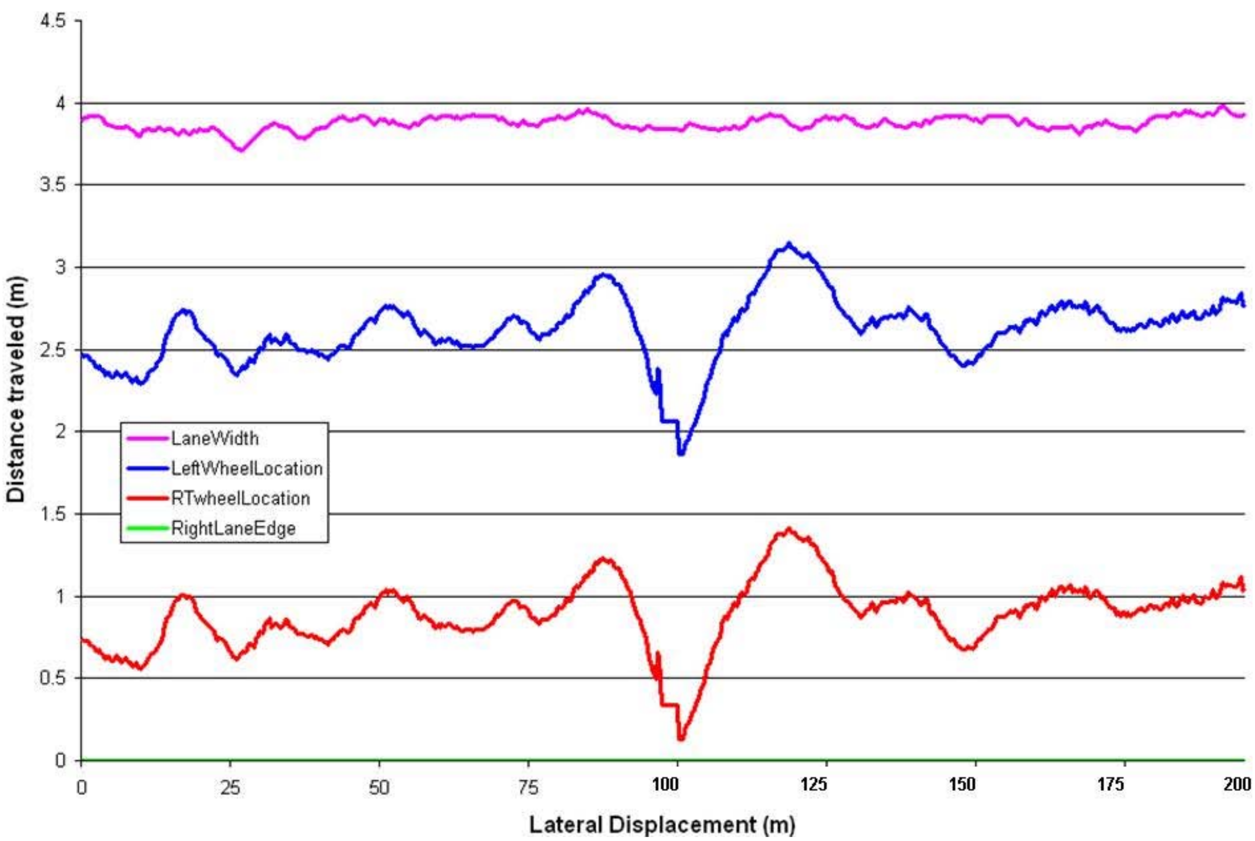
TTLD can also be calculated and used as the surrogate measure.

- **Lower boundary:** Lower boundary is the point where normal driving ends, which will need to be determined in the full-scale study. This category will need to be defined after examination of data in the full-scale NDS.
- **Upper boundary:** The upper boundary for lateral drift and a lane departure is where the vehicle's outside tires come within a certain tolerance distance ($X_{tol}$) of the lane line or lane boundary. A tolerance distance is necessary because lane tracking systems can only locate a vehicle within its lane to a certain level of accuracy. As a result, the tolerance distance reflects uncertainty in locating the vehicle. UMTRI used a tolerance distance of 0.1 m in its study. The distance that will be used in the SHRP 2 naturalistic study will depend on the accuracy of the lane tracking system used.
- **Data needs:** DTLD requires vehicle position relative to its lane. The GPS system that will be available with the full-scale system will not be accurate enough to locate a vehicle precisely within its lane. The only way to obtain vehicle lane position will be to use the lane tracking system that will be available with the vehicle instrumentation package. The planned lane position tracking system for the full-scale study has a stated accuracy of ±0.656 ft (0.2 m). This is within the range of a normal tire width and is likely to be adequate.
- **Limitations:** It will be difficult to calculate TTLC from any of the available SHRP 2 study variables, with the exception of the lane tracking system. The system is not expected to perform on gravel roads or in situations where either lane lines or some other lane delineation are not present (e.g., snow-covered roadway).

The vehicle trace from the UMTRI data, depicted in Figure 5.8, shows an example of a lateral drift, where the vehicle clearly drifted to the right but did not leave its lane.

## Lane Encroachment

- **Description:** This category includes incidents where a vehicle departs its original lane of travel and encroaches into an adjacent travel lane. This adjacent lane may have other vehicles traveling in the same or opposite directions. An encroachment is defined as one or more tires encroaching or crossing the edge of the lane line.
- **Crash surrogates:** The main hazard for an encroachment into an adjacent lane is a sideswipe or rear-end collision with




Source: UMTRI RDCW data set.

***Figure 5.8. Vehicle trace of nondeparture lateral drift.***

another vehicle. The main hazard for an encroachment into an oncoming lane is a head-on or opposing-direction sideswipe crash. The crash surrogate for a head-on, sideswipe, or rear-end collision with another vehicle is TTC. DTC can also be calculated. Level of risk can be defined by a threshold value as the point at which an evasive steering maneuver is required to avoid a collision.

If the subject vehicle does not encounter another vehicle after encroaching into one or more lanes of travel, the next proximate hazard is leaving the roadway. The crash surrogate in this case would be time to lane edge (TTLE) or distance to lane edge (DTLE).

- **Lower boundary:** This is the threshold for lateral drift, as described above.
- **Upper boundary:** The threshold between a lane encroachment and a shoulder encroachment is the point at which the vehicle's outside tires come within a certain tolerance distance ($X_{tol}$) of the lane line or lane boundary separating the lane and adjacent shoulder, similar to what was described for a lateral drift. The threshold between a lane encroachment and lane departure crash is the point at which the subject vehicle strikes another vehicle.
- **Data needs:** Calculation of time to collision requires vehicle speed, distance to adjacent or oncoming vehicle, coefficient of friction, grade, deceleration rate, and vehicle braking characteristics. Distance to adjacent vehicle will need to be determined using forward or side radar. Thus, time or distance to collision can only be calculated when a vehicle is within the tolerance of the radar systems.
- **Limitations:** The ability to determine time or distance to collision or lane edge depends on the accuracy of the lane tracking system and the accuracy with which and distance at which the instrumented vehicle radar system can track objects in its path.

### Shoulder Encroachment

- **Description:** This category includes incidents where a vehicle departs the traveled roadway surface onto a paved or unpaved shoulder. This is often referred to as a road departure or ROR incident.
- **Crash surrogates:** There can be several hazards once a vehicle leaves the traveled portion of the roadway. The first hazard encountered when leaving the traveled way is the shoulder itself. Specific shoulder hazards that might be encountered include differential friction between the roadway and unpaved shoulder and other shoulder irregularities (e.g., loose material, muddy shoulders) that may lead to loss of control or overturning of the vehicle.

Hazards are different for a paved shoulder and unpaved shoulder. The team met with several lane departure experts at the Iowa Department of Transportation (Iowa DOT) and FHWA, and it was decided that encroachments onto the shoulder under different circumstances present different levels of crash risk. An encroachment onto a paved shoulder introduces a lower level of risk, in the absence of hazards, than an encroachment onto an unpaved shoulder, even if TTC or time to shoulder edge (TTSE) is the same. The friction differential between the unpaved shoulder and paved roadway poses a risk for loss of control anytime the vehicle partially or fully leaves the paved roadway surface. This may be addressed by categorizing the encroachment into different levels of risk or considering time to paved shoulder edge as one crash surrogate and time to unpaved shoulder edge as another.

It is difficult to determine a crash surrogate for loss of control on the shoulder because it does not fit within any of the typical metrics used in crash surrogates. Changes of a certain magnitude in steering angle or yaw rate may be used to identify loss of control, but they are not crash surrogates per se. It may be necessary to define the next most likely sequence of events (overturn, return to travel lane, cross centerline) and then use the corresponding crash surrogate for that event.

Rollover potential is the crash surrogate when rollover is a possibility. Rollover potential is described in the section "Determining Rollover Potential" (p. 72).

The next hazard encountered when leaving the traveled way is collision with a fixed object. This can occur on the shoulder or when the vehicle leaves the shoulder. When a fixed object (e.g., tree, guardrail, mailbox, utility pole, bridge abutment) presents the most immediate hazard, the proposed crash surrogate is TTC or DTC.

Another hazard when leaving the roadway is that, once a vehicle leaves the shoulder, it may encounter an adverse slope, which may result in overturning. When the primary hazard is leaving the shoulder, the proposed crash surrogate is TTSE. Rollover risk may also be used.

Level of risk for most of the crash surrogates listed above can be defined by the actions that need to be taken to avoid a crash. The point between a lower risk and a higher risk event may be defined as the point at which a severe evasive action is required to avoid a crash. Once a vehicle leaves the roadway onto to the shoulder, the recovery options are braking to a stop before leaving the shoulder or striking an object, or steering back onto the original travel lane.

Evasive actions occur when the vehicle undergoes a steering or braking maneuver that exceeds normal steering or braking. AASHTO uses a deceleration rate of 11.2 ft/s$^2$ (0.35 $g$) for stopping distance because this deceleration is within the capability of most drivers to stay within their lane and control their vehicle when braking on wet surfaces (AASHTO, 2004). A value of 14.8 ft/s$^2$ (0.46 $g$) is used for emergency braking. VTTI used a lateral acceleration of ≥0.7 $g$ as a trigger that a lane departure incident had occurred (Dingus et al., 2006). Thus, between 0.35 $g$ and 0.7 $g$ is a good starting point for setting the threshold deceleration between an encroachment and a lane departure conflict. It will be necessary to examine a number of lane departure incidents and subjectively assess what constitutes an encroachment versus a lane departure conflict and then determine the boundary between normal and significant braking.

Time to collision ($t_{TTC}$) is a function of initial vehicle velocity, angle of departure ($\theta$), coefficient of friction ($f$) between the tires and shoulder, braking capabilities of the vehicle, driver reaction time, driver response, distance to the object ($d_{obj}$), grade, and deceleration rate ($a$). The time to collision ($t_{critical}$) that requires a critical deceleration rate ($a_{critical}$) in order for the vehicle to stop safely is the threshold between an encroachment and a lane departure conflict, as shown in Figure 5.9. The time to collision ($t_{norm}$) where a vehicle can stop safely with normal deceleration rates ($a_{norm}$), as shown in Figure 5.10, is given by the following:

$$t_{norm} = t_{TTC} - t_{critical} \tag{5.4}$$

Similarly, TTSE or distance to shoulder edge (DTSE) is a function of initial vehicle velocity, angle of departure ($\theta$), coefficient of friction ($f$) between the tires and shoulder, braking capabilities of the vehicle, driver reaction time, driver response, distance to the shoulder edge ($d_{shld}$), shoulder width ($w_{shld}$), grade, and deceleration rate ($a$). The distance the vehicle travels before crossing the edge of the shoulder (Figure 5.11) is given by the following:

$$d_{shld} = \frac{w_{shld}}{\sin(\theta)} \tag{5.5}$$

Less information was available about what constitutes a normal range of steering angles than was available for normal passenger vehicle deceleration rates. As a result, it may

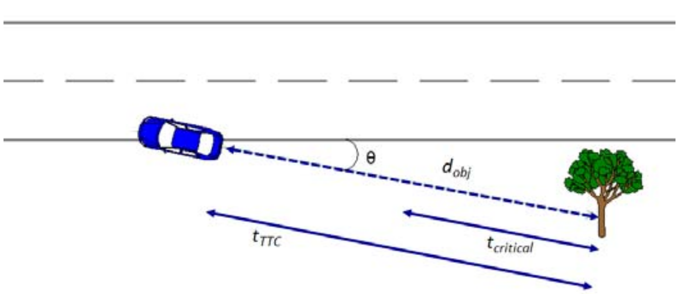



*Figure 5.9. Threshold requiring evasive deceleration.*

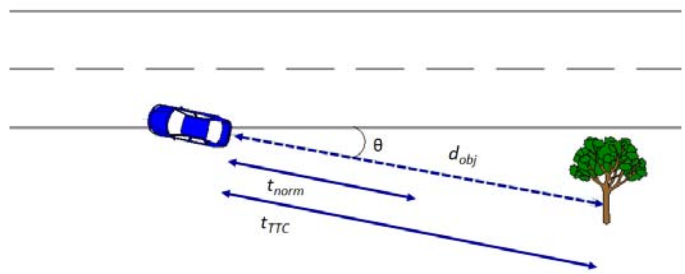



*Figure 5.10. Threshold requiring normal deceleration.*

also be necessary to examine a number of lane departure incidents and subjectively assess what constitutes a significant evasive action. The threshold defining a significant evasive action is the point at which a driver must employ excessive steering maneuvers in order to avoid the object or shoulder edge, as shown in Figures 5.12 and 5.13.

- **Lower boundary:** This is the threshold for lane encroachments as described above.
- **Upper boundary:** The threshold between a shoulder encroachment and a lane departure crash is when the vehicle physically strikes an object or physically rolls over (crash), which may defined as TTC or TTSE = 0.
- **Data needs:** Calculation of time or distance to collision or time or distance to shoulder edge requires vehicle speed, deceleration, angle of departure, shoulder width, distance to fixed object, coefficient of friction, grade, and deceleration rate.
- **Limitations:** In the full-scale NDS, friction will not be available. The expected spatial accuracy of roadside features is ±3.0 ft (0.914 m). The accuracy of the GPS, used to determine the vehicle's spatial position, is unknown, but the GPS will not have differential correction capabilities. Accuracy for a nondifferentially corrected GPS can be as low as ±15 m (49.2 ft). This would significantly affect the ability to calculate TTC. The vehicle instrumentation packages are expected to be able to determine distance and heading to objects using the forward or side radar. The vehicle instrumentation system will have a forward radar capable of tracking and storing information for the five objects closest to the vehicle. Objects can be identified and tracked for up to 200 m (656.2 ft) in front of the vehicle within ±0.324

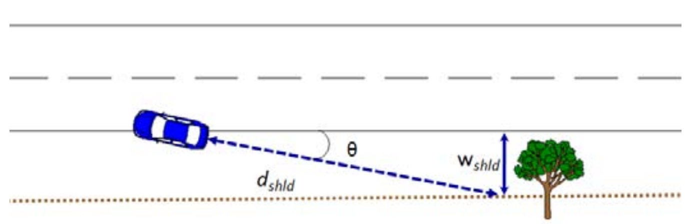



*Figure 5.11. Distance to edge of shoulder.*

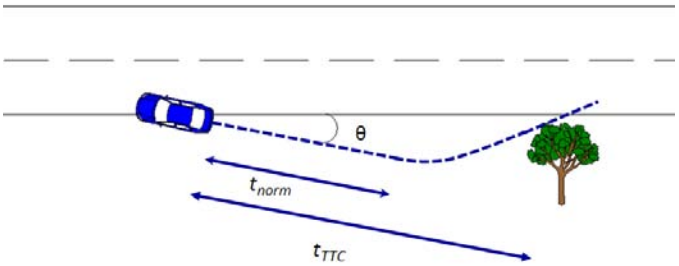



Figure 5.12. Threshold for normal steering.

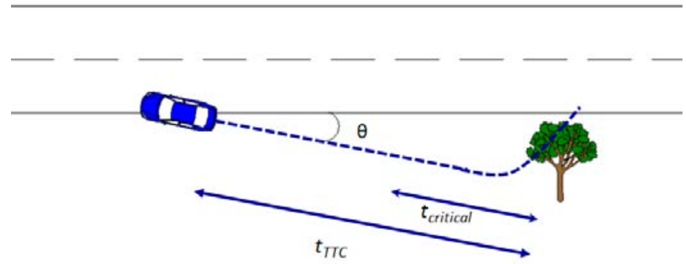


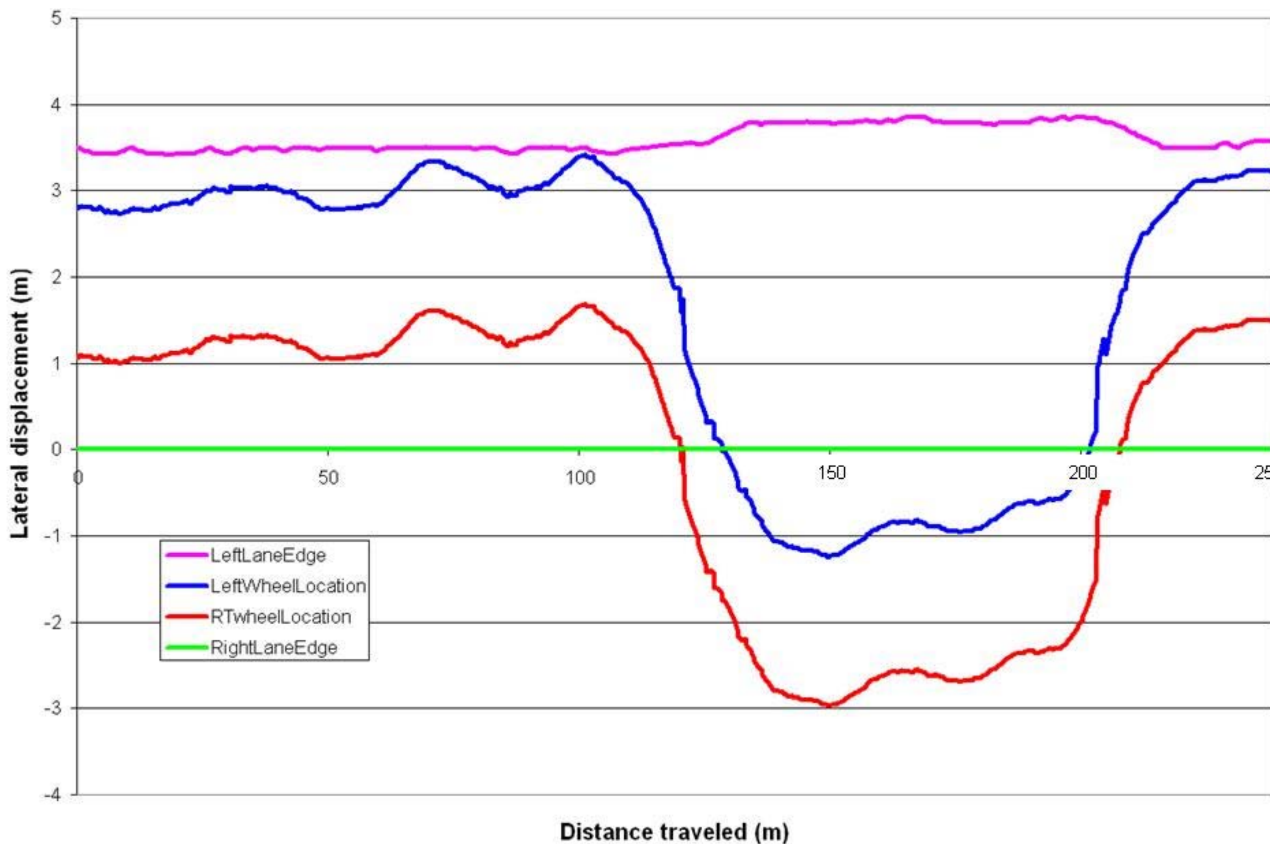

Figure 5.13. Threshold for evasive steering.

radians (18°) of the horizontal field of view, given that the field of view is centered on the test vehicle heading.

Figure 5.14 shows a typical vehicle trace for a nonconflict encroachment (UMTRI data). As shown, the vehicle leaves the roadway for some distance and then safely returns.

## Lane Departure Crash

- **Description:** A crash is defined as an incident where a vehicle strikes another vehicle or object (TTC or TTSE = 0). A vehicle overturning one or more times is also considered to be a crash. A crash may also be defined as a situation where the vehicle leaves the roadway and is forced to an unplanned stop. For instance, sliding off a roadway during a winter weather event and then sliding to a stop in the median may be considered a crash. This category includes all lane departure crashes, whether or not they are reported in a police document, that are observed in the naturalistic study. A reported collision is one where a crash report is filed. An unreported collision is one observed in the naturalistic study but for which a police accident report has not been filed; consequently, the crash would not show up in a crash database. For example, a driver leaves the roadway and strikes a mailbox but proceeds after recovering. In some states, there is no requirement to report property-damage-only crashes




Source: UMTRI RDCW data set.

Figure 5.14. Vehicle trace of nonconflict run-off-road incident.

unless the damage exceeds some value. The VTTI driving study found that out of 82 minor nonproperty-damage contact collisions, only 15 were reported to the police (Dingus et al., 2006).

- **Crash surrogates:** NA.
- **Lower boundary:** This is the threshold for lane or shoulder encroachments as previously described.
- **Upper boundary:** NA.
- **Data needs:** Identification of lane departure crashes will require setting triggers for vehicle kinematics that provide indications that a crash has occurred (e.g., sudden deceleration). It is expected that crashes will be identified as part of the data quality assurance (SHRP 2 Safety Project S06, Technical Coordination and Quality Control).
- **Limitations:** It is expected that VTTI will identify crashes during the full-scale instrumented vehicle data collection and that most crashes will be identified. However, the identification of crashes will be highly dependent on the thresholds used, and, as such, some crashes (e.g., vehicle sliding off the roadway during winter weather) may not be included.

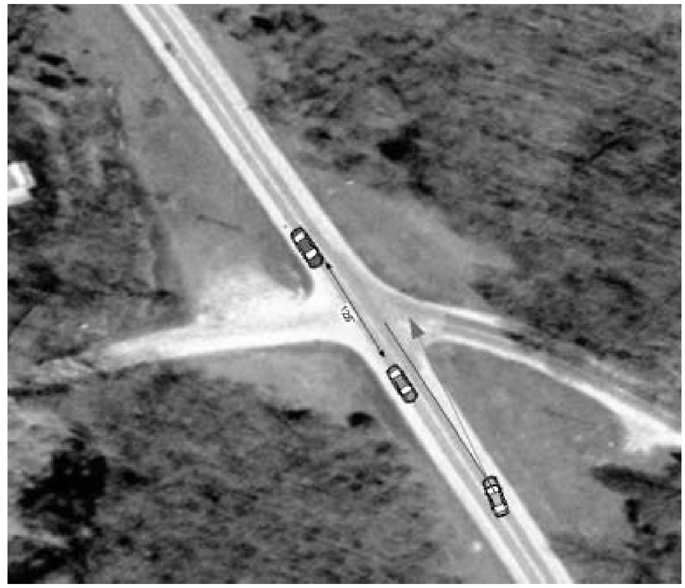


*Figure 5.15. Schematic of a vehicle departing its lane and its likely path if the driver does not correct lane departure.*

## Evaluating Incident Outcome with UMTRI Data

A number of lane departure incidents in the UMTRI data set were assessed so that they could be divided into categories and crash surrogates in order to test the system for its ability to categorize lane departures, as described in the previous sections.

For each lane departure, the hazard that the vehicle was most likely to encounter after departing its lane based on factors such as angle of departure, surrounding hazards, and vehicle speeds was determined. Hazards were those that presented the most imminent threat. The object (hazard) most likely to be struck was determined for each situation by estimating the anticipated vehicle path and by a visual inspection of the forward and aerial imagery. A hazard could also include an oncoming vehicle or the shoulder if no specific objects were in the vehicle's likely path.

Figure 5.15 shows an example of how potential hazards were determined for one vehicle. The subject's vehicle exited the roadway and encroached 2.1 ft onto a paved shoulder. If the vehicle were to continue along its path, the intersecting roadway provided the first hazard that the subject vehicle would encounter. If the vehicle returned to the roadway and overcorrected, the first hazard the vehicle would encounter was an oncoming vehicle, as shown in Figure 5.16. Each vehicle's position was determined for various points in time.

Figures 5.17 and 5.18 show other examples of how potential hazards can be determined. In this case, the vehicle departed its lane to the left and crossed the centerline. The forward imagery and aerial imagery for the location were examined for potential hazards (Figure 5.18). The most imminent hazard for the scenario if the vehicle were to continue on its current path (to the left) or overcorrect to the right was determined using vehicle speed, vehicle position, and location of potential hazards. At the vehicle's current speed and trajectory, if the vehicle did not correct its path, the most imminent hazard was the left guardrail. The shoulder is paved and poses a low hazard. If the vehicle overcorrected, an estimation of the potential paths off the right side of the roadway indicated that the most likely hazard was a mailbox. No oncoming vehicles were present, so collision with another vehicle was not a potential threat.

*Figure 5.16. Schematic of vehicle departing its lane and overcorrecting.*

Source: UMTRI RDCW data set.

***Figure 5.17. Forward image for where vehicle departed the roadway (cross centerline).***

## Identifying Lane Departure Incidents Using Existing Data Sets

The full-scale SHRP 2 naturalistic driving study will result in a tremendous amount of data. This amount will necessitate an automated method to identify lane departures. An automated method would entail selecting variables within the naturalistic driving data sets that are most likely to experience a significant change during a lane departure, and establishing a threshold value for these variables so they can be used as flags for potential lane departures.

### Triggers Used in Other Naturalistic Driving Studies

VTTI selected relevant variables and triggers to set thresholds between valid and invalid critical events (Dingus et al., 2006). The study used a sensitivity analysis to evaluate placement of triggers at various levels. If a trigger is set too low (Type I error, or lower sensitivity), a larger percentage of actual incidents is selected, as well as a larger number of nonincidents (false alarms). This results in longer and less useful data reduction time. Alternatively, if the trigger is set too high (Type II error), nonincidents are less likely to be selected, but a larger number of actual incidents may be missed as well.

VTTI used an iterative process to select triggers for valid incidents. Triggers were set to a lower sensitivity, and data reduction was used to evaluate resulting incidents. VTTI researchers used a normal distribution to depict how Type I and Type II errors could be minimized based on signal detection theory, as shown in Figure 5.19. The final triggers for variables related to lane departure incidents include the following:

- Lateral acceleration ≥0.7 *g*;
- Longitudinal acceleration ≥0.6 *g*;
- Longitudinal acceleration ≥0.5 and forward TTC ≤4 s;
- Longitudinal deceleration 0.4 *g* to 0.5 *g*, forward TTC ≤4 s, and distance to collision <100 ft; and
- Yaw rate ≥ |4°| change in heading within a 3-s window of time.

McLaughlin et al. (2009) evaluated ROR crashes and near crashes using the VTTI 100-car study. These researchers identified ROR maneuvers by evaluating steering wheel position, yaw rate, and braking, as shown in Figure 5.20.

The University of Iowa teen driver study used a trigger of 0.5 *g* for lateral acceleration to indicate when a potential incident had occurred (McGehee et al., 2007).

### Evaluation of Lane Departure Thresholds Using UMTRI Data

The naturalistic driving study data from UMTRI and VTTI were used to evaluate which variables may be the most useful

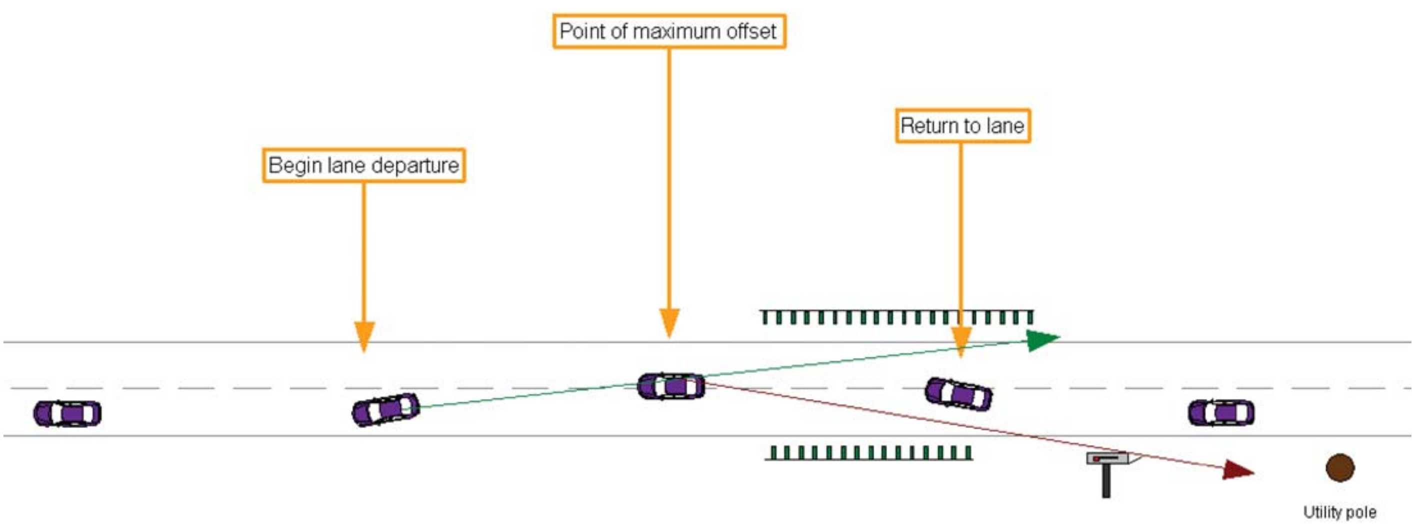
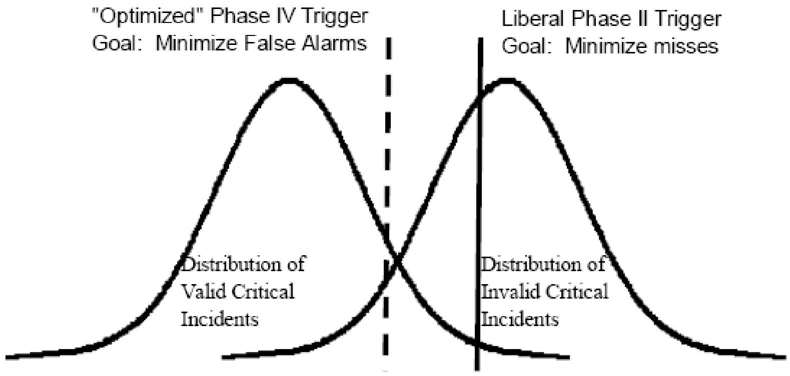



***Figure 5.18. Schematic of potential outcomes and hazards encountered.***




Source: Dingus et al., 2006.

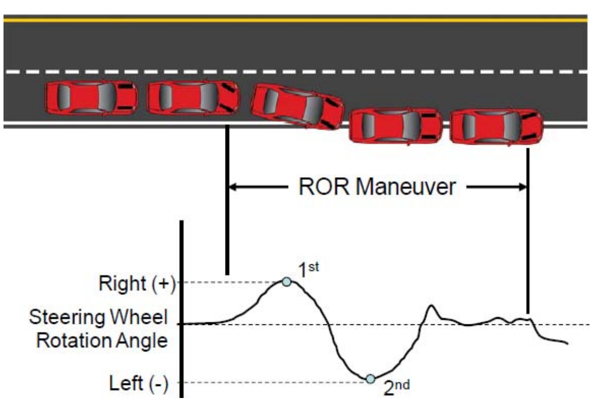
***Figure 5.19. Graphical depiction of setting trigger criteria using the distribution of valid events.***

in setting triggers to identify lane departure events and to assess what thresholds may be used. Data were reduced as described in Appendix A.

The UMTRI data indicated a number of encroachments, but no conflicts or crashes. Only data for rural, paved, two-lane roadways were included. The VTTI data provided near crashes and crashes, but no encroachments. Additionally, variables were not consistent between the two data sets. The two data sets therefore were evaluated separately. This section describes the evaluation using the UMTRI data, and the following section describes the evaluation using the VTTI data.

The first section below examines differences in kinematic variables between normal driving data and left- and right-side lane departures. The second section compares normal driving to assess variables that could be used to partition normal driving data. The third section discusses sample size issues.

It should be noted that left-side lane departure in curves is sometimes intentional and, rather than being due to an unintentional lane departure, is due to the driver intentionally "cutting the curve." While the researchers did not account for this specifically, this should be considered when evaluating lane departures.




Source: McLaughlin et al., 2009.

***Figure 5.20. Steering wheel angle relative to ROR maneuver.***

### *Examining Kinematic Variables*

The UMTRI data reduction resulted in 22 right-side and 51 left-side lane departure events for two-lane rural roads. An incident was a situation where the vehicle departed its lane by 0.1 m or more at some point. All incidents were considered to be encroachments because the vehicle departed its lane in each case but was not forced to take some evasive maneuver and did not lose control on the shoulder.

The continuous data surrounding each incident was extracted. Data for which no incident had occurred was termed "normal" driving data. Several variables were examined to determine whether they could be used to set thresholds between normal driving and lane departure incidents. The maximum positive and negative value for each incident was extracted for the following vehicle kinematic variables: lateral speed, lateral acceleration, yaw rate, forward acceleration, roll rate, and pitch rate.

The maximum negative and positive values for various vehicle kinematic variables for right-side and left-side events were compared with approximately 105,400 records (in 0.1-s intervals) of nonincident (normal) driving. Data for each kinematic variable (lateral speed, yaw rate, side acceleration, forward acceleration, roll rate, and pitch rate) for the normal data were graphed against the data for the lane departure events.

Figure 5.21 shows the distribution of data for the kinematic variable "lateral or side speed (in m/s)" for left- and right-side lane departures that are all approximately normal. The distribution for normal driving is the center distribution, shown in maroon. The distribution of maximum positive side acceleration for left-side lane departures is shown to the right in green, and the distribution of maximum negative side acceleration is shown to the left in blue. Data for left- and right-side

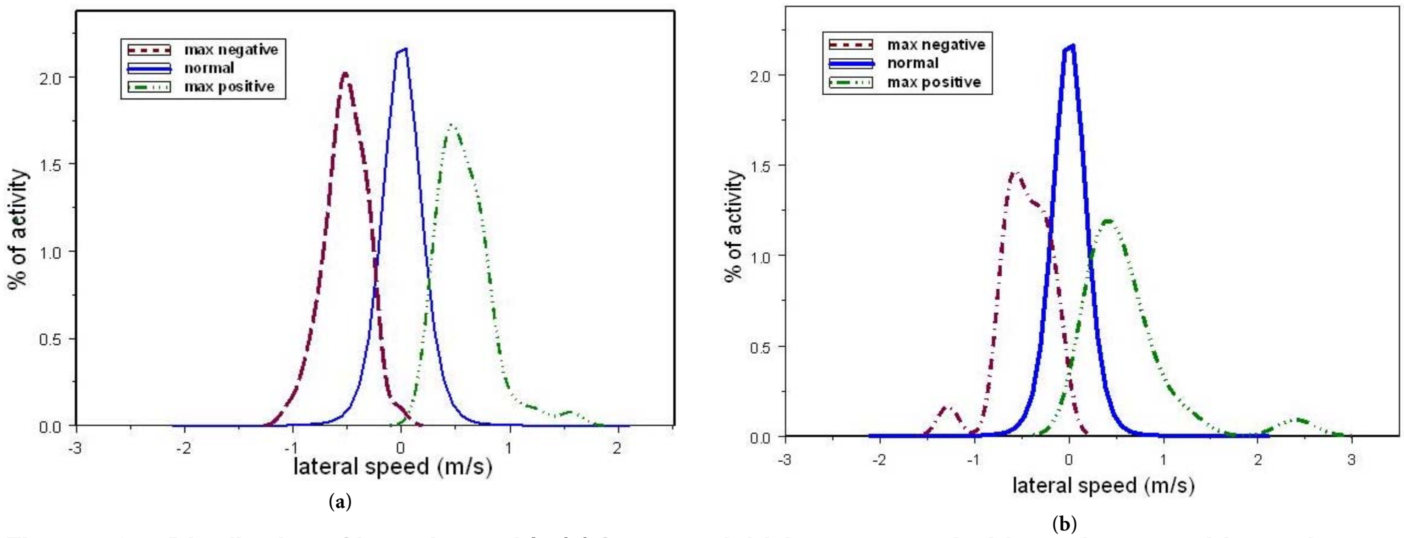



***Figure 5.21. Distribution of lateral speed (m/s) for normal driving compared with maximum positive and negative values from (a) left-side and (b) right-side lane departures.***

lane departures were evaluated separately because they have different kinematic signatures. The analysis showed differences in acceleration among the left- and right-side departures compared with normal driving ($p < 0.05$ for all comparisons).

Figure 5.22 shows the distribution of data for the kinematic variable "forward acceleration for left- and right-side lane departures in *g*'s." Figures 5.23 to 5.26 show the same information for lateral or side acceleration in *g*'s, roll rate in degrees per second, pitch rate in degrees per second, and yaw rate in degrees per second, respectively.

Because the values for the left side and right side were discrete, a ranking test was used. The Wilcoxon Rank Sum Test was used to determine whether the normal driving data were statistically different from events data. The test determines whether two independent samples of observations are from the same distribution. It evaluates the sign and magnitude of the rank of differences between pairs, and assesses whether two independent samples have similar rankings. In all cases except maximum low yaw rate for right-side lane departures, the test showed that the data were statistically different.

Although the distributions for most variables were determined to be different at the 95% level of significance, a significant amount of overlap exists, as shown in Figures 5.21 to 5.26. Thus, setting a higher threshold to ensure that a larger

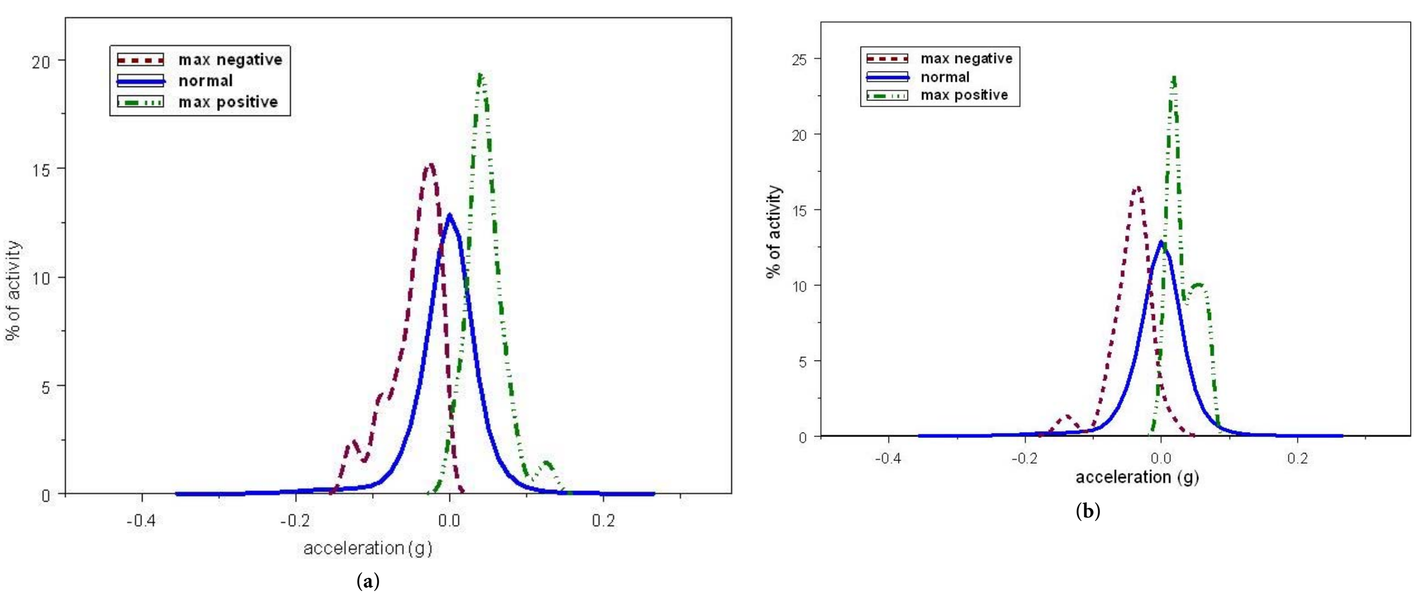
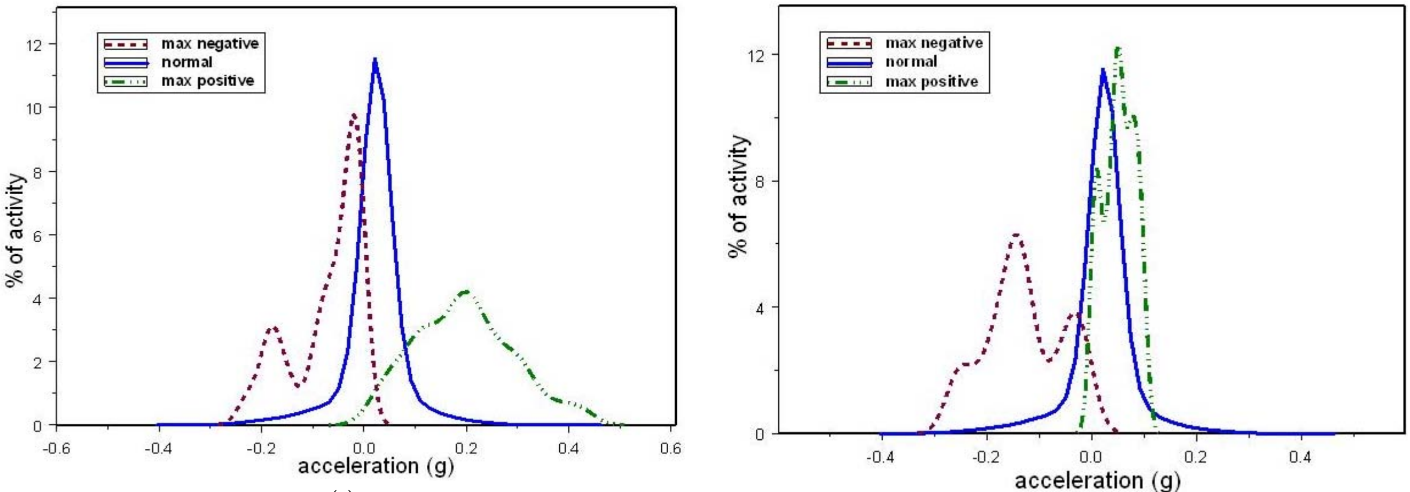
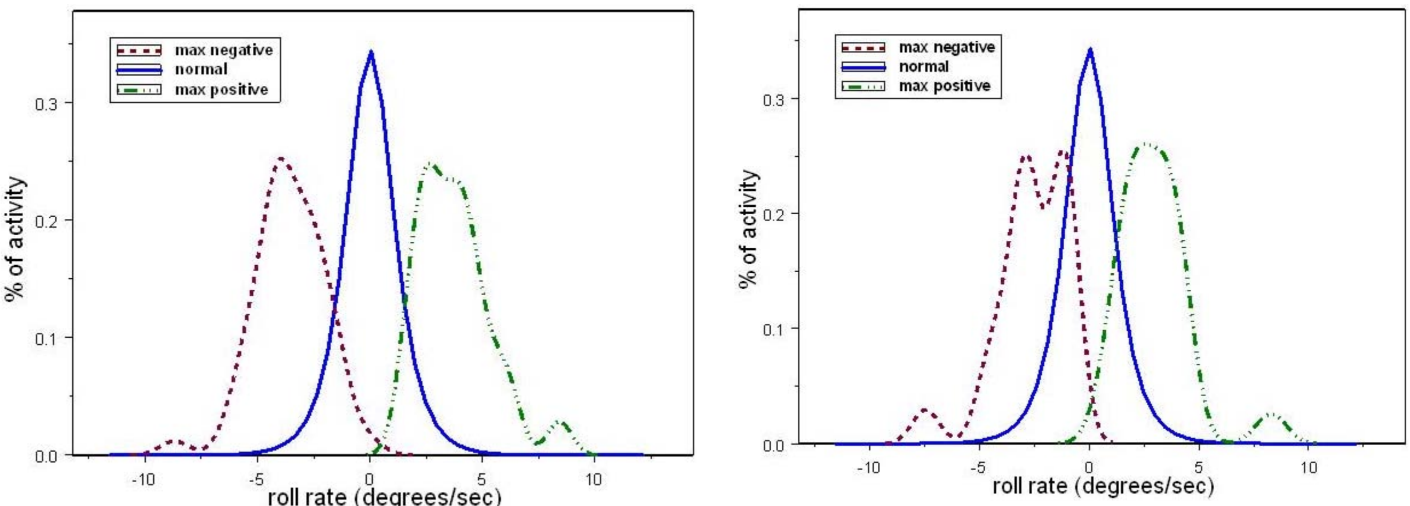



***Figure 5.22. Distributions of forward acceleration (g) for normal driving compared with maximum positive and negative values from (a) left-side and (b) right-side lane departures.***



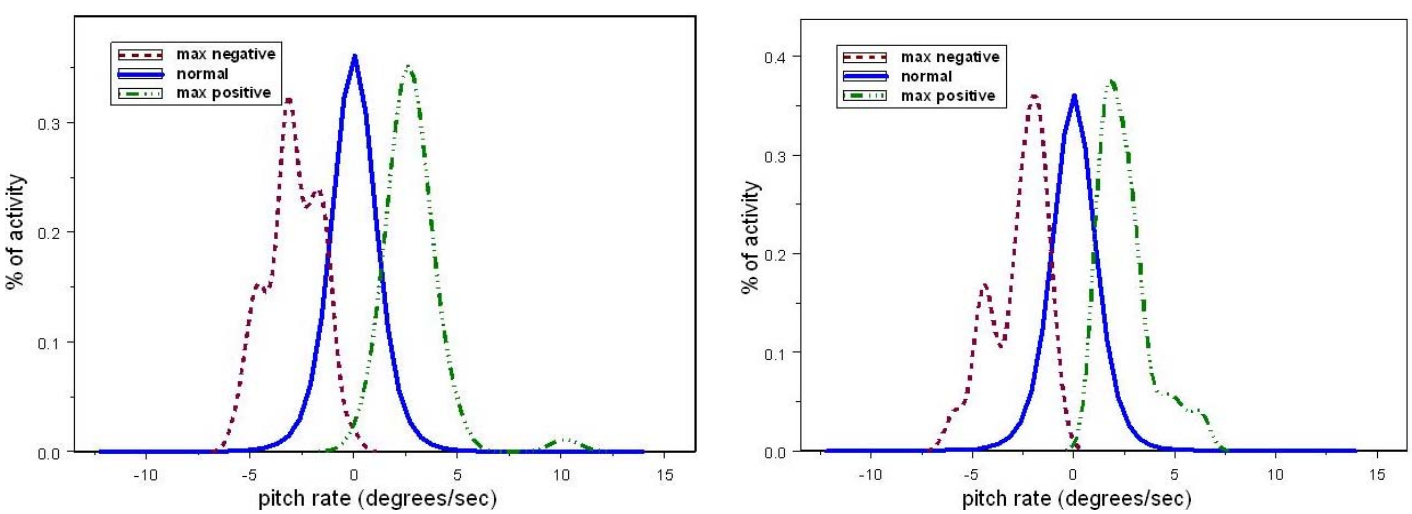

(a) (b)

***Figure 5.23. Distributions of lateral acceleration (g) for normal driving compared with maximum positive and negative values from (a) left-side and (b) right-side lane departures.***




(a) (b)

***Figure 5.24. Distributions of roll rate (degrees/s) for normal driving compared with maximum positive and negative values from (a) left-side and (b) right-side lane departures.***




(a) (b)

***Figure 5.25. Distributions of pitch rate (degrees/s) for normal driving compared with maximum positive and negative values from (a) left-side and (b) right-side lane departures.***

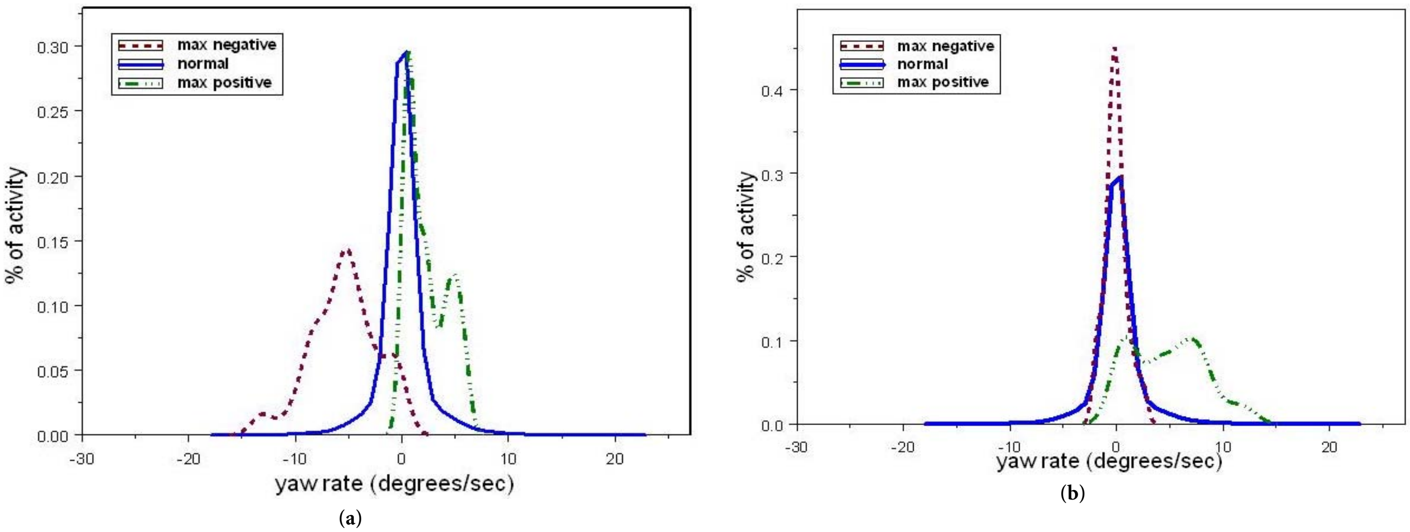



***Figure 5.26. Distributions of yaw rate (degrees/s) for normal driving compared with maximum positive and negative values from (a) left-side and (b) right-side lane departures.***

number of normal driving conditions are not included may result in a threshold that will likely miss a large number of events. The alternative is also true: setting a lower threshold to include most events may also result in the inclusion of a large number of nonevents, which will result in more unnecessary data reduction. This is similar to what VTTI found as its study evaluated methods to set appropriate triggers, as discussed in the section "Triggers Used in Other Naturalistic Driving Studies" (p. 65).

To summarize the data, differences exist in vehicle kinematic values for lane departure events and for normal driving. However, although the data can be shown to be different and statistically significant, a considerable overlap still exists. This indicates the difficulty in setting thresholds low enough to include all incidents but still high enough so that a large amount of nonincident data does not have to be evaluated.

Tables 5.2 and 5.3 show the range of maximum and minimum values for each kinematic variable. The values for lateral speed in Table 5.2 can be interpreted to mean that the continuous data for each left-lane departure included at least one value between –1.04 m/s and –0.03 m/s and at least one value with a lateral speed of 0.20 m/s or higher. If these values represented

***Table 5.2. Range of Maximum Negative and Positive Values for Left-Lane Departure Events for UMTRI Data***

| Lateral Speed (m/s) | | Yaw Rate (deg/s) | | Side Acceleration (*g*) | |
|---|---|---|---|---|---|
| **Max Negative** | **Max Positive** | **Max Negative** | **Max Positive** | **Max Negative** | **Max Positive** |
| –1.04 to –0.03 | 0.20 to 1.56 | –13.3 to –0.2 | 0.10 to 6.15 | –0.23 to –0.01 | 0.02 to 0.42 |
| **Forward Acceleration (m/s²)** | | **Roll Rate (deg/s)** | | **Pitch Rate (deg/s)** | |
| –0.13 to –0.03 | 0.03 to 0.13 | –8.70 to –0.06 | 1.45 to 8.54 | –5.64 to –0.13 | 0.48 to 10.19 |

***Table 5.3. Range of Maximum Negative and Positive Values for Right-Lane Departure Events for UMTRI Data***

| Lateral Speed (m/s) | | Yaw Rate (deg/s) | | Side Acceleration (m/s²) | |
|---|---|---|---|---|---|
| **Max Negative** | **Max Positive** | **Max Negative** | **Max Positive** | **Max Negative** | **Max Positive** |
| –1.28 to –0.08 | 0.22 to 2.40 | –1.90 to –0.05 | 0.15 to 12.20 | –0.26 to –0.01 | 0.01 to 0.1 |
| **Forward Acceleration (m/s²)** | | **Roll Rate (deg/s)** | | **Pitch Rate (deg/s)** | |
| –0.14 to –0.02 | 0.01 to 0.07 | –7.43 to –0.76 | 0.69 to 8.29 | –5.86 to –0.81 | 1.05 to 6.26 |

a large number of lane departure incidents, they could be used as a starting point to set threshold values for flagging lane departures. However, as indicated, there were not enough samples of incidents to determine what threshold values should be set. It will be necessary to set thresholds after examination of a much larger number of incidents in the full-scale study. However, initial results suggest that for left-side lane departures, roll rate, yaw rate, side acceleration, and side speed are likely to be good candidates to identify events. Results suggest that for right-side lane departures, yaw rate, side acceleration, and lateral speed are good candidates to identify encroachments. Ayers et al. (2004), for instance, indicated that yaw rate is the first indication that a potential vehicle movement may be occurring.

### *Selecting Parameters to Partition Driving Environments*

The analysis in the previous sections included all incidents and normal driving data that were extracted for two-lane rural roads. In reality, vehicle kinematic variables will differ among different driver, roadway, and environmental factors; in the large-scale study, differences in vehicle operation under different roadways and environments should be considered and statistically controlled. The Wilcoxon Rank Sum Test was used to compare whether distributions of lateral offset were different under several driving scenarios.

Although only a limited amount of data was available, an exploratory analysis of differences in normal data for different situations was made to get a sense of how data might be partitioned. The lane offset variable, which indicates the offset of the vehicle's center from the center of the lane, was used for comparison. Normal driving data on a tangent section was compared with normal driving on a right-hand and left-hand curve (orientation of curve from the perspective of the driver—e.g., a right-hand curve to the right). As shown in Figure 5.27, lateral offset differs from a tangent section to a right- or left-hand curve, and lateral offset on a right-hand curve also differs from lateral offset on a left-hand curve. Data were not sufficient to compare lateral offset for different curve radii, but this variable is expected to have a large impact. Results indicated that differences were statistically significant at the 95% level of significance.

Differences in lateral offset were also compared for daytime versus nighttime driving. As indicated in Figure 5.28, the mean lateral offsets in the two situations were similar, but the results of the Wilcoxon Rank Sum Test indicate that the distributions were different at the 95% level of significance. Data were compared for tangent sections only.

Differences between drivers were also compared. Lateral offset by driver for normal driving is shown in Figures 5.29 and 5.30. Data were compared for tangent sections only. As shown, lateral offset among drivers varies significantly. Distribution of lateral offset was compared among all drivers and differences were statistically significant for all driver pairs except for Drivers 14 and 48, Drivers 14 and 60, and Drivers 64 and 85. This is the expected result because different drivers have different driving styles.

Lateral offset was compared for several situations, as described in the previous paragraphs. Differences were noted

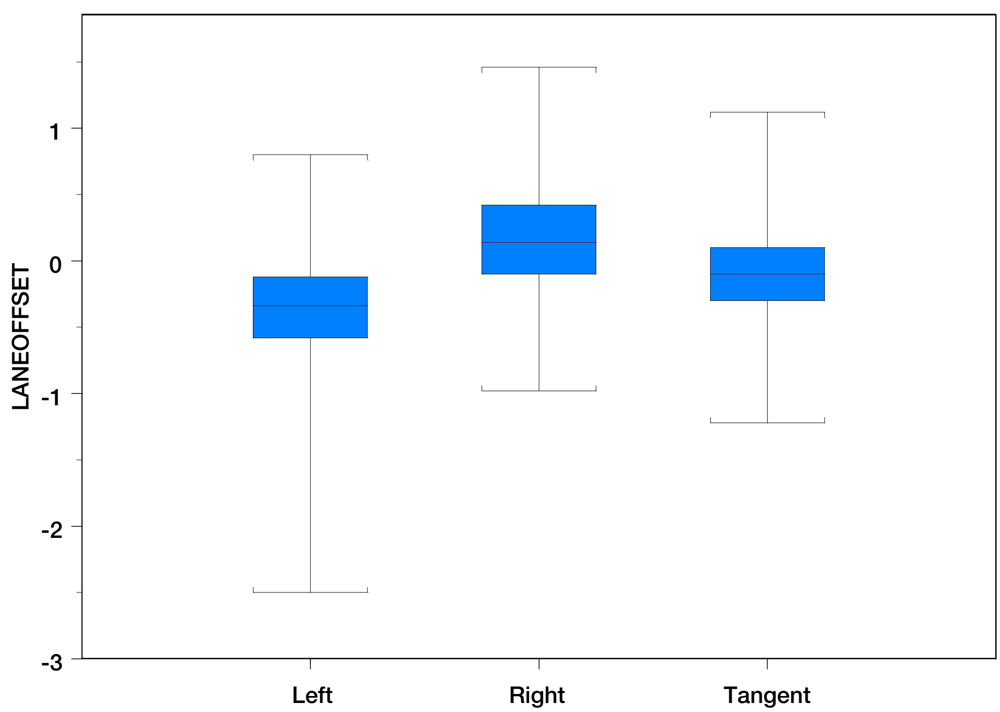



*Figure 5.27. Vehicle offset for tangent section versus right and left curve.*

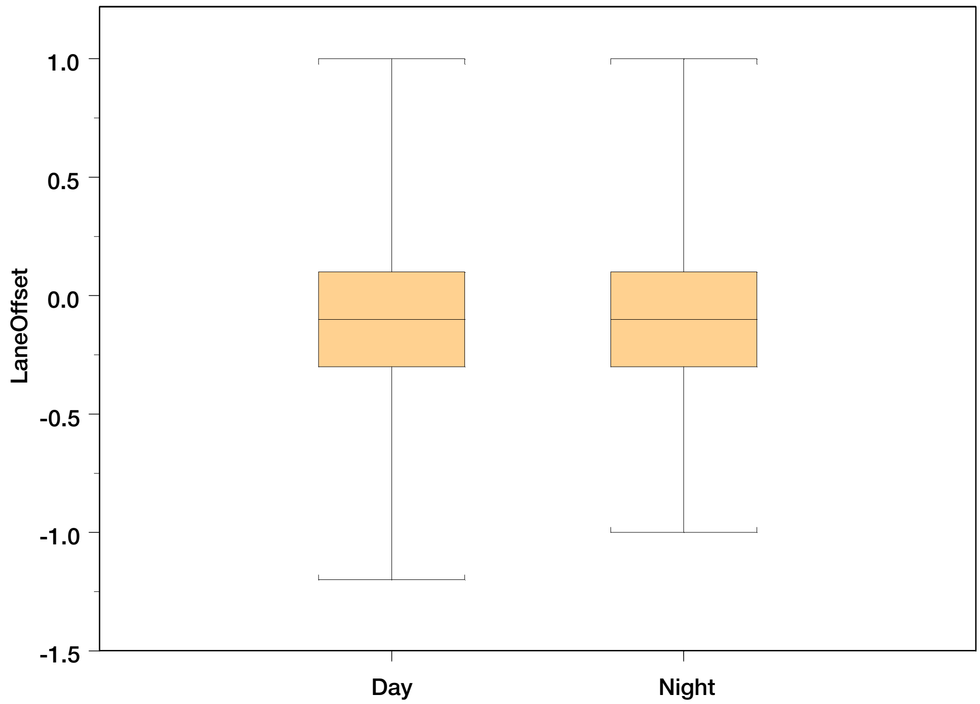



*Figure 5.28. Vehicle offset for daytime versus nighttime driving on tangent sections.*

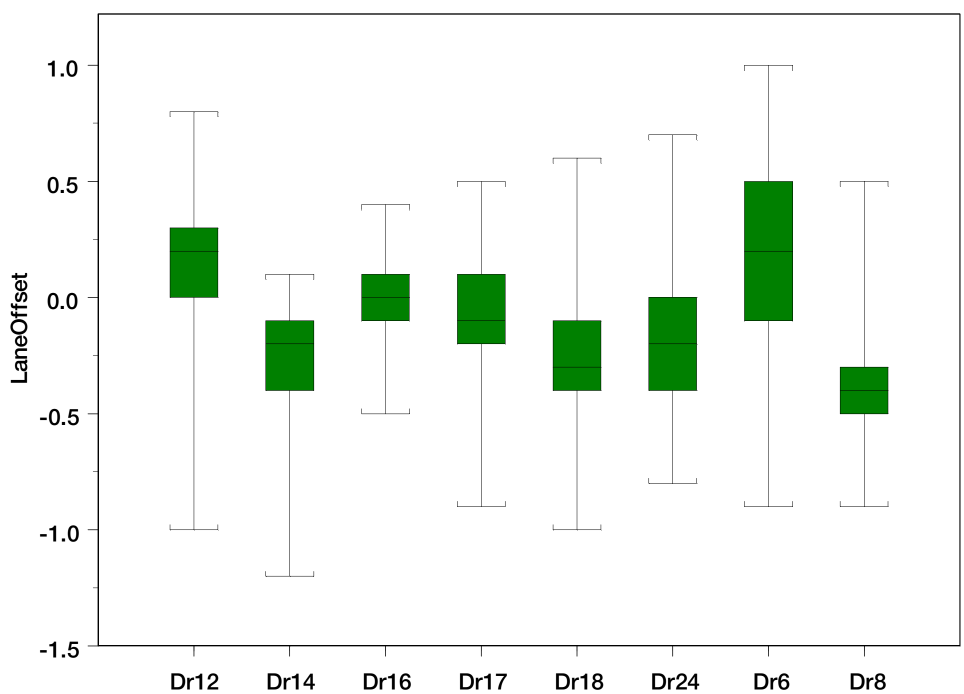



*Figure 5.29. Vehicle offset for Drivers 6, 8, 12, 14, 16, 17, 18, and 24 for tangent sections.*

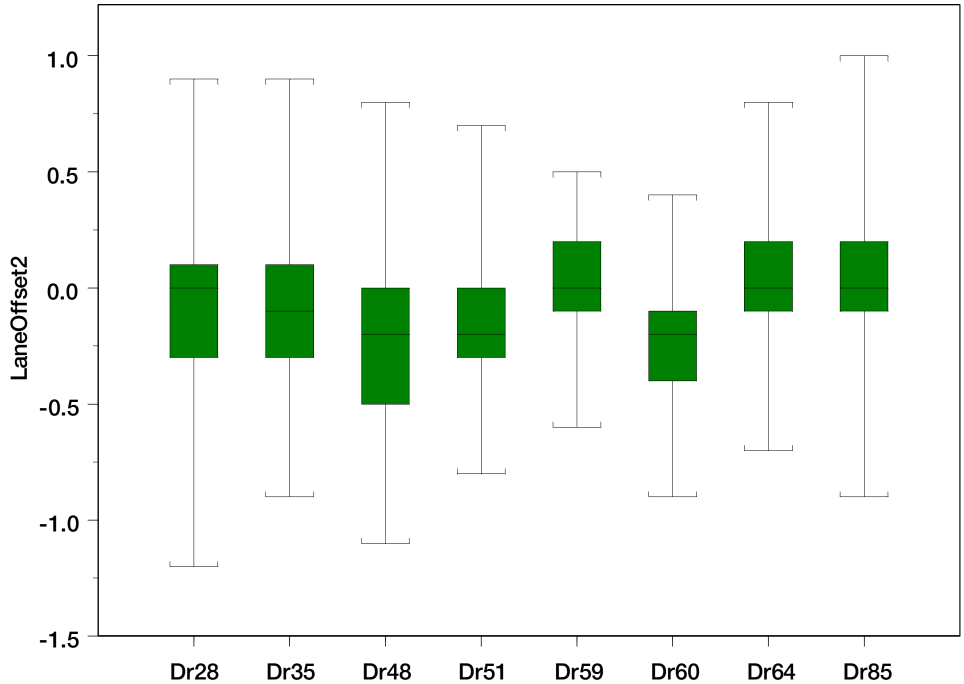



***Figure 5.30. Vehicle offset for Drivers 28, 35, 48, 51, 59, 60, 64, and 85 for tangent sections.***

between driving on a tangent and on a left- or right-hand curve, between nighttime and daytime driving, and between individual drivers. As indicated, differences are expected as to what constitutes normal driving behavior. Normal driving can be stratified by a large number of variables. Assuming the focus of lane departures will be on rural roadways, the minimum roadway and environmental characteristics should include the following:

- Roadway type (e.g., two lane, four-lane undivided, four-lane divided);
- Tangent versus curve;
- Radius of curve (may be aggregated to ranges of curve radii);
- Paved versus unpaved shoulders;
- Narrow versus wide shoulders;
- Dry versus wet versus snow- or ice-covered roadways;
- Nighttime versus daytime driving, including presence of overhead street lighting;
- Presence of rumble strips;
- Posted speed limit; and
- Roadway surface (paved versus dirt or gravel).

### Evaluation of Lane Departure Thresholds Using VTTI Data

The VTTI and UMTRI data were evaluated separately because different data were available from each. Additionally, the UMTRI data had only data for encroachments, while the VTTI data had only data for crashes and near crashes. The disadvantage of the VTTI data was that it included all roadway types and only had a limited sample size (n = 29). Additionally, no exposure (normal) driving data were available to the research team, so event thresholds could not be compared against normal driving conditions.

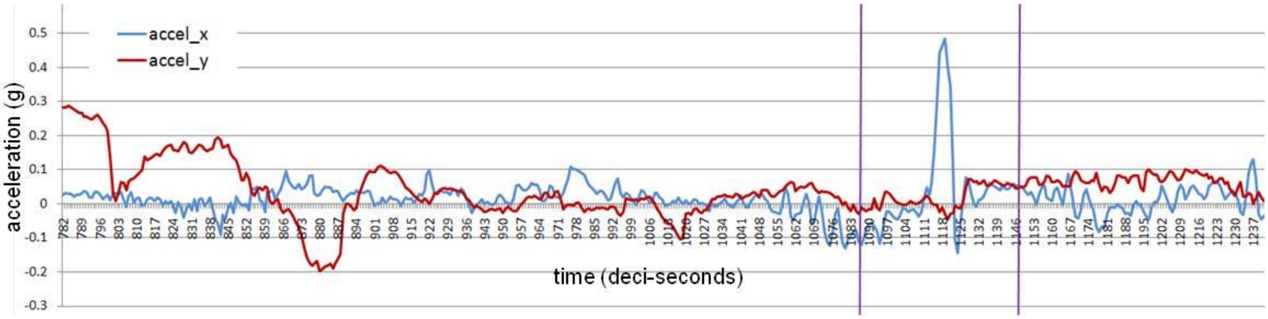
Thresholds for the crash and near-crash incidents available from the VTTI data were evaluated. Continuous data were available for 29 lane departure incidents. The crash and near-crash events showed distinct changes in forward and side acceleration, which were the only two variables that could be used to evaluate the data. Several examples are shown in Figures 5.31 and 5.32.

Table 5.4 shows the range of maximum and minimum values for each kinematic variable for each lane departure event. The values for side acceleration in Table 5.2 indicate that all left-lane departures had at least one negative value that was –0.01 *g* and at least one positive value that was greater than 0.02 *g*. As shown in Table 5.3, each right-side lane departure had at least one negative value for side acceleration that was –0.01 *g* or lower and one positive that was 0.01 *g* or higher. These values could therefore be used as starting points in identifying lane departure events.

## Determining Rollover Potential

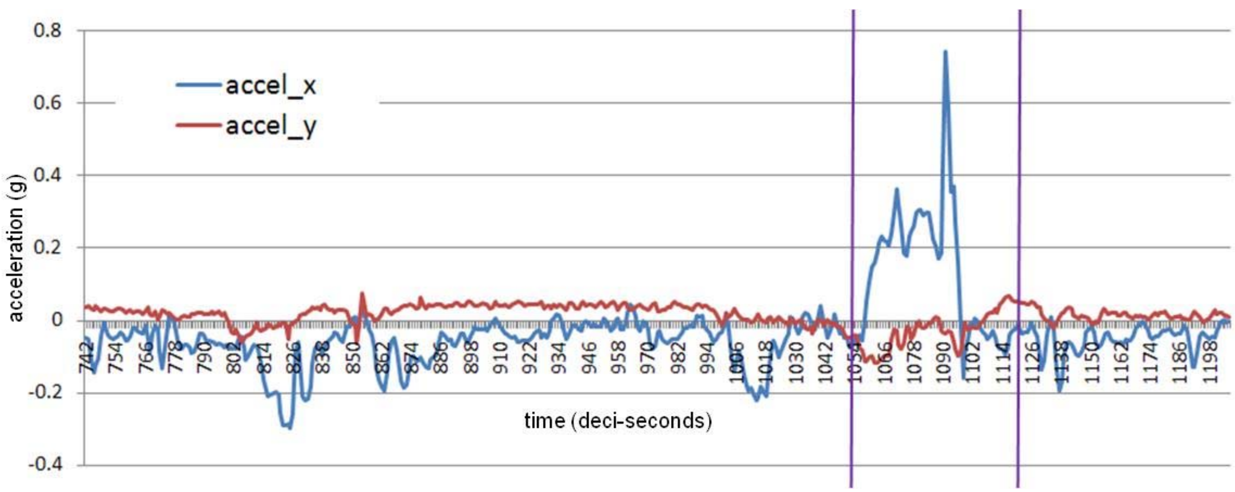
### Background

On average, approximately 274,000 light vehicles were involved in rollover crashes annually between 1999 and 2003 (NHTSA,




Source: VTTI data set.

*Figure 5.31. Forward and side acceleration trace for near crash on two-lane roadway.*




Source: VTTI data set.

*Figure 5.32. Forward and side acceleration trace for near crash on two-lane roadway.*

2005b). Although rollover crashes made up only 2% of crashes, they accounted for almost one-third of light vehicle occupant fatalities (including 59% of sport-utility vehicle fatalities) in 2003. Rollover crashes accounted for 10,182 fatalities for passenger vehicle occupants in 2007 (Insurance Institute for Highway Safety, 2009).

Most rollovers occur when a driver loses control of a vehicle and the vehicle begins to slide sideways. At this point, a curb, guardrail, tree stump, or soft or uneven ground on the side of the roadway can "trip" the vehicle and cause it to roll over. Rollovers are also caused by a driver turning too aggressively either at high velocity or with a sharp turning radius, causing the vehicle to tip up and then roll over. Rollovers can also take place after a collision with another vehicle.

## Estimating Rollover Propensity

The risk of a vehicle being involved in a rollover depends on a number of factors, including the vehicle's center of gravity, vehicle design, friction between surface and tires, steering input, roadway geometry, and vehicle speed. As a result, determining the likelihood that a vehicle will roll over can be fairly complicated. The simplified methods that follow, however, have been used to estimate roll propensity.

***Table 5.4. Range of Maximum Negative and Positive Values for VTTI Crash and Near-Crash Events***

| | Side Acceleration (*g*) | | Forward Acceleration (*g*) | |
|---|---|---|---|---|
| | **Max Negative** | **Max Positive** | **Max Negative** | **Max Positive** |
| Left-lane departure | –0.68 to –0.05 | 0.03 to 0.23 | –0.80 to –0.03 | 0.03 to 0.74 |
| Right-lane departure | –2.69 to –0.01 | 0.08 to 0.29 | –1.90 to –0.11 | 0.05 to 0.95 |

**Metric**

| Design Speed (km/h) | Maximum e (%) | Limiting Values of f | Total (e/100 + f) | Calculated Radius (m) | Rounded Radius (m) |
|---|---|---|---|---|---|
| 20 | 4.0 | 0.18 | 0.22 | 14.3 | 15 |
| 30 | 4.0 | 0.17 | 0.21 | 33.7 | 35 |
| 40 | 4.0 | 0.17 | 0.21 | 60.0 | 60 |
| 50 | 4.0 | 0.16 | 0.20 | 98.4 | 100 |
| 60 | 4.0 | 0.15 | 0.19 | 149.1 | 150 |
| 70 | 4.0 | 0.14 | 0.18 | 214.2 | 215 |
| 80 | 4.0 | 0.14 | 0.18 | 279.8 | 280 |
| 90 | 4.0 | 0.13 | 0.17 | 375.0 | 375 |
| 100 | 4.0 | 0.12 | 0.16 | 491.9 | 490 |
| 20 | 6.0 | 0.18 | 0.24 | 13.1 | 15 |
| 30 | 6.0 | 0.17 | 0.23 | 30.8 | 30 |
| 40 | 6.0 | 0.17 | 0.23 | 54.7 | 55 |
| 50 | 6.0 | 0.16 | 0.22 | 89.4 | 90 |
| 60 | 6.0 | 0.15 | 0.21 | 134.9 | 135 |
| 70 | 6.0 | 0.14 | 0.20 | 192.8 | 195 |
| 80 | 6.0 | 0.14 | 0.20 | 251.8 | 250 |
| 90 | 6.0 | 0.13 | 0.19 | 335.5 | 335 |
| 100 | 6.0 | 0.12 | 0.18 | 437.2 | 435 |
| 110 | 6.0 | 0.11 | 0.17 | 560.2 | 560 |
| 120 | 6.0 | 0.09 | 0.15 | 755.5 | 755 |
| 130 | 6.0 | 0.08 | 0.14 | 950.0 | 950 |
| 20 | 8.0 | 0.18 | 0.28 | 12.1 | 10 |
| 30 | 8.0 | 0.17 | 0.25 | 28.3 | 30 |
| 40 | 8.0 | 0.17 | 0.25 | 50.4 | 50 |
| 50 | 8.0 | 0.16 | 0.24 | 82.0 | 80 |
| 60 | 8.0 | 0.15 | 0.23 | 123.2 | 125 |
| 70 | 8.0 | 0.14 | 0.22 | 175.3 | 175 |
| 80 | 8.0 | 0.14 | 0.22 | 228.9 | 230 |
| 90 | 8.0 | 0.13 | 0.21 | 303.6 | 305 |
| 100 | 8.0 | 0.12 | 0.20 | 393.5 | 395 |
| 110 | 8.0 | 0.11 | 0.19 | 501.2 | 500 |
| 120 | 8.0 | 0.09 | 0.17 | 666.6 | 665 |
| 130 | 8.0 | 0.08 | 0.18 | 831.3 | 830 |
| 20 | 10.0 | 0.18 | 0.28 | 11.2 | 10 |
| 30 | 10.0 | 0.17 | 0.27 | 26.2 | 25 |
| 40 | 10.0 | 0.17 | 0.27 | 46.6 | 45 |
| 50 | 10.0 | 0.16 | 0.26 | 75.7 | 75 |
| 60 | 10.0 | 0.15 | 0.25 | 113.3 | 115 |
| 70 | 10.0 | 0.14 | 0.24 | 160.7 | 160 |
| 80 | 10.0 | 0.14 | 0.24 | 209.9 | 210 |
| 90 | 10.0 | 0.13 | 0.23 | 277.2 | 275 |
| 100 | 10.0 | 0.12 | 0.22 | 357.7 | 360 |
| 110 | 10.0 | 0.11 | 0.21 | 453.5 | 455 |
| 120 | 10.0 | 0.09 | 0.19 | 596.5 | 595 |
| 130 | 10.0 | 0.08 | 0.18 | 738.9 | 740 |
| 20 | 12.0 | 0.18 | 0.30 | 10.5 | 10 |
| 30 | 12.0 | 0.17 | 0.29 | 24.4 | 25 |
| 40 | 12.0 | 0.17 | 0.29 | 43.4 | 45 |
| 50 | 12.0 | 0.16 | 0.28 | 70.3 | 70 |
| 60 | 12.0 | 0.15 | 0.27 | 104.9 | 105 |
| 70 | 12.0 | 0.14 | 0.26 | 148.3 | 150 |
| 80 | 12.0 | 0.14 | 0.26 | 193.7 | 195 |
| 90 | 12.0 | 0.13 | 0.25 | 255.0 | 255 |
| 100 | 12.0 | 0.12 | 0.24 | 327.9 | 330 |
| 110 | 12.0 | 0.11 | 0.23 | 414.0 | 415 |
| 120 | 12.0 | 0.09 | 0.21 | 539.7 | 540 |
| 130 | 12.0 | 0.08 | 0.20 | 665.0 | 665 |

**US Customary**

| Design Speed (mph) | Maximum e (%) | Limiting Values of f | Total (e/100 + f) | Calculated Radius (ft) | Rounded Radius (ft) |
|---|---|---|---|---|---|
| 15 | 4.0 | 0.175 | 0.215 | 70.0 | 70 |
| 20 | 4.0 | 0.170 | 0.210 | 127.4 | 125 |
| 25 | 4.0 | 0.165 | 0.205 | 203.9 | 205 |
| 30 | 4.0 | 0.160 | 0.200 | 301.0 | 300 |
| 35 | 4.0 | 0.155 | 0.195 | 420.2 | 420 |
| 40 | 4.0 | 0.150 | 0.190 | 563.3 | 565 |
| 45 | 4.0 | 0.145 | 0.185 | 732.2 | 730 |
| 50 | 4.0 | 0.140 | 0.180 | 929.0 | 930 |
| 55 | 4.0 | 0.130 | 0.170 | 1190.2 | 1190 |
| 60 | 4.0 | 0.120 | 0.160 | 1505.0 | 1505 |
| 15 | 6.0 | 0.175 | 0.235 | 64.0 | 65 |
| 20 | 6.0 | 0.170 | 0.230 | 116.3 | 115 |
| 25 | 6.0 | 0.165 | 0.225 | 185.8 | 185 |
| 30 | 6.0 | 0.160 | 0.220 | 273.6 | 275 |
| 35 | 6.0 | 0.155 | 0.215 | 381.1 | 380 |
| 40 | 6.0 | 0.150 | 0.210 | 509.6 | 510 |
| 45 | 6.0 | 0.145 | 0.205 | 660.7 | 660 |
| 50 | 6.0 | 0.140 | 0.200 | 836.1 | 835 |
| 55 | 6.0 | 0.130 | 0.190 | 1065.0 | 1065 |
| 60 | 6.0 | 0.120 | 0.180 | 1337.8 | 1340 |
| 65 | 6.0 | 0.110 | 0.170 | 1662.4 | 1660 |
| 70 | 6.0 | 0.100 | 0.160 | 2048.5 | 2050 |
| 75 | 6.0 | 0.090 | 0.150 | 2508.4 | 2510 |
| 80 | 6.0 | 0.080 | 0.140 | 3057.8 | 3060 |
| 15 | 8.0 | 0.175 | 0.255 | 59.0 | 60 |
| 20 | 8.0 | 0.170 | 0.250 | 107.0 | 105 |
| 25 | 8.0 | 0.185 | 0.245 | 170.8 | 170 |
| 30 | 8.0 | 0.160 | 0.240 | 250.8 | 250 |
| 35 | 8.0 | 0.155 | 0.235 | 348.7 | 350 |
| 40 | 8.0 | 0.150 | 0.230 | 465.3 | 465 |
| 45 | 8.0 | 0.145 | 0.225 | 502.0 | 500 |
| 50 | 8.0 | 0.140 | 0.220 | 760.1 | 760 |
| 55 | 8.0 | 0.130 | 0.210 | 963.5 | 965 |
| 60 | 8.0 | 0.120 | 0.200 | 1204.0 | 1205 |
| 65 | 8.0 | 0.110 | 0.190 | 1487.4 | 1485 |
| 70 | 8.0 | 0.100 | 0.180 | 1820.9 | 1820 |
| 75 | 8.0 | 0.090 | 0.170 | 2213.3 | 2215 |
| 80 | 8.0 | 0.080 | 0.160 | 2675.6 | 2675 |
| 15 | 10.0 | 0.175 | 0.275 | 54.7 | 55 |
| 20 | 10.0 | 0.170 | 0.270 | 99.1 | 100 |
| 25 | 10.0 | 0.165 | 0.265 | 157.8 | 160 |
| 30 | 10.0 | 0.160 | 0.280 | 231.5 | 230 |
| 35 | 10.0 | 0.155 | 0.255 | 321.3 | 320 |
| 40 | 10.0 | 0.150 | 0.250 | 428.1 | 430 |
| 45 | 10.0 | 0.145 | 0.245 | 552.9 | 555 |
| 50 | 10.0 | 0.140 | 0.240 | 696.8 | 695 |
| 55 | 10.0 | 0.130 | 0.230 | 879.7 | 880 |
| 60 | 10.0 | 0.120 | 0.220 | 1094.6 | 1095 |
| 65 | 10.0 | 0.110 | 0.210 | 1345.8 | 1345 |
| 70 | 10.0 | 0.100 | 0.200 | 1838.8 | 1840 |
| 75 | 10.0 | 0.090 | 0.190 | 1980.3 | 1980 |
| 80 | 10.0 | 0.080 | 0.180 | 2378.3 | 2380 |
| 15 | 12.0 | 0.175 | 0.295 | 51.0 | 50 |
| 20 | 12.0 | 0.170 | 0.290 | 92.3 | 90 |
| 25 | 12.0 | 0.165 | 0.285 | 146.7 | 145 |
| 30 | 12.0 | 0.160 | 0.280 | 215.0 | 215 |
| 35 | 12.0 | 0.155 | 0.275 | 298.0 | 300 |
| 40 | 12.0 | 0.150 | 0.270 | 396.4 | 395 |
| 45 | 12.0 | 0.145 | 0.265 | 511.1 | 510 |
| 50 | 12.0 | 0.140 | 0.260 | 643.2 | 845 |
| 55 | 12.0 | 0.130 | 0.250 | 809.4 | 810 |
| 60 | 12.0 | 0.120 | 0.240 | 1003.4 | 1005 |
| 65 | 12.0 | 0.110 | 0.230 | 1228.7 | 1230 |
| 70 | 12.0 | 0.100 | 0.220 | 1489.8 | 1490 |
| 75 | 12.0 | 0.090 | 0.210 | 1791.7 | 1790 |
| 80 | 12.0 | 0.080 | 0.200 | 2140.5 | 2140 |

Note: In recognition of safety considerations, use of $e_{max}$ = 4.0% should be limited to urban conditions.

Source: AASHTO, 2001. 

*Figure 5.33. Maximum side friction factors.*

One of the main methods to assess rollover risk is to use the static stability factor (SSF) (NHTSA, 2005a). SSF is given by Equation 5.6:

$$\mathrm{SSF} = \frac{t}{2h} \tag{5.6}$$

where

$t$ = vehicle track width and
$h$ = height to center of gravity.

Gillespie (1992) expanded the concept to develop the relationship between forces acting on the vehicle and a measure of vehicle stability against rollover. The relationship is given by

$$\frac{a_{threshold}}{g} = \frac{t}{2h} + \varphi \tag{5.7}$$

where

$a_{threshold}$ = maximum side acceleration sustained before a vehicle engages in rollover and
$\varphi$ = cross slope (for flat roads) or superelevation (for curves).

In designing horizontal curves, the radius is calculated (AASHTO, 2001) using the formula

$$R = \frac{v^2}{127(e+f)g} \tag{5.8}$$

and

$$\varphi = \frac{v^2}{127Rg} - f \tag{5.9}$$

where

$v$ = advisory speed or design speed limit in m/s,
$f$ = safe side friction coefficient,
$g$ = 9.81 m/s$^2$, and
$R$ = radius of the curve expressed in m.

Substituting Equation 5.9 into Equation 5.7 yields the following:

$$\text{Rollover threshold (RT)} = \frac{a_{threshold}}{g} = \frac{t}{2h} + e$$
$$= \frac{t}{2h} + \frac{v^2}{127Rg} - f \tag{5.10}$$

Friction ($f$) can be obtained using Figure 5.33.

## Summary

The value of the naturalistic driving study is the ability to gain insight on crashes that may not be observed using other data collection approaches (e.g., crash databases, test tracks, driving simulators). Thus, even though the number of crashes may not be as representative, naturalistic studies do capture many useful safety and crash surrogates that may not be observed in police-reported crashes but can provide more insight into what can precipitate a crash—before a crash actually occurs.

This report has outlined information that may be used to develop crash surrogates for lane departures. Selection of crash surrogates for lane departures is not an easy task because multiple hazards can be present for each lane departure, and different surrogates may need to be specified depending on the most likely hazard. Lane departures are partitioned into categories and the surrogates are defined for the hazards most likely to be encountered.

Additionally, the team evaluated lane departures in the UMTRI and VTTI NDSs and identified some starting points for setting triggers for the full-scale study.

CHAPTER 6

# Analytical Tools and Initial Analysis of Lane Departure Research Questions

This chapter outlines several exploratory analytical approaches that were used to evaluate the existing naturalistic driving study data and that may be appropriate for analyzing the data that will result from the full-scale naturalistic driving study data to answer a variety of lane departure research questions.

The first is a data mining approach (classification and regression tree analysis). The second uses odds ratio and logistic regression. The third approach describes how the exploratory method used in the second can be expanded for the full-scale study to account for repeated measurements. The fourth approach is a time series analysis. Each approach uses data sampled in a different way. Each is described in a separate section, with the following information provided for each section:

- Background information that describes the general methodology;
- Details on how the approach was used to conduct an initial analysis of existing naturalistic driving studies;
- Results from the initial analysis of existing data;
- Considerations for the full-scale study, with a particular focus on data reduction and sampling;
- Limitations of the method using the existing data; and
- Expected limitations and advantages for the full-scale study.

Information common to all methods, such as data sampling approaches and data reduction, is provided in separate sections.

## Objective

The objective of this analysis plan was to develop and explore methodologies to answer research questions relating to lane departure crashes. The focus was to identify which roadway, driver, vehicle, and environmental factors are the best explanatory variables in predicting an increased likelihood of lane departures and lane departure crashes.

Improved data about actual events that lead to a lane departure crash or a noncrash incident will be extremely valuable in developing a better understanding of what negative factors lead to crashes and near misses, as well as of the factors that result in more positive subsequent events and outcomes. Understanding why crashes did not occur yields as much useful information as evaluating why they did occur. In both cases, factors that cause a vehicle to initially leave the roadway and the relationship between road, environment, vehicle, and human factors and subsequent events and outcomes can be studied. Dingus et al. (2006) reported that analysis of near crashes from the VTTI naturalistic driving study was valuable, as it demonstrated drivers successfully performing evasive maneuvers.

The intent of answering lane departure research questions is to provide roadway agencies and other practitioners with information about which factors positively or negatively influence the likelihood of a lane departure. A better understanding of roadway factors will allow agencies to better address safety in roadway design and assess the benefits of various countermeasures, such as rumble strips, flattening or better delineating curves, mandating paved shoulders on reconstruction and rehabilitation projects, and policy. A better understanding of driver factors related to lane departures will allow agencies to make better policy decisions, such as addressing younger driver training and licensing. A better understanding of environmental factors will enable agencies, for instance, to make informed winter maintenance decisions and determine trade-offs in application of street lighting.

## Audience

The primary audiences who can utilize the information obtained from answering lane departure research question in the full-scale study are state, county, and local transportation agencies and policy makers. Consequently, the information obtained should be in a format that can be used to make informed decisions about improved highway design during initial design and during reconstruction and rehabilitation. The information can also be used to select appropriate roadway countermeasures and guide policy decisions. Hence, the outcome of this lane departure analysis should provide quantitative relationships between lane departure crash likelihood and explanatory factors so that agencies can estimate the benefits and costs of implementing countermeasures.

Given the likely audience, results presented in the form of "rumble strips reduce lane departures by 20%" or "drivers are four times more likely to be involved in a fatal/major injury crash on a two-lane roadway with 12-ft lanes and 6-ft gravel shoulders than on a two-lane roadway with 12-ft lanes and 2-ft paved shoulders" would be the most useful for comparing alternatives.

Consequently, analysis methods that provide crash reduction factors or odds ratios may be the most beneficial for providing specific information that can be used in assessing the costs and benefits of different designs or countermeasures. Highway engineers and policy makers have some familiarity with these types of analyses, and the results of these types of analysis can be communicated to the general public. However, it will only be possible to create crash reduction factors if sufficient crashes are available in the full-scale study.

The information in this report is geared toward those who will conduct or review lane departure analyses using the naturalistic driving study.

## Data Availability in Full-Scale Naturalistic Driving Study to Answer Lane Departure Research Questions

This section provides a brief discussion of the data expected to be available in the full-scale naturalistic driving study to answer lane departure research questions as they relate to the discussion on analytical methods in this section. Chapter 4 provides an in-depth review of the most recently available information and discusses the availability of data in the full-scale study to answer lane departure research questions. The accuracy, frequency of collection, and resolution that are expected to be necessary to address lane departure research questions is presented, and comments regarding the adequacy of the expected data collection are provided.

Dynamic driver and vehicle data are expected to be collected by the vehicle instrumentation DAS at 10 Hz (0.1-s intervals). Data will be reported at this level of resolution. Some data elements may be collected at a higher resolution (at a rate higher than 10 Hz) and will be aggregated to the 10-Hz level. Other data will be collected at a lower frequency or resolution and will be reported at 10 Hz.

Sensors available in the DAS that will monitor drivers include left-side, right-side, and head position driver video and a passive alcohol sensor. There has been some discussion about a "head position tracking system" being provided in the DAS. It is unknown at this time whether this will be available or whether head position tracking will be completed for all data or only for a subset of the data.

Dynamic vehicle factors from the DAS include forward/side radar; collection of vehicle kinematics (e.g., speed, acceleration, side acceleration); vehicle spatial position; and forward, side, and back video.

The final data set available to researchers is expected to consist of a spatial data set that contains individual vehicle/driver activity data at 10 Hz (1 row or frame per 0.1 s) and that information from other sources will be linked to that database and reported at the same level, even if those data are not collected at the same frequency. Static driver and vehicle variables may be either linked to the data set or provided in a relational database that can be joined to the spatial driver/vehicle data set.

Roadway data will be collected by the mobile mapping system (SHRP 2 Safety Project S04B) or will be available from existing state databases. The mobile mapping system will only collect data from a sample of roadways in a given study area. As a result, the same roadway data and the same data accuracy and resolution may not be available for all roadways. If the source of data is state databases, some differences will result across study areas.

Roadway data, when available, is expected to be linked to the vehicle data in the full-scale study using spatial overlay. Most of the roadway data will be collected at a lower rate than 10 Hz but will be reported with the final data set at that level. For instance, shoulder width may only be measured once per mile. If linked with the vehicle data, shoulder width will be included as a field reported for each 0.1 s, but the value would be consistent for all 0.1-s observations between each 1-mi sample interval.

Roadway variables that are not provided will need to be extracted from the outside vehicle imagery, aerial images, or other sources.

Two dynamic environmental factors will be included in the DAS: time and outside temperature. All other environmental factors will need to be extracted from the outside vehicle imagery or other sources.

Certain static characteristics, such as driver age, driver gender, vehicle type, and vehicle track width, will also be available. If they are not included as data fields in the continuous data, they can be linked and included as data fields in the continuous data.

No dynamic driver factors will be provided with the final data set except for readings from a passive alcohol sensor. As indicated, some head tracking information may be available. All other driver factors will have to be reduced from the driver video (e.g., distractions). Reduction of driver data at the

0.1-s interval would require a tremendous amount of resources. Therefore, data applications using continuous data would likely need to reduce driver data at a lower resolution (e.g., once per minute). This information can then be linked to the continuous data. Some automation can be used, but applications would need to be developed for this.

Driver factors relevant to driver distraction that would need to be reduced include the following: head position, which serves as a measure of eyeglance location; distractions (e.g., cell phone use, talking to passengers); hand position on steering wheel or other location; and measures of fatigue, such as head drooping or yawning.

## Data Segmentation Approaches

Modeling relies on obtaining the necessary data at the appropriate level of accuracy, frequency, and resolution. Data can be extracted in different ways depending on the application. Researchers for SHRP Safety Project S02, Integration of Analysis Methods and Development of Analysis Plan (Boyle et al., 2010), developed a model segmentation approach that can be applied to answering research questions for the full-scale naturalistic driving study data, as described below.

This approach is included in the analysis plan because it was decided that this was a useful structure for presenting the ways data were collected for the four analyses described in the following sections. The data segmentation is as follows:

- **Continuous (frame):** At this level, data are modeled at the rate at which they were collected, resulting in very large sample sizes. This is similar to the raw data set that would result from the instrumented vehicle DAS. The data will be quality assured, and some review of the data will be necessary.

  The instrumented vehicle data is expected to be collected at 10 Hz (0.1-s intervals). The term "continuous" is used, although in reality the data are discrete because they represent data aggregated to a set amount of time (i.e., data are aggregated to 0.1-s intervals). However, for all intents and purposes, the data can be considered as continuous.
- **Sequential blocks:** At this level, data are sampled and aggregated to blocks or epochs in which they are summarized over consecutive time periods. For instance, a 5-min sampling rate would indicate that data over each 5-min period are summarized into one observation. Data from different data fields can be aggregated over the block of time in different ways. For instance, the data for a particular field could be averaged, it could be summed, minimum and maximum values could be provided, or the number of times a particular value occurs could be reported. Data can be aggregated for any time period up to the trip level.
- **Sample based:** Data at this level are sampled at regular time intervals but are not aggregated. For instance, driver head pose may be sampled and reduced by the researchers every 2 min. Data at this level represents a snapshot in time.
- **Event:** Data at this level are aggregated for an incident (e.g., lane departure) or some other event of interest. Event data are aggregated for a set amount of time around an "event" to one observation per event (e.g., 30 s before the event start to 30 s after). An incident could be a crash, near crash, lane departure, and so forth. An example of an event of interest is vehicle activity in the vicinity of signalized intersections where one or more approaches have a posted speed limit of 50 mph or higher (high-speed signalized intersections). An event differs from a block in that it contains data only when an incident or event of interest has occurred.

Examples of data at the continuous, sequential block, and sample-based levels are shown in Figures 6.1 to 6.3. The variable speed is "evaluated" for a theoretical database. As shown in Figure 6.1, data at the continuous level are used at the rate at which they are reported in the naturalistic driving study. As a result, one observation is present for each 0.1-s of driving (one row). Figure 6.2 shows data collected at the sequential block level. A 1-min sampling interval was selected, and data were aggregated over each 1-min period. Each minute of data would provide one observation. An example of the sample-based approach is shown in Figure 6.3 for the same data set. Data are sampled at 1-min intervals. As a result, one row of data is extracted for each 1-min sample period. Speed would be reported for the 0.1-s interval extracted. One observation would be present for each 1-min period, but the data would reflect the 0.1-s intervals only.

## General Information About Data Reduction

This section provides general information about how the existing naturalistic driving study data were reduced for the analyses described in this chapter. Data for rural driving was requested from UMTRI for the road departure crash warning (RDCW) field operation test (FOT). A description of the data request and detailed description of the data received and other data sets used is provided in Chapter 3. A detailed description of the data reduction process is provided in Chapter 4.

Data were provided for subjects during the period when the RDCW system was functioning and recording data but not providing feedback to drivers. UMTRI provided data for rural roadways for 44 drivers. Vehicle activity data were provided in a Microsoft Access database and were provided as continuous data. Each row of data represented 0.1 s of vehicle driving for one driver during a trip. Forward imagery was

| Driver | Trip | Time | Radius (ft) | Shoulder Width (ft) | Ay (g) | Speed (mph) |
|---|---|---|---|---|---|---|
| 19 | 6 | 0.1 | tangent | 3.1 | 0.11 | 57.1 |
| 19 | 6 | 0.2 | tangent | 3.1 | 0.11 | 57.1 |
| 19 | 6 | 0.3 | tangent | 3.1 | 0.12 | 57.2 |
| 19 | 6 | 0.4 | tangent | 3.1 | 0.11 | 57.1 |
| 19 | 6 | 0.5 | tangent | 3.1 | 0.11 | 57.3 |
| 19 | 6 | 0.6 | tangent | 3.1 | 0.11 | 57.2 |
| 19 | 6 | 0.7 | tangent | 3.1 | 0.10 | 57.1 |
| 19 | 6 | 0.8 | tangent | 3.1 | 0.11 | 57.1 |
| 19 | 6 | 0.9 | tangent | 3.1 | 0.12 | 57.1 |
| 19 | 6 | 1.0 | tangent | 3.1 | 0.11 | 57.2 |
| 19 | 6 | 1.1 | tangent | 3.1 | 0.11 | 57.5 |
| 19 | 6 | 1.2 | tangent | 3.1 | 0.11 | 57.6 |
| 19 | 6 | 1.3 | tangent | 3.1 | 0.12 | 57.1 |
| 19 | 6 | 1.4 | tangent | 3.1 | 0.12 | 56.8 |
| 19 | 6 | 1.5 | tangent | 3.1 | 0.11 | 56.6 |
| 19 | 6 | 1.6 | tangent | 3.1 | 0.11 | 56.2 |
| 19 | 6 | 1.7 | tangent | 3.1 | 0.17 | 56.0 |
| 19 | 6 | 1.8 | tangent | 3.1 | 0.18 | 55.9 |
| 19 | 6 | 1.9 | tangent | 3.1 | 0.19 | 55.8 |
| 19 | 6 | 2.0 | 1000 | 3.1 | 0.21 | 55.4 |
| 19 | 6 | 2.1 | 1000 | 3.1 | 0.20 | 55.0 |
| 19 | 6 | 2.2 | 1000 | 3.1 | 0.19 | 49.9 |
| 19 | 6 | 2.3 | 1000 | 3.1 | 0.17 | 49.5 |
| 19 | 6 | 2.4 | 1000 | 3.1 | 0.15 | 49.0 |
| 19 | 6 | 2.5 | 1000 | 3.1 | 0.11 | 49.0 |
| 19 | 6 | 2.6 | 1000 | 3.1 | 0.11 | 49.0 |
| 19 | 6 | 2.7 | 1000 | 3.1 | 0.11 | 49.0 |
| 19 | 6 | 2.8 | 1000 | 3.1 | 0.11 | 49.1 |
| 19 | 6 | 2.9 | 1000 | 3.1 | 0.11 | 49.1 |
| 19 | 6 | 3.0 | 1000 | 3.1 | 0.11 | 49.2 |

***Figure 6.1. Example of data set showing data segmented at the continuous level.***

provided in most cases at 2 Hz (two rows per second, or one image per five rows of vehicle trace data). During times when the RDCW system alert reported that a lane departure may have occurred, forward imagery was provided at 10 Hz (10 rows per second, or 1 image per row of vehicle data trace data) for the 4 s before and 4 s after an alert was recorded.

Several variables used in the analysis were provided with the data set: driver number, trip number, time since start of trip, driver age, driver gender, vehicle spatial position, heading, brake on or off, cruise control on or off, vehicle offset from center of lane, lane width, vehicle track width, speed, lateral speed, lateral acceleration, side acceleration, yaw rate, roll rate, pitch rate, wiper status, headlamp status, road type, posted speed limit, advisory speed, AADT, and number of thru lanes. In some cases, advisory speed and posted speed limit were not included and had to be obtained from the forward imagery.

A number of variables that were not provided in the UMTRI data could be extracted or created from either the UMTRI data or from other available data sources. Other data sources included aerial imagery, a roadway database, and a crash database for Michigan. (The databases are described in Chapter 3; extraction of data elements is discussed in Appendix A.)

Because large amounts of data were provided and data reduction became a time-consuming task, it was decided to focus on rural, two-lane, paved roadways. Only paved roadways were considered because the lane tracking system did not function well on unpaved surfaces, such as gravel.

In order to determine what other variables should be extracted, the team conducted a comprehensive literature review and compiled a list of potential variables that have been shown to affect the likelihood and severity of lane departure crashes (see Chapter 2).

| Driver | Trip | Time | Radius (ft) | Shoulder Width (ft) | Ay (g) | Speed (mph) |
|---|---|---|---|---|---|---|
| 19 | 6 | 0.1 | tangent | 3.1 | 0.11 | 57.1 |
| 19 | 6 | 0.2 | tangent | 3.1 | 0.11 | 57.1 |
| 19 | 6 | 0.3 | tangent | 3.1 | 0.12 | 57.2 |
| 19 | 6 | 0.4 | tangent | 3.1 | 0.11 | 57.1 |
| 19 | 6 | 0.5 | tangent | 3.1 | 0.11 | 57.3 |
| 19 | 6 | 0.6 | tangent | 3.1 | 0.11 | 57.2 |
| 19 | 6 | 0.7 | tangent | 3.1 | 0.10 | 57.1 |
| 19 | 6 | 0.8 | tangent | 3.1 | 0.11 | 57.1 |
| 19 | 6 | 0.9 | tangent | 3.1 | 0.12 | 57.1 |
| 19 | 6 | 1.0 | tangent | 3.1 | 0.11 | 57.2 |
| 19 | 6 | 1.1 | tangent | 3.1 | 0.11 | 57.5 |
| 19 | 6 | 1.2 | tangent | 3.1 | 0.11 | 57.6 |
| 19 | 6 | 1.3 | tangent | 3.1 | 0.12 | 57.1 |
| 19 | 6 | 1.4 | tangent | 3.1 | 0.12 | 56.8 |
| 19 | 6 | 1.5 | tangent | 3.1 | 0.11 | 56.6 |
| 19 | 6 | 1.6 | tangent | 3.1 | 0.11 | 56.2 |
| 19 | 6 | 1.7 | tangent | 3.1 | 0.17 | 56.0 |
| 19 | 6 | 1.8 | tangent | 3.1 | 0.18 | 55.9 |
| 19 | 6 | 1.9 | tangent | 3.1 | 0.19 | 55.8 |
| 19 | 6 | 2.0 | 1000 | 3.1 | 0.21 | 55.4 |
| 19 | 6 | 2.1 | 1000 | 3.1 | 0.20 | 55.0 |
| 19 | 6 | 2.2 | 1000 | 3.1 | 0.19 | 49.9 |
| 19 | 6 | 2.3 | 1000 | 3.1 | 0.17 | 49.5 |
| 19 | 6 | 2.4 | 1000 | 3.1 | 0.15 | 49.0 |
| 19 | 6 | 2.5 | 1000 | 3.1 | 0.11 | 49.0 |
| 19 | 6 | 2.6 | 1000 | 3.1 | 0.11 | 49.0 |
| 19 | 6 | 2.7 | 1000 | 3.1 | 0.11 | 49.0 |
| 19 | 6 | 2.8 | 1000 | 3.1 | 0.11 | 49.1 |
| 19 | 6 | 2.9 | 1000 | 3.1 | 0.11 | 49.1 |
| 19 | 6 | 3.0 | 1000 | 3.1 | 0.11 | 49.2 |

Avg speed: 57.2 mph

Avg speed: 56.5 mph

Avg speed: 50.4 mph

***Figure 6.2. Example of data set showing data segmented at the sequential block level.***

All of the data elements that the team determined were important from the literature and could be obtained from one of the available databases (vehicle data, aerial imagery, roadway data, forward imagery, and crash database) were extracted. In several cases, data were obtained from the merging of two or more databases. For instance, curve radius and direction were determined by overlaying the vehicle database with aerial imagery and determining the start and end point in the vehicle data that corresponded to each curve, while curve radius was measured using the aerial imagery.

The original continuous vehicle activity data from UMTRI were provided in a database with each row representing 0.1 s of activity for a particular driver/vehicle. When other variables were extracted from the various data sets, they were linked to the continuous data even if they were extracted at a lower resolution. For instance, shoulder width was determined for a homogenous roadway section. All vehicle activity along that section would have been selected, and a data field "ShldWidth" would be populated with the single measurement for shoulder width.

A summary of the variables used in the different analyses is provided in Tables 6.1 and 6.2. A number of other variables were extracted, such as type of curve advisory signing and visibility, but were not included in the analyses because of low sample sizes.

Lane departures were determined by calculating vehicle wheel path using vehicle offset, lane width, and track width, as described in Appendix A. A lane departure was defined as a vehicle wheel path crossing over the right (right-side lane departure) or left (left-side lane departure) lane line and encroaching upon either the shoulder or the adjacent lane by 0.1 m or more. The threshold 0.1 m was used as a buffer because there is some uncertainty in estimation of wheel path. In all cases, the vehicle departed the lane and then returned to the initial lane of travel without losing

| Driver | Trip | Time | Radius (ft) | Shoulder Width (ft) | Ay (g) | Speed (mph) |
|---|---|---|---|---|---|---|
| 19 | 6 | 0.1 | tangent | 3.1 | 0.11 | 57.1 |
| 19 | 6 | 0.2 | tangent | 3.1 | 0.11 | 57.1 |
| 19 | 6 | 0.3 | tangent | 3.1 | 0.12 | 57.2 |
| 19 | 6 | 0.4 | tangent | 3.1 | 0.11 | 57.1 |
| 19 | 6 | 0.5 | tangent | 3.1 | 0.11 | 57.3 |
| 19 | 6 | 0.6 | tangent | 3.1 | 0.11 | 57.2 |
| 19 | 6 | 0.7 | tangent | 3.1 | 0.10 | 57.1 |
| 19 | 6 | 0.8 | tangent | 3.1 | 0.11 | 57.1 |
| 19 | 6 | 0.9 | tangent | 3.1 | 0.12 | 57.1 |
| 19 | 6 | 1.0 | tangent | 3.1 | 0.11 | 57.2 |
| 19 | 6 | 1.1 | tangent | 3.1 | 0.11 | 57.5 |
| 19 | 6 | 1.2 | tangent | 3.1 | 0.11 | 57.6 |
| 19 | 6 | 1.3 | tangent | 3.1 | 0.12 | 57.1 |
| 19 | 6 | 1.4 | tangent | 3.1 | 0.12 | 56.8 |
| 19 | 6 | 1.5 | tangent | 3.1 | 0.11 | 56.6 |
| 19 | 6 | 1.6 | tangent | 3.1 | 0.11 | 56.2 |
| 19 | 6 | 1.7 | tangent | 3.1 | 0.17 | 56.0 |
| 19 | 6 | 1.8 | tangent | 3.1 | 0.18 | 55.9 |
| 19 | 6 | 1.9 | tangent | 3.1 | 0.19 | 55.8 |
| 19 | 6 | 2.0 | 1000 | 3.1 | 0.21 | 55.4 |
| 19 | 6 | 2.1 | 1000 | 3.1 | 0.20 | 55.0 |
| 19 | 6 | 2.2 | 1000 | 3.1 | 0.19 | 49.9 |
| 19 | 6 | 2.3 | 1000 | 3.1 | 0.17 | 49.5 |
| 19 | 6 | 2.4 | 1000 | 3.1 | 0.15 | 49.0 |
| 19 | 6 | 2.5 | 1000 | 3.1 | 0.11 | 49.0 |
| 19 | 6 | 2.6 | 1000 | 3.1 | 0.11 | 49.0 |
| 19 | 6 | 2.7 | 1000 | 3.1 | 0.11 | 49.0 |
| 19 | 6 | 2.8 | 1000 | 3.1 | 0.11 | 49.1 |
| 19 | 6 | 2.9 | 1000 | 3.1 | 0.11 | 49.1 |
| 19 | 6 | 3.0 | 1000 | 3.1 | 0.11 | 49.2 |

row speed: 57.1 mph

row speed: 57.5 mph

row speed: 55.0 mph

***Figure 6.3. Example of data set showing data segmented at the sample level.***

control or making sudden evasive maneuvers. This type of lane departure was referred to as an encroachment in the discussion on crash surrogates in Chapter 5. The UMTRI data set did not provide any near crash or crashes. It should be noted that some of the left-side lane departures may have been cases of drivers intentionally "cutting the curve." It may be possible to ascertain this from the driver's face video and from the driver's hand position on the steering wheel. However, the team did not have access to this type of information.

The data reduction resulted in a total of 22 right-side lane departure and 51 left-side lane departure events for two-lane rural roads. It also resulted in over 113,000 observations (0.1-s data frames) of normal driving.

Data for which lane departure incidents occurred were modeled as continuous data in the data mining and time series analysis approaches and were summarized by event for the approach using an odds ratio. In this case, data for a block of time around left- or right-side lane departures were summarized as an "event." The start point for each lane departure was determined by identifying the point at which the vehicle began deviating from its path toward the edge of the lane, as shown in Figure 6.4. The end point of the event was the point after the vehicle returned to the roadway and corrected its path. The start and end times were noted at those points, and the continuous data for each event were extracted. A lane departure event included time spent drifting from the roadway or lane, time off the roadway or lane, and time returning to the original lane of travel. The average lane departure was approximately 8.0 s (80 instances of 0.1-s observations). Depending on the amount of time included, each event was weighted accordingly.

Data for which no lane departure had occurred were used to represent normal driving data. Shankar et al. (2008) referred to exposure measures as "controls." Events are situations of interest (crash, near crash), and controls are situations where the outcome is absent (normal driving). Risk can be determined by dividing the number of events by total exposure ("control") for a cohort.

Normal driving data were used as continuous data for the data mining and times series analyses and were summarized

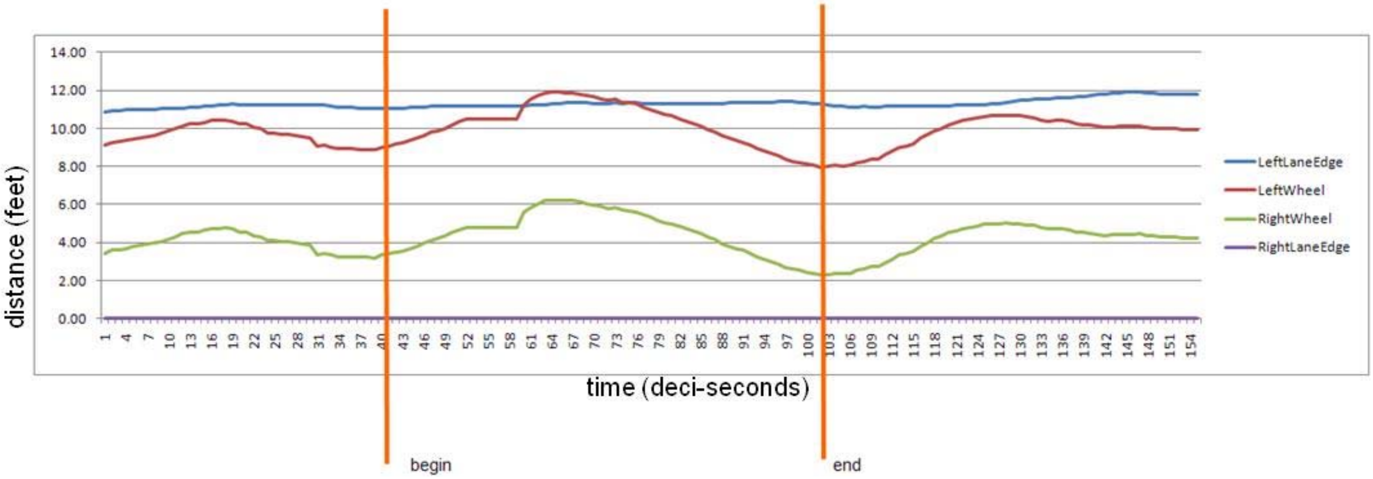
*Table 6.1. Description of Driver, Vehicle, and Environmental Variables*

| Variable | Source | Description | Variable Type |
|---|---|---|---|
| **Driver Variables** | | | |
| Age | Provided with data set | Age of driver | Numeric |
| Gender | | 1 = male, 2 = female | Categorical |
| OvrSpd5 | Calculated from speed and posted speed limit | Fraction of time driver exceeds the posted speed limit by 5 mph on rural, two-lane roads | Numeric |
| OvrSpd10 | | Fraction of time driver exceeds the posted speed limit by 10 mph on rural, two-lane roads | Numeric |
| OvrAdvSpd5 | Calculated from speed and posted speed limit | Fraction of time driver exceeds the advisory curve speed by 5 mph on rural, two-lane roads | Numeric |
| OvrAdvSpd10 | | Fraction of time driver exceeds the advisory curve speed by 10 mph on rural, two-lane roads | Numeric |
| **Vehicle Variables** | | | |
| Spd | Provided with data set | Vehicle forward speed (m/s) | Numeric |
| LatSpd | | Vehicle side speed (m/s) | Numeric |
| $A_x$ | | Forward acceleration (m/s$^2$) | Numeric |
| $A_y$ | | Side acceleration (m/s$^2$) | Numeric |
| RollRate | | Rate of roll (deg/s) | Numeric |
| PitchRate | | Pitch rate (deg/s) | Numeric |
| YawRate | | Rate of yaw (deg/s) | Numeric |
| Following | Extracted from forward video | Subjective measure of vehicle following<br>0: Not following<br>1: Following<br>2: Following closely | Categorical |
| **Environmental Variables** | | | |
| Time | Extracted from forward video and time | Indicates time of day<br>0: Day<br>1: Dawn/dusk/night<br>There was no overhead lighting on any of the roadways, so all nighttime driving was dark/unlighted. | Categorical |
| EnvCond | Extracted from forward video | Prevailing atmospheric conditions<br>0: Clear (no precipitation)<br>1: Light to moderate rain<br>2: Heavy rain<br>3: Light to moderate snow<br>4: Heavy snow<br>5: Fog | Categorical |
| RoadSurf | Extracted from forward video and wiper status | 0: Dry<br>1: Wet<br>There was no snow on any of the roadways. | Categorical |

*Table 6.2. Description of Roadway and Other Variables*

| Variable | Source | Description | Variable Type |
|---|---|---|---|
| **Roadway Variables** | | | |
| Radius | Extracted from aerial imagery | Curve radius in m | Numeric |
| CurveType | Extracted from forward video | Direction of curve from perspective of driver<br>0: No curve<br>1: Right curve<br>2: Left curve | Categorical |
| LaneWidth | Provided | Lane width in m | Numeric |
| ShldWidth | Extracted from forward video | Shoulder width in m | Numeric |
| ShldType | | Type of shoulder present<br>1: Paved<br>3: Gravel<br>4: Earth<br>6: No shoulder<br>7: Partially paved | Categorical |
| PvMCond | | Pavement marking condition<br>0: Highly visible<br>1: Visible<br>2: Obscure | Categorical |
| DwyDen | | Density of driveways to the right (driveways/m) | Numeric |
| **Other Variables** | | | |
| AADT | Provided | Annual average daily traffic for roadway segment in vehicles per day | Numeric |
| OnDen | Extracted from forward video | Density of on-coming vehicle (vehicles/m) | Numeric |
| Conflict | Extracted from forward video and vehicle data | Indicates type of vehicle event<br>11: Normal driving<br>21: Right-side lane departure<br>31: Left-side lane departure | Categorical |
| Angle | | Angle that vehicle left roadway during departure | Numeric |
| MaxOff | Extracted from vehicle data | Maximum distance vehicle encroached into adjacent lane or shoulder during lane departure | Numeric |
| CrshDen | Extracted from Michigan crash database and aerial imagery | Density of lane departure crashes along roadway segment (crashes/m) | Numeric |

into epochs, which are similar to events for the odds ratio analysis. Epochs were selected by driver and trip when roadway and environmental conditions were consistent. When a change in roadway occurred, a new epoch was created. For instance, data for a driver traveling along a specific roadway during a particular trip would be partitioned each time roadway conditions changed. Data along a tangent section would be marked as one epoch if the roadway cross section did not change. When the vehicle encountered a curve, a new epoch would be created that contained all of the vehicle activity on the curve. At the end of the curve, a new epoch would be created for the next tangent section. Data could not be partitioned by driver characteristics because dynamic driver characteristics were not available and static driver variables such as age and gender did not change. In most cases, environmental conditions were consistent across a roadway section, so it was not necessary to




Source: UMTRI RDCW data set.

***Figure 6.4. Begin and end point for event.***

consider changes in environmental conditions. Data were summarized for each epoch. The length of each epoch was different because drivers spent different amounts of time driving on a particular type of roadway. The number of 0.1-s intervals for each epoch was included as a weighting factor.

Information about normal driving is useful because it can be used to represent exposure. One of the strengths of the naturalistic driving studies is that a substantial amount of normal driving will be available, which can be used to determine a driver's exposure for a particular set of circumstances. Currently, there is no realistic method to obtain exposure data for an individual driver, and it is even more difficult to obtain detailed exposure for a cohort of drivers.

The most common measures to calculate exposure for a driver cohort is to use number of licensed drivers partitioned by age or some other characteristic. However, the use of number of licensed drivers assumes that all drivers drive an equal number of miles and may overestimate or underestimate involvement if the driver group has different travel trends. For instance, older drivers may drive substantially less than drivers in other age groups. VMT by age group is a better measure because it demonstrates actual exposure, but it is difficult to obtain on a local or even state level. National studies, such as the National Personal Transportation Study (NPTS), have developed VMT fractions by age group, but national statistics may not be representative of state and local areas.

Actual VMT by age group can be extracted from naturalistic driving study data for a given set of conditions. This will provide a unique opportunity to study risk by driver sub-population groups. Using the naturalistic driving study, the amount of driving a driver or group of drivers engages in on a particular roadway type can be used as a measure of exposure.

## Analysis Approach 1: Data Mining

Three different analysis approaches were used to model the UMTRI data. The first was a data mining approach, described in this section.

### Description

Data mining is the process of analyzing data to uncover patterns and establish relationships. Data mining processes may include the following (Search SQL Server, 2009):

- Association, which involves looking for patterns consisting of events that are connected to each other;
- Sequencing, which involves looking for patterns consisting of events where one event leads to another;
- Classification, which involves looking for new patterns;
- Clustering, which involves organizing groups of facts; and
- Forecasting, which involves looking for patterns that can be used to make predictions.

Data mining is the exploration and analysis of large amounts of data to discover meaningful patterns and rules in the data that are not evident. The process can be automated or semiautomated (Collier et al., 1998). The discovery of patterns leads to additional knowledge. Data mining is useful for large data sets where patterns cannot easily be uncovered by human analysts. It also allows analysis of data that may never have been analyzed using other techniques. It can be used for both prediction and description (Tan et al., 2006).

## Sampling Approach

A sample-based approach using a sampling interval of 0.1 s was used to model the data. As a result, every 10th observation (0.1-s frame) was selected. The sample included both normal data and left- and right-side lane departures.

## Response Variables

Two models were developed, one with a response variable for right-side lane departures and the other with a response variable for left-side lane departures.

## Explanatory Variables

All of the driver, roadway, and environmental variables in Tables 6.1 and 6.2 were evaluated in different combinations. Variables that were expected to be correlated were not evaluated at the same time.

## Modeling Approach and Results

### *Description of Classification and Regression Tree Model*

A classification and regression tree model was the data mining modeling approach selected. Classification methods assign objects to predefined categories (Tan et al., 2006). Tree-based models are used for both classification and regression. A tree-based analysis uses a response variable ($Y$) that can be either quantitative or qualitative and a set of classification or predictor variables ($X_i$) that may be a mixture of ordinal or nominal variables. For classification trees the response is categorical, and for regression trees the dependent variable is quantitative (Nagpual, 2009). Classification and regression trees use algorithms to determine a set of if-then logical split conditions that divide the data into subsets. One of the advantages of regression tree analysis over traditional regression analysis is that it is a nonparametric method that does not require assumptions of a particular distribution and is more resistant to the effects of outliers; splits usually occur at nonoutlier values. Tree models are nonlinear, indicating that there is no assumption about the underlying relationships between the response and explanatory variables. In addition, independent variables do not have to be specified in advance. A regression tree selects only the most important independent variables and the values of those variables that result in the maximum reduction in deviance. Another advantage is that results are invariant with respect to monotone transformations of the independent variables. Thus, the researcher does not have to test a number of transformations to find the "best" fit (Roberts et al., 1999). The regression tree also allows relationships between variables to be uncovered that may not be determined using other methods (StatSoft, 2008). For instance, shoulder width may be relevant in determining whether a right-side lane departure results in a lane departure crash on curves of a certain radius but may not be relevant for tangent sections or curves with larger radii.

S-PLUS Statistical Software's (Version 8.0.4) classification and regression tree analysis was used to evaluate the data. Regression tree rules are determined by a procedure known as recursive partitioning, which iteratively generates a tree structure by splitting the sample data set into two subsets according to two rules. First, the independent variable that produces the maximum reduction in variability is identified. Next, the value of the variable that results in the maximum reduction in variability is selected (Wolf et al., 1998).

Figure 6.5 shows an example of a classification and regression tree analysis that was used to determine the factors related to high accelerations at intersections in a model to predict vehicle activity for emissions modeling (Hallmark et al., 2002). As indicated, the tree split on the variables queue position, approach grade, and distance to the nearest downstream signalized intersection. A vehicle that was on a segment with a downstream distance of less than 902 ft, that was in a queue position less than 2 (first in queue), and that was on approach with a grade of −1% or lower had an acceleration of 10.85 ft/s$^2$ starting up from the intersection stop line. As shown, grade was only relevant for vehicles in queue positions 1, 2, and 3, and distance to the nearest downstream signalized intersection was only relevant for vehicles in queue position 1.

In growing a regression tree, the binary partitioning algorithm recursively splits the data in each node until the node is homogenous or until a minimum criterion such as number of observations is met. If left unconstrained, a regression tree can grow until it results in a complex model with a single observation at each terminal node that explains all the deviance.

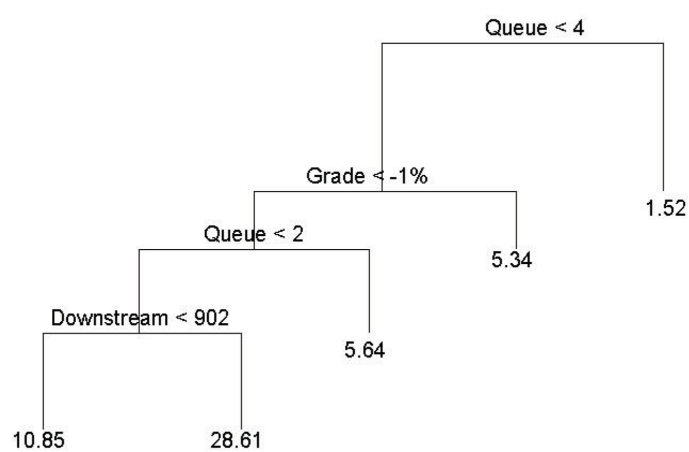



***Figure 6.5. Example of a classification and regression tree used to model vehicle acceleration for emissions modeling.***

However, for application purposes, it is desirable to create an end product that balances the model's ability to explain the maximum amount of deviation with a simpler model that is easy to interpret and apply. To simplify the final model, the user can set values such as the minimum number of observations present before a split occurs or minimum deviance allowed at each node. Default values may also be used. Three other functions in S-PLUS can be used to simplify the tree, as described below.

Pruning reduces the nodes on a tree by successively snipping off the least important splits. The equation to determine the importance of a subtree using a cost-complexity measure is as follows (Insightful Corporation, 2007):

$$D_k(T') = D(T') + k \cdot size(T') \qquad (6.1)$$

where

$D_k(T')$ is the deviance of the subtree $T'$,
$k$ is the cost-complexity parameter, and
$size(T')$ is the number of terminal nodes of $T'$.

Cost complexity pruning selects the subtree $T'$ which minimizes $D_k(T')$ over all subtrees.

The second function that can be used to simplify the model is shrinking, which reduces the number of effective nodes. This is accomplished by shrinking the fitted value of each node toward its parent node using the following algorithm (Insightful Corporation, 2007):

$$\hat{y}(node) = k \cdot \mathrm{N}(node) + (1-k) \cdot \hat{y}(parent) \qquad (6.2)$$

where

$k$ is the shrinking parameter ($k$ may be a scalar or a vector, $0 < k < 1$),
$\mathrm{N}$ ($node$) is the usual fitted value for a node, and
$\hat{y}(parent)$ is the shrunken fitted value for the node's parent.

Snipping allows the user to interactively remove nodes and try various modifications to the original model. The effects of using any of the procedures (pruning, shrinking, snipping, modifying the minimum number of observations, modifying the minimum node size, or modifying the minimum node deviance) can be evaluated by observing normal probability plots of the residuals for the tree object, comparing residual mean deviance for different models, or inspecting a plot of the reduction in deviance with the addition of nodes. The residual mean deviance (rmd) is an indicator of regression tree fit and is the statistic reported rather than the traditional $r^2$ value in linear regression analysis. The rmd is the mean deviance of the data samples in the terminal nodes of an estimated tree model. A lower value for rmd indicates a better fit (Roberts et al., 1999).

### *Analysis Approach*

Classification and regression tree analysis methods were used to identify variables with the most explanatory power in influencing the occurrence of a right- or left-side lane departure. A separate model was created for right- and left-side departures.

For each model, a variety of explanatory variables were evaluated in different combinations. Variables that did not appear in one of the main branches of the regression tree being evaluated were removed, and other combinations of variables were evaluated. Initial models resulted in complex trees. For example, the initial left-side lane departure model is shown in Figure 6.6. As indicated, the tree model is complex. S-PLUS plots the tree structure so that the more important the parent split, the farther the children node pairs are from the parents. This information and a plot of the deviance as a function of the number of nodes and cost-complexity parameters were used to evaluate the most relevant splits. The snip and prune tree functions in S-PLUS were used to develop the final models shown in Figures 6.7 and 6.8.

### *Results*

As shown in Figure 6.7, the most relevant explanatory variables for left-side lane departures were radius of curve, driver age, and shoulder type. The values at the end of the tree nodes indicate the type of lane departure. The value 11 was used for normal driving, and the value 31 was used to indicate that a left-side lane departure had occurred. The numbers showed trends only: the higher the node value, the more likely a left-side lane departure would occur. The values do not correspond to an actual probability and are an artifact of the model used to develop the regression trees. With significantly more data, the model could have been developed so that the probability of lane departure was the node value. Because there was not a substantial amount of data, the tree should only be interpreted to show a general pattern and break points for variables where relationships are emerging. For instance, for the left-side lane departure, age was relevant when curve radius was less that 1,081 ft but was not relevant for curve radii greater than this.

As indicated in Figure 6.8, the most relevant explanatory variables for right-side lane departures were also radius of curve, driver age, and shoulder type. The values at the ends of the tree nodes indicate the type of lane departure. The value 11 was used for normal driving, and the value 21 was used to indicate that a right-side lane departure had occurred. The higher the node value, the more likely a right-side lane departure would occur.

## Estimating Sample Size for the Full-Scale Study

As indicated, only limited data were available to evaluate the lane departure research questions. In this section, sample size

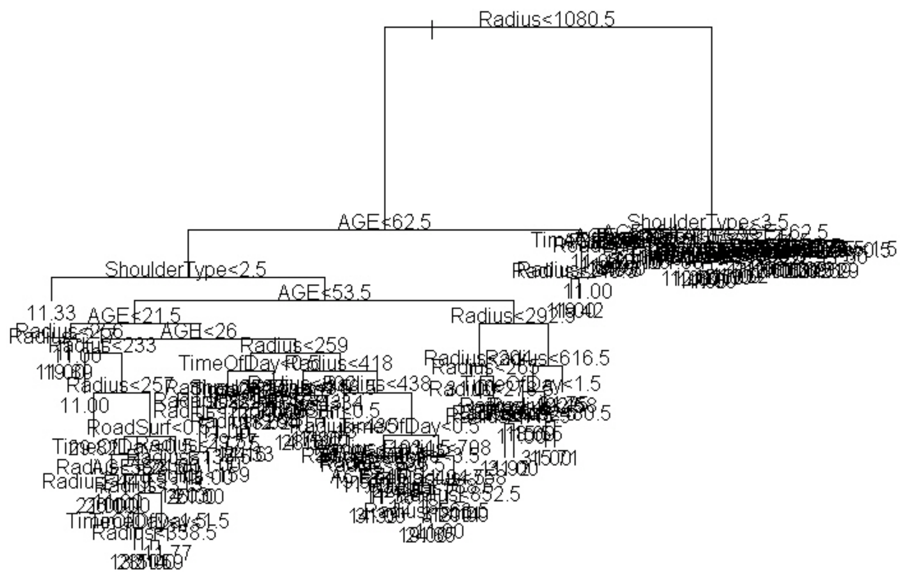



*Figure 6.6. Initial tree model for left-side lane departures.*

for the full-scale study is addressed using classification and regression trees.

Appropriate sample size for classification and regression trees is not easily determined. Sample size depends on factors such as number of variables, deviance at each node, complexity of the model, and minimum specified node size. Most model packages set some minimum default node size. In S-PLUS, the default node size is five observations.

Morgan et al. (2009) evaluated methods to test sample size for decision-tree analysis. The authors indicate that when data sets are too large, decision trees may overfit. They also found that accuracy decreases as sample size increases. From the data they evaluated, they found that relatively stable patterns emerged between 8,000 and 16,000 samples with models that had a large number of variables to evaluate. The authors also describe other work to evaluate sample size.

## Application to Full-Scale Study

A sample-based approach is expected to be the best data sampling method for classification and regression tree analysis. Use of continuous data would require reduction of a

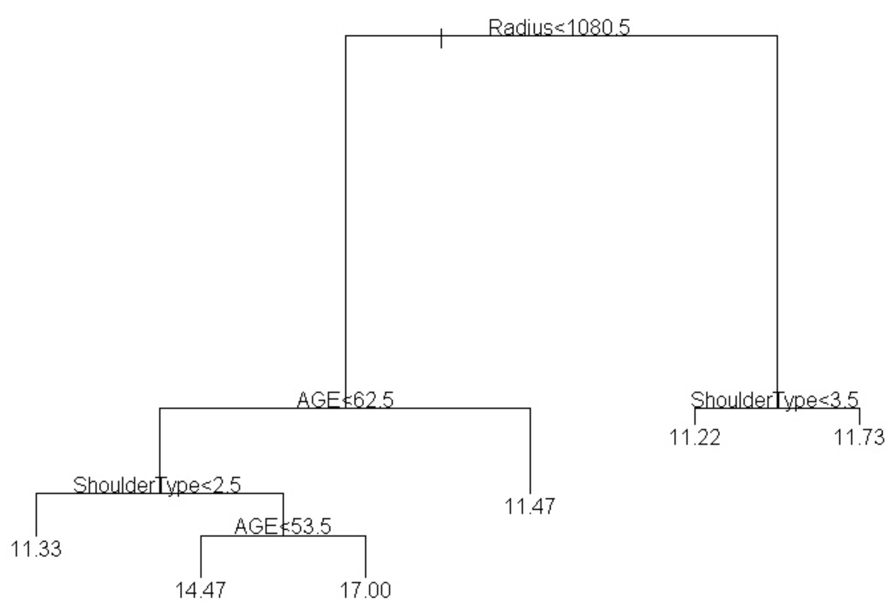



*Figure 6.7. Final tree model for left-side lane departures.*

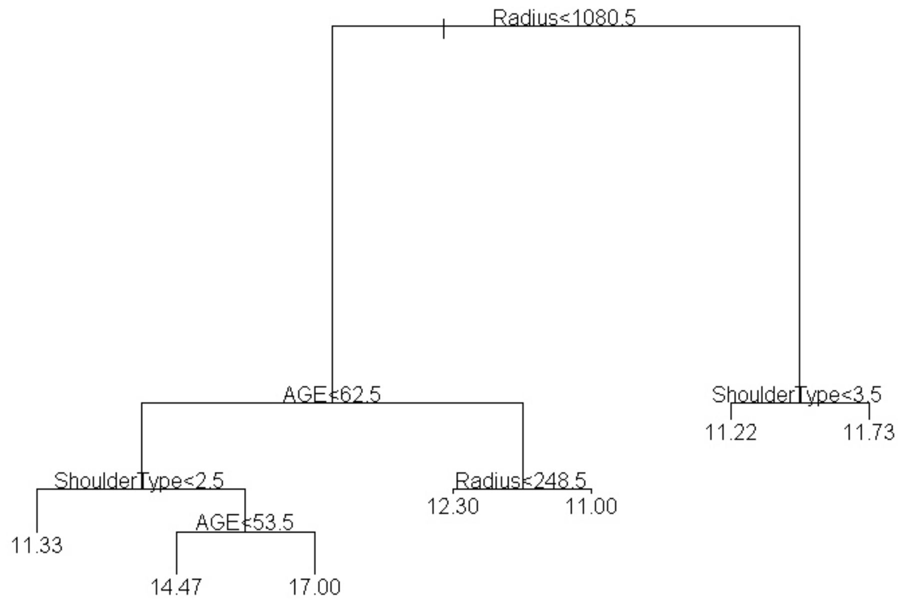



***Figure 6.8. Final tree model for right-side lane departures.***

large amount of data, which would be extremely resource-intensive.

The main advantage to data mining is that it will be useful in uncovering relationships in the data that may not be found using other methods. Additionally, data mining can also evaluate a large amount of data using an automated process.

One disadvantage is that this modeling approach is not common among practitioners. It will be necessary to interpret results so that practitioners can incorporate the information into decision-making models, such as comparing the costs and benefits of a particular countermeasure.

## Analysis Approach 2: Odds Ratio and Logistic Regression

The second analysis approach was to calculate odds ratios, as described in this section.

### Description

In this approach, both a simple odds ratio test and logistic regression were used to identify factors related to lane departures. An odds ratio compares the probability of an event happening with the probability of the same event not happening. Logistic regression evaluates the association between a binary response and explanatory variables. The natural logarithm of the odds is related to explanatory variables using a linear model.

The difference between the approaches used here and the case-control used later is in the assumptions that are made for modeling. The odds ratio and the logistic regression approaches used here assume random independent sampling from a specific event (either left-lane-departure or right-lane-departure) and random independent sampling of normal driving epochs.

Each epoch was treated as an individual observation. Since the same driver may have been represented in more than one epoch, correlations between epochs may have existed. However, since this was an exploratory analysis, these two approaches used all available epochs and made the assumption of independence for simplicity. In future analyses, when larger data sets are available, researchers could test if the correlation is zero. If the correlation is not zero, several adjustments could be considered. For example, researchers could model the correlation structure or use the paired case-control approach.

### Data Sampling Approach

Data were reduced as described in the previous section on data mining (p. 84). Each lane departure event or normal driving epoch was modeled as one observation. However, the models were weighted by the number of 0.1-s intervals for each event or epoch.

### Response Variables

Occurrence of a lane departure was the response variable. Right-side and left-side lane departures were modeled separately.

### Explanatory Variables

A number of explanatory variables were available, as shown in Tables 6.1 and 6.2. When variables were highly likely to be correlated, only one variable was evaluated. For instance, ambient conditions and roadway surface condition are highly correlated. Road surface condition was used because it is more likely to have an impact on whether a driver has a lane departure. Variables were not included in the simple odds ratio when there were not enough observations to calculate an odds ratio.

## Simple Odds Ratio Modeling Approach and Results

Simple odds ratios were calculated using Equation 6.3:

$$\mathrm{OR} = \frac{\mathrm{RD}_j / \mathrm{RD}_k}{\mathrm{ND}_j / \mathrm{ND}_k} \tag{6.3}$$

where

- OR = odds ratio,
- $\mathrm{RD}_j$ = number of observations for situation $j$ where lane departure occurs,
- $\mathrm{RD}_k$ = number of observations for situation $k$ where lane departure occurs,
- $\mathrm{ND}_j$ = number of observations for situation $j$ where no lane departure occurs, and
- $\mathrm{ND}_k$ = number of observations for situation $k$ where no lane departure occurs.

The 95% confidence interval was calculated using Equations 6.4, 6.5, and 6.6:

$$\text{CI of OR is } \exp\left(\log(\mathrm{OR}) \pm 1.96 * \mathrm{sd}\right) \tag{6.4}$$

and

$$\text{standard deviation of } \log(\text{odds ratio}) = \left(1/A + 1/B + 1/C + 1/D\right)^{0.5} \tag{6.5}$$

$$\text{standard deviation of } \log(\text{odds ratio}) = \left(1/\mathrm{RD}_j + 1/\mathrm{RD}_k + 1/\mathrm{ND}_j + 1/\mathrm{ND}_k\right)^{0.5} \tag{6.6}$$

The simple odds ratio only allows for two responses within a variable (e.g., rumble strips present or not). Therefore, when a variable had several responses, an odds ratio was calculated for each response if there were sufficient values. For instance, curve type had three responses: no curve, left-hand curve, and right-hand curve. As a result, presence of left-hand and right-hand curves was compared against tangent sections.

The results of this approach are presented in Table 6.3. Categories were created for numeric variables such as radius, as shown in Table 6.3, to create two responses. When numeric variables could not easily be combined into categories, they were not included.

As indicated, radius of curvature was highly relevant in the occurrence of right- and left-side departures. Left-side and right-side lane departures were 10.9 times and 29.2 times more likely to occur on curves with a very small radius (less than 200 m) than on a tangent section. Lane departures were also much more likely to occur on other curve radii as shown. Curve direction from the perspective of the driver (left curve vs. right curve) was also relevant in determining the occurrence of both right- and left-side lane departures. The odds of a left-side lane departure on a left-hand curve were 5.1 greater than on a tangent section, and for a right-hand curve the odds were 2.95 greater. The odds of having a right-side lane departure

***Table 6.3. Results of Simple Odds Ratio***

| | Left-Side Departure vs. Normal | | Right-Side Departure vs. Normal | |
|---|---|---|---|---|
| **Variable** | **Odds Ratio** | **Confidence Interval** | **Odds Ratio** | **Confidence Interval** |
| Radius < 200 m vs. tangent | 10.9 | (9.7, 12.3) | 29.2 | (25.4, 33.5) |
| 400 m > radius ≥ 200 m vs. tangent | 32.8 | (30.1, 35.9) | 10.9 | (9.6, 12.4) |
| 600 m > radius ≥ 400 m vs. tangent | 19.7 | (17.8, 21.8) | 22.1 | (18.8, 25.9) |
| 600 m ≥ radius vs. tangent | 20.4 | (18.4, 22.7) | 13.6 | (11.3, 16.4) |
| Left-hand curve vs. tangent | 5.1 | (4.78, 5.54) | 3.84 | (3.37, 4.38) |
| Right-hand curve vs. tangent | 2.95 | (2.71, 3.20) | 6.60 | (5.93, 7.34) |
| Wet vs. dry roadway | 0.97 | (0.9, 1.1) | Not enough samples | |
| Day vs. night/dusk | 1.8 | (1.7, 1.9) | 0.38 | (0.3, 0.4) |
| Male vs. female | 1.28 | (1.19, 1.38) | 1.14 | (1.02, 1.28) |
| Gravel vs. paved/partially paved | 9.83 | (8.37, 11.54) | 0.16 | (0.14, 0.18) |
| Earth vs. paved/partially paved | 5.36 | (4.55, 6.32) | 0.07 | (0.06, 0.08) |

were 3.8 for left-hand curves and 6.6 for right-hand curves. Weather and time of day did not appear to be relevant because the odds ratio was close to 1.0. Men were slightly more likely than women to be involved in both types of lane departures (1.3 for left-side lane departure, 1.1 for right-side lane departure). Shoulder type appeared to be relevant for left-side but not right-side lane departures. It should be noted that a left-side lane departure on a curve can be intentional (cutting the curve). It was not possible to distinguish between intentional and unintentional lane departures, except for events like a vehicle changing lanes to avoid a parked car or object in the roadway. These lane departures were removed but others, as indicated, could not be identified.

## Logistic Regression Modeling Approach and Results

Multivariate logistic regression was used to examine factors associated with the risk of both left- and right-side lane departures using data summarized from the UMTRI data set, as described in the previous section on data mining (p. 84). In each model, the records for lane departures were used as the cases, while records without lane departures (normal driving) were used as the controls. Separate models were created for left- and right-side lane departures. Both models were created using the LOGISTIC procedure in the SAS/STAT 9.2 software package.

The response variable was presence of a lane departure, $Z$, given as 0 if there is no lane departure (normal driving) and 1 if a lane departure occurred.

The models for right- and left-side lane departures were created using the following logic. Occurrence of a lane departure $Z$ is the response variable. $Z$ is a Bernoulli variable with $p = P(Z = 1)$ as the probability of occurrence of a lane departure. Therefore, $p/(1 - p)$ is the odds of a lane departure happening. In order to link the odds of a lane departure to the explanatory variables investigated ($X$'s), the logit link function was used. Hence, a connection between the probability of a lane departure and the linear combination of predictor variables ($X$'s) using Equation 6.7:

$$\text{logit}(p) = \log(p/(1-p)) = \beta_0 + \beta_1 X_1 + \ldots \beta_k X_k \tag{6.7}$$

Stepwise selection was used to determine which variables were relevant and should be included in the model. For each step, a covariate was added to the model if the significance level for entry was met (0.1 was used). Then the chi-square statistic was computed. If the covariate satisfied the significance level (0.1), it was included in the model. The Akaike information criteria (AIC) and Schwarz criterion (SC) were used to compare models and determine which variables to include in the final model.

Only a small sample of left- and right-side lane departures was available. As a result, it was not possible to evaluate the significance of all variables and test correlations between variables. In order to build a model that best represented the data, the decision to remove variables from the model was based on whether correlation among input variables was expected. The maximum likelihood (ML) method was used to calculate the coefficient estimates, and the Wald statistic was used to test the significance of covariates.

The variable Observation was used as the frequency variable in the model, which indicated the frequency of occurrence of each observation. This variable was used to weight the model. Odds ratios were used to assess whether a specific condition was more or less likely to result in a lane departure. An odds ratio greater than 1 indicated that the odds of a lane departure occurring are higher, and an odds ratio less than 1 revealed lower odds.

### *Left-Side Lane Departures*

Equation 6.8 describes the final model for left-side lane departures. The estimated log (odds) is given by

$$\begin{aligned}\text{Log}(\text{odds}) = 1.6107 - 0.0105 * \text{AGE} + 0.1682 \\ * \text{I}[\text{GENDER} = 1(\text{male})] - 0.00025 * \text{Radius} \\ - 0.7823 * \text{LaneWidth} - 0.3067 \\ * \text{I}[\text{TimeOfDay} = 0(\text{day})] - 20.9528 \\ * \text{CrashDensity}\end{aligned} \tag{6.8}$$

Model statistics are provided in Table 6.4, and the odds ratio estimates are shown in Table 6.5.

The first variable in Equation 6.8, AGE, is driver age. As indicated, as driver age increases, the odds for a left-side lane departure decrease, which indicates that involvement in left-side lane departures decreases with age.

Results for GENDER show that drivers involved in left-side lane departures are 1.4 times more likely to be male drivers (condition 1) than female drivers (condition 2).

The variable Radius is the radius of a curve. A very large value of 9999 was used for tangent sections, and the variable was modeled as a continuous variable. Table 6.5 shows that as radius increases, the likelihood of a lane departure decreases.

A negative correlation between LaneWidth and the likelihood of a left-side lane departure suggests that as lane width increases, the odds for the left-lane departure decrease.

The result for the variable TimeOfDay is a comparison of the odds of having a left-side lane departure during the day compared with at night. As shown, the odds ratio of having a left-lane departure during the day (condition 0) compared with at night (condition 1) is 0.542, indicating that a left-side lane departure is less likely to happen during the day. Alternatively, the odds of a lane departure at night compared with during the day are 1/0.542 = 1.85.

The last variable, CrashDensity, indicates the odds of a lane departure based on the density of lane departure crashes along the segment. As shown, the probability of having a left-side lane departure decreases as the density of lane departure crashes increases. The results of this variable are counter to what was expected. It was expected that roadway sections with a high density of lane departure crashes would be more likely to have lane departures.

### Right-Side Lane Departure Crashes

Equation 6.9 describes the final model for right-side lane departure events. The estimated log (odds) of a right-side lane departure is given by

$$\begin{aligned}\text{Log}(\text{odds}) = 0.1679 + 0.0427 * \text{AGE} - 0.00025 * \text{Radius} - 1.4042 \\ * \text{LaneWidth} - 0.8994 * I(\text{ShldType} = 3) - 2.9345 \\ * I(\text{ShldType} = 4) + 1.8799 * I(\text{ShldType} = 6) \\ + 0.365 * I(\text{ShldType} = 7) - 0.2864 * I(\text{Time} = 0) \\ + 81.9775 * \text{CrshDen} + 1.8348 * \text{OvrSpd10} \quad (6.9)\end{aligned}$$

Model statistics are provided in Table 6.6, and the odds ratio estimates are shown in Table 6.7.

The positive estimate for the variable Age in Equation 6.9 indicates that as age increases, the odds of a right-side lane departure also increase. This is the opposite of the result for left-side lane departures.

The negative estimate for the variable Radius indicates that as radius increases, the odds of having a right-side lane departure decrease. As a result, the likelihood of having a right-side lane departure is greater on curves with smaller radii.

The coefficient for LaneWidth indicates that as the lane width increases, the odds for a right-side lane departure decrease.

The variable ShldType indicates that the type of shoulder is significant. Nonpaved shoulder types were compared with paved shoulders (condition 1). The odds of having a right-side lane departure when gravel shoulders (condition 3) were present compared with when paved shoulders were present are 0.08. The odds of having a right-side lane departure when earth shoulders (condition 4) or partially paved shoulders (condition 7) were present compared with when paved shoulders were present are 0.01 and 0.29, respectively. Hence, the odds of a right-side lane departure are greater on paved shoulders than on earth, gravel, or partially paved shoulders. The odds of having a right-side lane departure when very narrow shoulders (condition 6) were present compared with when paved shoulders were present are 1.34. However, the confidence interval contains 1, so the difference is not statistically significant. In addition, there were very few observations where no shoulder was present.

The variable Time represents time of day. The odds of a right-side lane departure during the day (condition 0) compared with at night (condition 1) are 0.564. Alternatively, the odds of a

***Table 6.4. Model Fit Statistics for Left-Side Lane Departure***

| Criterion | Intercept Only | Intercept and Covariates | |
|---|---|---|---|
| AIC | 35815.581 | 29956.217 | |
| SC | 35825.176 | 30023.384 | |
| –2 Log L | 35813.581 | 29942.217 | |
| **Association of Predicted Probabilities and Observed Responses** | | | |
| Percent concordant | 80.0 | Somers' D | 0.624 |
| Percent discordant | 17.6 | Gamma | 0.640 |
| Percent tied | 2.4 | Tau-a | 0.047 |
| Pairs | 442473680 | c | 0.812 |
| **R-Square** | | | |
| R-square | 0.0526 | Max-rescaled R-square | 0.1873 |
| **Hosmer and Lemeshow Goodness-of-Fit Test** | | | |
| Chi-square | DF | Pr > ChiSq | |
| 445.9524 | 8 | <.0001 | |

***Table 6.5. Results for the Left-Side Lane Departure Model***

| Variable | Condition | Estimate | Std Error | *p*-value | OR 95% Lower | OR Estimate | OR 95% Upper |
|---|---|---|---|---|---|---|---|
| Age | 1 | –0.0105 | 0.00123 | <.0001 | 0.987 | 0.990 | 0.992 |
| Gender | 1 vs. 2 | 0.1682 | 0.0204 | <.0001 | 1.292 | 1.400 | 1.517 |
| Radius | 1 | –0.00025 | 3.637E-6 | <.0001 | 1.000 | 1.000 | 1.000 |
| LaneWidth | 1 | –0.7823 | 0.0712 | <.0001 | 0.398 | 0.457 | 0.526 |
| TimeOfDay | 0 vs. 1 | –0.3067 | 0.0178 | <.0001 | 0.505 | 0.542 | 0.581 |
| CrashDensity | 1 | –20.9528 | 3.5571 | <.0001 | <0.001 | <0.001 | <0.001 |

**Table 6.6. Model Fit Statistics for the Right-Side Lane Departure Model**

| Criterion | Intercept Only | Intercept and Covariates | |
|---|---|---|---|
| AIC | 17921.931 | 11629.157 | |
| SC | 17931.503 | 11734.452 | |
| –2 Log L | 17919.931 | 11607.157 | |
| **Association of Predicted Probabilities and Observed Responses** | | | |
| Percent concordant | 94.0 | Somers' D | 0.884 |
| Percent discordant | 5.6 | Gamma | 0.887 |
| Percent tied | 0.3 | Tau-a | 0.029 |
| Pairs | 183668320 | c | 0.942 |
| **R-Square** | | | |
| R-square | 0.0578 | Max-rescaled R-square | 0.3717 |
| **Hosmer and Lemeshow Goodness-of-Fit Test** | | | |
| Chi-square | DF | Pr > ChiSq | |
| 1510.8238 | 8 | <.0001 | |

right-side lane departure at night can be computed by 1/0.0564 = 1.77. Hence, the odds of having a lane departure at night are 1.77 times the odds of having one during the day.

The variable CrshDen is the number of actual lane departure crashes per meter for the section of roadway where the vehicle activity took place. The result shows that as crash density increases, the odds of having a right-side lane departure increase dramatically. This is also the opposite of what was found for left-side lane departures.

The last variable, OvrSpd10, indicates the amount of time a driver spends going 10 mph over the speed limit. The results indicate that drivers who spend more time traveling 10 mph or more over the speed limit increase their odds of having a right-side lane departure.

## Sample Size for Full-Scale Study

As indicated, only limited data were available for evaluating the lane departure research questions. In this section, a method to estimate sample size for the full-scale study is presented for the logistic regression.

A literature review regarding sample size indicated that there are various schools of thought on determining sample size for logistic regression. References include Hosmer and Lemeshow (2000, 339–347), Agresti (2002, 242–243), and Hsieh et al. (1998, 1623–1634).

Calculation of sample size for logistic regression can be complicated because multiple logistic regression analysis is nonlinear. Hsieh et al. (1998) suggest a method for simplifying sample-size calculation. Based on their method, the following describes an example calculation of sample size for the two logistic regression models presented in the previous section.

### Left-Side Lane Departure

In order to calculate sample size for left-side lane departures, one of the explanatory variables that is of interest is first chosen (e.g., LaneWidth, termed as $X$).

The sample size is calculated according to the following equations:

$$n = \left[z_{1-\alpha} + z_{1-\beta}\exp\left(-\tau^2/4\right)\right]^2\left(1+2\pi\delta\right)\Big/\left[\left(\pi\tau^2\right)*\left(1-\rho^2\right)\right] \quad (6.10)$$

$$\delta = \left[1+\left(1+\tau^2\right)\exp\left(5\tau^2/4\right)\right]/\left[1+\exp\left(-\tau^2/4\right)\right] \quad (6.11)$$

where

$\pi$ is the estimated probability when all the continuous variables are at their means, calculated as 0.03184.

**Table 6.7. Results for the Right-Side Lane Departure Model**

| Variable | Condition | Estimate | Std Error | *p*-value | OR 95% Lower | OR Estimate | OR 95% Upper |
|---|---|---|---|---|---|---|---|
| Age | 1 | 0.0427 | 0.00286 | <.0001 | 1.038 | 1.044 | 1.050 |
| Radius | 1 | –0.00025 | 6.192E-6 | <.0001 | 1.000 | 1.000 | 1.000 |
| LaneWidth | 1 | –1.4042 | 0.1140 | <.0001 | 0.196 | 0.246 | 0.307 |
| ShldType | 3 vs. 1 | –0.8994 | 0.0647 | <.0001 | 0.069 | 0.083 | 0.101 |
| ShldType | 4 vs. 1 | –2.9345 | 0.1048 | <.0001 | 0.008 | 0.011 | 0.014 |
| ShldType | 6 vs. 1 | 1.8799 | 0.1220 | <.0001 | 0.967 | 1.338 | 1.850 |
| ShldType | 7 vs. 1 | 0.3650 | 0.0519 | <.0001 | 0.252 | 0.294 | 0.343 |
| Time | 0 vs. 1 | –0.2864 | 0.0402 | <.0001 | 0.482 | 0.564 | 0.660 |
| CrshDen | 1 | 81.9774 | 7.9325 | <.0001 | >999.999 | >999.999 | >999.999 |
| OvrSpd10 | 1 | 1.8347 | 0.0696 | <.0001 | 5.464 | 6.264 | 7.180 |

$\tau$ is the effect of $X$ at the mean level of the other predictors. For example, to determine the necessary sample size for detecting that the effect of a one standard deviation increase in lane width results in a 50% increase in the odds of left lane departure, with all other continuous variables at their mean values, then $\tau = \log(1.5)$.
$z_{1-\alpha}$ and $z_{1-\beta}$ are the $(1 - \alpha)$ and $(1 - \beta)$ standard normal quantiles, respectively.
$\alpha$ is the level of significance, which is 0.05 here.
$1 - \beta$ is the power, which is 0.9 here.
$\rho$ is the multiple correlation of $X$ and the remaining covariates in the model.

The $R^2$ in linear regression can be used to measure $\rho$, which is 0.1621 here. Inserting all of the above values into Equation 6.11 provides an estimated sample size for the logistic regression, which is 1,663.

#### *Right-Side Lane Departure*

Similarly for the right-side lane departure model, sample size is calculated using

$\pi = 0.01634$
$\rho = 0.1505$
$\tau = \log(1.5)$

The sample size is determined to be 3,109 using Equations 6.10 and 6.11.

### Application to Full-Scale Study

In order to apply logistic regression to the full-scale naturalistic driving study, the following approach may be considered. A sequential block approach may be used to reduce the data. The first step would be to identify all lane departure crashes, near crashes, and encroachments that meet the requirements of the research question. For instance, only right-side lane departures on four-lane, rural, divided roadways may be included. A set amount of time, an epoch, would be determined based on the average length of lane departure. For instance, the epoch could comprise 3 s before the lane departure and 3 s after, resulting in a 6-s interval. Normal driving data could be sampled at regular intervals (e.g., 5 min) and data aggregated for that epoch. For instance, if a 6-s epoch was selected, all lane departure events would be extracted, data for vehicle activity meeting the criteria would be sampled every 5 min, and 6 s of data would be reduced for that interval. Driver, roadway, and environmental conditions would need to be consistent across the epoch. For instance, if the driver were traveling on a tangent section at the beginning of the epoch and then encountered a curve after 2 s, the epoch would have to be adjusted to include just the tangent section or the curve.

Logistic regression analysis is ideal for the naturalistic driving study because normal driving data will be provided that can be used to account for exposure. Historically, it has been difficult to account for driver activity under a range of situations to determine if one situation is overrepresented. For instance, it is commonly accepted that crashes are more likely during a winter weather event. However, it is very difficult to determine what fraction of time drivers spend driving on snowy or icy roads, so it is difficult to determine whether crashes under these conditions are overrepresented.

Additionally, the results of logistic regression can be expressed as odds ratios, which can easily be explained to lay persons and used by transportation agencies.

## Analysis Approach 3: Logistic Regression for Correlated Data

The previous section described an analysis approach using logistic regression to evaluate the odds of having a left- or right-side lane departure based on a small sample of available data. Because the sample size was small, it was difficult to address issues such as the correlation between data that occurs when repeated samples are taken from the same situation (e.g., repeated samples for the same driver, same trip). This section provides an alternate approach using logistic regression considering correlated data.

### Description

This approach considers matched control and event samples to avoid confounders. For each selected case epoch, several matched controlled periods of the same length are sampled from the same driver and same trip. Other covariates not selected in the model (either because of not recorded or not enough data to have a good estimate) are assumed to be constant in the same trip of the same driver. For example, a sleepy driver in a trip is sleepy the whole trip, not just at the end of the trip. The effects of these covariates are thus eliminated in this matched case-control model. Further, each epoch is assumed to be separated (not adjacent), so there is an assumption of independence within each matched case-control set.

The conditional logistic regression model focuses on estimating the differences within these matched sets. The goal is to understand the association between covariates (either environmental, driver related, roadway related, or vehicle related) on the probability of an event. Each period includes a response variable that takes on the values $Y = 1$ (event) or $Y = 0$ (no event) and candidate explanatory variables whose distribution of values within the epoch can be summarized using, for example, the observed range (max-min) of values of the covariate. Consider $Y_{ij} \mid X_{ij} \sim \text{Bernoulli}(p_\theta(X_{ij}))$, the response in the $j$th sample of the $i$th driver. Assume that $j = 1$ corresponds to the case. Let $Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})$ and $X_i = (X_{i1}, X_{i2}, \ldots, X_{in_i})$.

The likelihood function is given by $L_i(\theta|Y_i) = f(Y_i|X_i) / [\Sigma_{\text{permutation of } yi} f(Y|X_i)] = [\Pi_{j\in \text{control}} \exp(x - {}_{ij}\beta)] / [\Sigma_{\text{choose}(ni,mi)} \Pi_{j\in \text{control}} \exp(x_{ij}\beta)]$.

The following example combines left-lane and right-lane departures as events. The significant positive model coefficients suggest that there exists a correlation between an event and time periods when a driver exhibits a large variation in lateral speed and lateral acceleration.

Data set:

| Driver | 6 | 8 | 12 | 24 | 48 | 51 | 60 | **Total** |
|---|---|---|---|---|---|---|---|---|
| Number of sampled periods (1 case, others control) | 11 | 3 | 21 | 15 | 16 | 6 | 16 | 88 |

If the variable LaneOffset is available:

| | coef | exp (coef) | se (coef) | *z* | *p* |
|---|---|---|---|---|---|
| max(LaneOffset) – min(LaneOffset) | 8.37 | 4317 | 3.05 | 2.74 | 0.0061 |
| max(AY) – min(AY) | 6.63 | 755 | 3.01 | 2.20 | 0.0280 |

Likelihood ratio test = 19.4 on 2 df, $p$ = 6.21e-05, $n$ = 88.

If the variable LaneOffset is not available:

| | coef | exp (coef) | se (coef) | *z* | *p* |
|---|---|---|---|---|---|
| max(LATERALSPEED) – min(LATERALSPEED) | 4.91 | 136.0 | 2.44 | 2.01 | 0.044 |
| max(AY) – min(AY) | 3.63 | 37.7 | 2.83 | 1.28 | 0.200 |

Likelihood ratio test = 14.2 on 2 df, $p$ = 0.000817, $n$ = 88.

The correlation within is an important issue for the longitudinal data. Not dealing with this correlation can cause biased estimates and underestimated standard deviation (suppose positively correlated). The following sections describe another model that assumes hierarchical structure to deal with the correlation within trips and nested in drivers. While the matched case-control method assumes independent matched set and tries to eliminate the correlation, the following method puts the correlation structure in the model.

## Sample

Given the example above, suppose now that periods are selected from both run-off-road (ROR) and non-ROR events under some common fixed condition, such as the same curvature or the same weather conditions. Samples are selected from all qualifying periods.

## Difference from the Matched Case-Control (the Conditional Logistic Model)

For the periods in the preceding section, the data set is constructed by sampling from the population of cases and the population of controls, even though observations may be correlated (e.g., they could be sampled from the same trip).

## Question of Interest

The question that may be answered with this approach is: What factors may be associated with the risk of ROR events?

## Response Variable

The response variable is a ROR event.

## Covariates

Any variable for which measurements are available can be included in the model as an independent variable. These might include, for example, driver characteristics, environmental conditions, and road conditions.

## Model

The model is described by the following: Let $Y_{ij} = 1$ if the $j$th sample (period) for the $i$th driver has ROR = 1. Assume that the distribution of $Y_{ij}$ is Bernoulli with $\text{logit}(P(Y_{ij} = 1|X_{ij})) = X_{ij}\beta + Z_{ij}\gamma_i$. Here, $X_{ij}$ denotes the covariates corresponding to the $j$th period for the $i$th driver, and $\gamma_i$ is a (multivariate) normal distributed random variable that is driver-specific. This random variable permits accounting for the correlation between observations within the same driver data.

## Model Example

For the example, continuous periods longer than 5 s (ROR event periods or nonevent periods) are selected. For each period, the middle 5 s are selected and the variables of interest within each period are summarized. Then, consider a random intercept for “driver” and a second random effect to represent “trip nested within driver” in the mixed-effect model and use forward selection of covariates based on the AIC, as shown in the following:

Fixed effects (R: glmer)

| | Estimate | Std. Error | Z value | Pr(>\|z\|) |
|---|---|---|---|---|
| (Intercept) | –8.662 | 1.880 | –4.61 | 4.1e-06 |
| Mean of shoulder width | 0.995 | 0.400 | 2.48 | 0.013 |
| Max(LaneOffset) – min(LaneOffset) | 3.346 | 1.986 | 1.68 | 0.092 |
| Max(YawRate) – min(YawRate) | 0.592 | 0.330 | 1.80 | 0.073 |

Estimated using generalized estimating equations (GEE) (R: geepack: geeglm)

| | Estimate | Std. Error | Wald | Pr(>\|W\|) |
|---|---|---|---|---|
| (Intercept) | –7.721 | 2.112 | 13.36 | 0.00026 |
| Mean of shoulder width | 0.819 | 0.500 | 2.69 | 0.10119 |
| Max(LaneOffset) – min(LaneOffset) | 2.419 | 1.544 | 2.45 | 0.11728 |
| Max(YawRate) – min(YawRate) | 0.681 | 0.248 | 7.51 | 0.00614 |

The statistical model R was used to estimate the example, and the functions used are shown in parentheses.

Note that by introducing the random effects into the model, inferences about the association between lane offset and the probability of an event and between yaw rate and the probability of an event are impacted and changed from statistically significant to statistically insignificant.

## Sample Size

Estimating sample size in generalized linear mixed models is, in general, not a straightforward endeavor. Dang et al. (2008) and Liu and Liang (1997) derived the exact form of the sample size estimator for the two-sample problem with correlated binary responses and exchangeable correlation structure by finding the approximating variance of the regression coefficient. Maas and Hox (2005) presented a simulation result for models with one random coefficient and one random slope at different sample sizes.

## Sample Size Calculation Using a Generalized Estimating Equations Method and a Simple Example

The following provides an example sample size calculation. Consider an additional explanatory variable, OvrSpd5, that is associated with driver behavior (frequency of driving over the speed limit). A mixed model was fit using GEE as described for the example above, and the regression coefficient associated with OvrSpd5 was not found to be significantly different than zero. Because the p-value for the hypothesis that the regression coefficient is equal to zero is 0.9338, the null hypothesis $H_0$: $\beta_4 = 0$ is not rejected. If the alternative hypothesis ($H_0$: $\beta_4 = \beta_a$) happens to be true, then we would like to have enough power to reject the null hypothesis. Assume that the estimated coefficient $\beta_4 = -0.0136$ is correct and that the standard error 0.1543 is correct under the current sample size. We want to increase the sample size to reduce the standard error enough so that we can achieve Type I error <0.05 when the true value of $\beta_4$ is 0 and the power of the test is at least 0.8 when the true value of $\beta_4$ is –0.0136. The coefficients for this example are described as follows:

Coefficients (R: geepack: geeglm)

| | Estimate | Std. Error | Wald | Pr(>\|W\|) |
|---|---|---|---|---|
| (Intercept) | –7.3837 | 5.1027 | 2.09 | 0.1479 |
| Mean of shoulder width | 0.8218 | 0.4830 | 2.90 | 0.0888 |
| Max(LaneOffset) – min(LaneOffset) | 2.3889 | 1.6127 | 2.19 | 0.1385 |
| Max(YawRate) – min(YawRate) | 0.6792 | 0.2604 | 6.80 | 0.0091 |
| OvrSpd5 | –0.0136 | 0.1643 | 0.01 | 0.9338 |

To obtain an estimate of the appropriate sample size under those conditions, some assumptions need to be made. These are as follows:

1. There is a fixed number of drivers indexed by $s = 1, 2, \ldots, S$.
2. Each driver has repeated measurements $t = 1, 2, \ldots, T$.
3. The correlation structure is "exchangeable." This means that every pair of samples in the same subgroup has the same correlation.
4. We are interested in testing the hypothesis $H_0$: $H_A = h_0$ versus $H_1$: $H_\beta \neq h_0$. $H$ is (0, 0, 0, 0, 1) and $h_0$ in this example is 0.
5. We let $b$ denote the point estimate of $\beta$ and let $\text{cov}(b) = T^{-1}V_b$ (where the covariance matrix can either be model based or can be estimated using robust methods). Then, the Wald test statistic $Q = T(Hb - h_0)'[HV(b)H']^{-1}(Hb - h_0)$ is asymptotically distributed as a $\chi^2(p, \lambda)$ random variable with df $= p$ and a noncentrality parameter $\lambda = 0$ under $H_0$ and $\lambda = \lambda_{H1}$ under $H_1$. The power of the test is $P_{H1}(Q > \chi^2(p, 0)_{1-\alpha})$, where $\alpha$ is the significance level. The sample size is calculated by finding the minimum $n$ such that the $P_{H1}(Q > \chi^2(p, 0)_{1-\alpha})$ achieves the desired power level.

## Comparison of the Logistic Regression Model

Both methods are trying to deal with the correlation structure. The matched case-control method uses fewer samples; the mixed effect model can use more samples but needs to make assumptions on the correlation structure and estimate extra coefficient for correlations. Currently, both models can extract information from this pilot data set. For a larger scale of study, when researchers can afford to deal with the correlation structure, the mixed effect model may provide more information.

## Analysis Approach 4: Time Series Analysis

Different from more common case-control study, the naturalistic data provides more than just counts of events. The purpose of using a dynamic model that puts interests on each 0.1 s includes modeling the pattern of driving and providing information "on" (while) driving.

For example, we know lane offset is correlated to lane departure events. When the time window is small, the car is out of the road during the event, so the measurement of lane offset is different from normal driving. When the time window is slightly larger, the averaged lane offset has not crossed the edge of the road, and there is no difference between "driver feels comfortable to stay close to edge at this section" and "driver is going to cross the edge next second." The "random" (distribution) explains the different outcomes with the same explanation variables as "randomness."

We could also look at the data in another perspective. We look at one instant while driving and think the following actions as the results of current status, driver's decision and operation, environment effect, and some randomness. Assume the current status is fixed and observed. Other factors are changed over time. If we could build a model, we could forecast a few seconds ahead. We might be able to determine some conflicts—for example, in danger but not reacted, or in danger and reacted but not enough.

This example is rather simplified. In the larger study, this model needs several longer continuous mechanical-recoded data that are known in closed situations (e.g., similar lane type) to train the basic model and test on the shorter manually collected data (e.g., from video, radius).

The fourth analysis approach used continuous data in a time series model, as described in this section.

### Description

The main advantage of applying a time series analysis to naturalistic driving study data is that it allows relationships between variables across time to be incorporated into the model. As a result, relationships can be established between, for example, driver distraction in previous time periods and probability of a lane departure or crash in a subsequent time period. A time series model can also be used to model outcome. Current methods, which use crash data to analyze the impact of countermeasures on safety, have only accomplished their goals by waiting for the system to fail (i.e., a crash occurs). In contrast, a time series analysis allows positive outcomes to be evaluated and relationships between positive outcomes and roadway, driver, or environmental features to be determined.

Time series models are extensions of regression models, where the errors are assumed to be correlated; thus, selection of independent variables to be included in the model and the form of the association between independent and dependent variables in the model can be addressed in a standard fashion.

### Sampling Approach

To demonstrate this analysis approach, data were modeled using continuous data (i.e., each observation represents 0.1 s of vehicle activity). All of the variables listed in Tables 6.2 and 6.3 were available but, because of the complexity of a time series model, only a few initial variables were included to demonstrate proof of concept.

### Response Variables

The response variable is vector-valued. The first element of the vector is a variable associated with *movement* (e.g., lane offset, yaw rate), whereas the second element is a variable associated with the *operation* of the vehicle (e.g., acceleration). We use $Y_1(t)$ and $Y_2(t)$ to denote the first and second elements of the response vector.

### Explanatory Variables

The model can include continuous, block-summarized, or static covariates. It can also include smoothed functions of independent variables that may vary by periods. We use $X(t)$ to denote the value of the vector of covariates at time $t$. There is no restriction on the number and type of covariates that can be included in the model. In particular, we can include driver, roadway, vehicle, or environmental variables and explore the association between them and the response variable while at the same time accounting for the correlation of consecutive observations obtained from the same process.

### Modeling Approach and Results

We assume that $E(Y_1(t)) = g(Y_1(<t), X(\leq t), Y_2(<t))$. That is, the mean of the movement response variable at any given time $t$ depends on movement at times preceding $t$, on the covariates up to and including their value at time $t$, and on the operation response variable at times prior to $t$. To complete the specification of the model, it is necessary to define the kernel of the model (i.e., the functional form for $g$ in the equation), the size of the lag (the number of observation periods for which correlation coefficients will be estimated), and the structure of the error term in the model.

One widely used model that permits accounting for the autocorrelation in the observed data is the Autoregressive

*Table 6.8. Estimates of the Parameters in the ARMA(3,2) Model Based on the Data Collected During 102.5 s During the Second Trip of Driver 6 in the Data Set Coefficients*

| | ar1 | ar2 | ar3 | ma1 | ma2 | Intercept | LATERALSPE ($t-1$) | AY($t-1$) |
|---|---|---|---|---|---|---|---|---|
| | 2.8021 | 2.6402 | 0.8379 | –1.8103 | 0.9189 | 0.2742 | –1.1262 | 2.0016 |
| SE | 0.0263 | 0.0524 | 0.0262 | 0.0192 | 0.0167 | 0.2382 | 0.1481 | 0.8626 |
| $\sigma^2$ estimated as 0.000323: log likelihood = 2660.57, AIC = –5303.15 | | | | | | | | |

Moving Average Model of order $p$ and $q$ (ARMA($p$,$q$)). The standard version of the ARMA($p$,$q$) model has a linear kernel and the general form

$$Y_1(t)=a_1Y_1(t-1)+\cdots+a_pY_1(t-p)+e(t)+b_1e(t-1)$$
$$+\cdots+b_qe(t-q)+\beta' X(t) \quad (6.12)$$

where $Y_1(t)$ might denote, for example, lane offset at time $t$; $e(t)$ is a random error term, often distributed as a normal random variable; and $X(t)$ is a vector of explanatory variables. The coefficients $a_i$, $b_j$, and $\beta$ are unknown and must be estimated from the data. The model includes $p$ lagged terms for the autoregressive part of the model, and $q$ lagged terms for the moving average portion of the model.

As an example, consider a specific driver from the available data set. An ARMA(3,2) model was fit to the continuous observations obtained over 102.5 s in Trip 2 of Driver 6. The coefficients $p = 3$ and $q = 2$ and the two covariates were chosen using the AIC.

Suppose that we wish to predict the location of the vehicle at time = $t + 1$, given information about the location, lateral speed, and lateral acceleration of the vehicle at time = $t$ and earlier. In this example, we only use time-dependent covariates. The example that follows presents a more general model, where other types of covariates, such as shoulder surface and shoulder width, are also included.

The model was fit using the statistical software R. Table 6.8 shows the estimated model parameters and their standard errors. We note, for example, that lateral speed at time $t$ is negatively and significantly associated with lane offset at time $t+1$, and the reverse is true for lateral acceleration. These two explanatory variables appear to be good predictors of lane offset.

Figure 6.9 shows two graphs. On the top panel, the standardized residuals over time computed from the model are

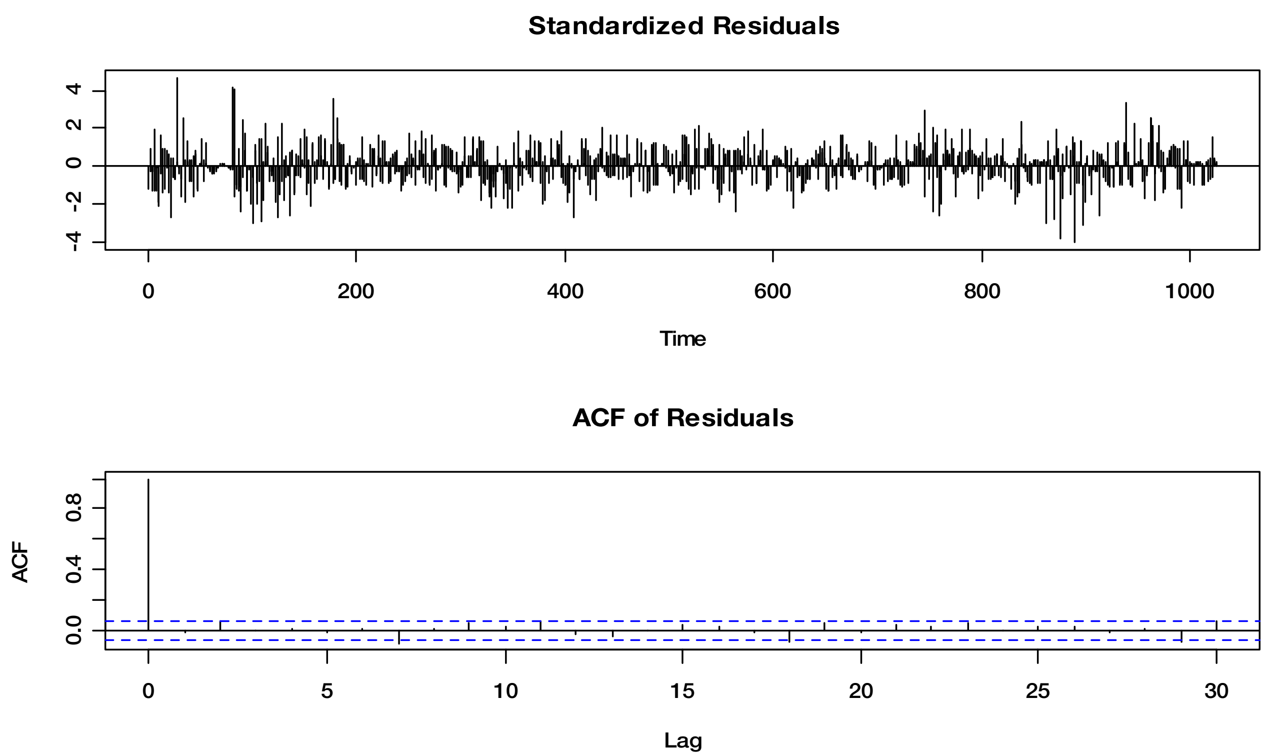



*Figure 6.9. Diagnostic plots for the ARMA(3,2) model fit to Driver 6.*

displayed. Because the residuals are standardized, we expect that about 99% of them will be within three standard deviations of their mean zero. The plot suggests that there are very few residuals that exceed the value 3 (in absolute value), so we are comfortable concluding that there appear to be no outliers in this particular data set and with respect to this model. Further, there seems to be no obvious pattern in the residuals, even though they are plotted in time order. This is consistent with the plot shown in the bottom panel of Figure 6.9. In this plot, we show the autocovariance function estimated from the estimated residuals. If the order of the model is correct, then we expect to see no significant autocorrelation among residuals. From these two diagnostic plots, we conclude that the model appears to fit the data reasonably well and that the autocorrelation and moving average structure in the model account for the correlation between observations collected over time.

Because the goal was to predict the lane offset at a future time given information available now, we predict the lane offset for this driver during this trip for the 3 s that follow the end of the trip. The predicted lane offset and the 1 standard deviation bands are shown in Figure 6.10.

However, to obtain lane offset predictions, we used a naive approach, in that we assumed that both lateral speed and lateral acceleration remained fixed at the values observed at the end of the trip. We know that lateral speed and lateral acceleration also change during the prediction period, however. The simple prediction approach can be extended to allow for evolution of the covariates over time, but to do so we must explicitly include lateral speed and lateral acceleration as vector-valued response variables and model each variable as a function of the other two.

For a second example, we can use the data collected for Driver 51, whose trip included two curves. It can be anticipated that the association between lane offset and curve will depend not only on curve characteristics such as length and radius but also on the location of the vehicle within the curve. In the original data set, we have information about whether the driver is entering a curve to the left or to the right, and we also know the length of the curve. The data set includes a variable called CURVE, which takes on the value 1 during all time periods in which the driver is taking a right curve, the value 2 during all time periods in which the driver takes a turn to the left, and the value 0 whenever the driver's vehicle is on a straight road.

Figure 6.11 shows the value of the variable CURVE for Driver 51. We have changed the labels to −1, 0, and 1 to denote left curve, no curve, and right curve, respectively. The three dashed curves colored red, blue, and green in the figure correspond to three different smooth functions that depend on the length of the curve. All three smooth representations (or summaries) of the curves improve the fit of the time series model relative to the model that includes the static −1, 0, 1 labels. The function drawn in red is best, at least in the AIC sense.

The function that smoothes out the effect of a curve over the period during which the driver is negotiating it can perhaps be improved by including the radius of the curve (in addition to the length) in the smoothing function.

Using the same ARMA(3,2) model but now with an additional explanatory variable consisting of the smooth curve indicator, we fitted the indicator corresponding to the red trajectory in Figure 6.11. Table 6.9 shows the estimated model parameters and their standard errors. Note that when we include the curve indicator into the model, lagged yaw rate is no longer statistically significant. Lateral speed continues to be negatively and significantly associated with lane offset.

As in the earlier example, we can explore residual plots and autocovariance plots to carry out model diagnostics. Figure 6.12 shows the time-ordered estimated residuals (top panel) and the autocovariance function for the estimated residuals (bottom panel). We see from the top panel that the proportion of standardized residuals with very high or very low values is

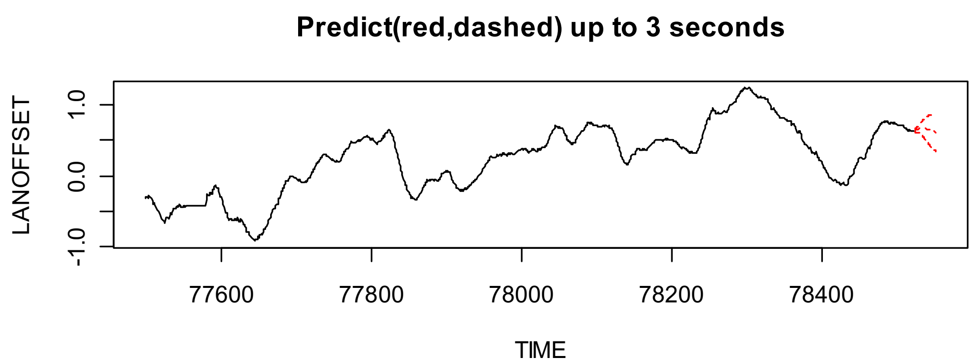



***Figure 6.10. Observed lane offset for Driver 6 in Trip 2 (black solid curve), one-step-ahead prediction of the 3 s beginning at the end of the trip (middle red dashed curve), and the ±1 standard deviation region around the prediction (top and bottom red dashed curves).***

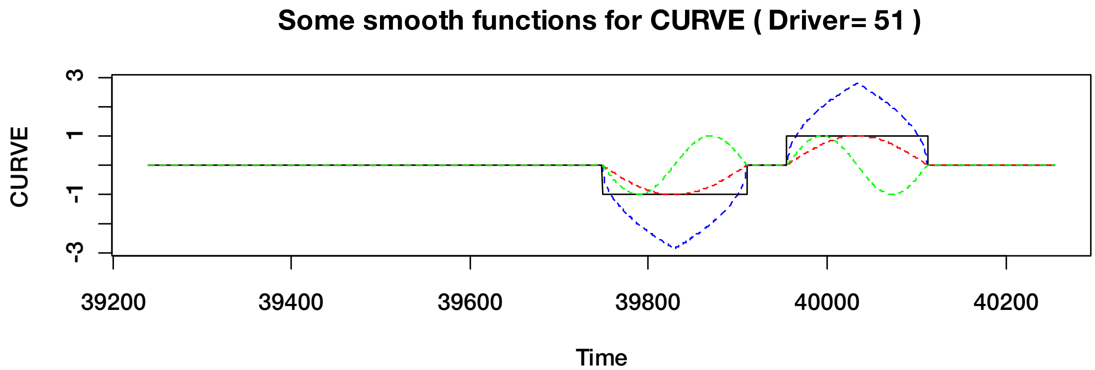



*Figure 6.11. Original curve indicator (black solid line) and three smooth functions of curve length.*

*Table 6.9. Estimates of the Parameters in the ARMA(3,2) Model Based on the Data Collected for Driver 51 in the Data Set*

| | ar1 | ar2 | ar3 | ma1 | ma2 | Intercept | LATERAL SPE (t − 1) | AY (t − 1) | Smooth (Curve) |
|---|---|---|---|---|---|---|---|---|---|
| | 0.8558 | 0.9891 | –0.8541 | 0.2144 | –0.7747 | –0.2199 | –1.2364 | –0.5456 | 0.9797 |
| SE | 0.0662 | 0.0137 | 0.0649 | 0.0833 | 0.0832 | 0.0975 | 0.4287 | 3.2924 | 0.2394 |
| $\sigma^2$ estimated as 0.00443: log likelihood = 1309.5, AIC = –2599 | | | | | | | | | |

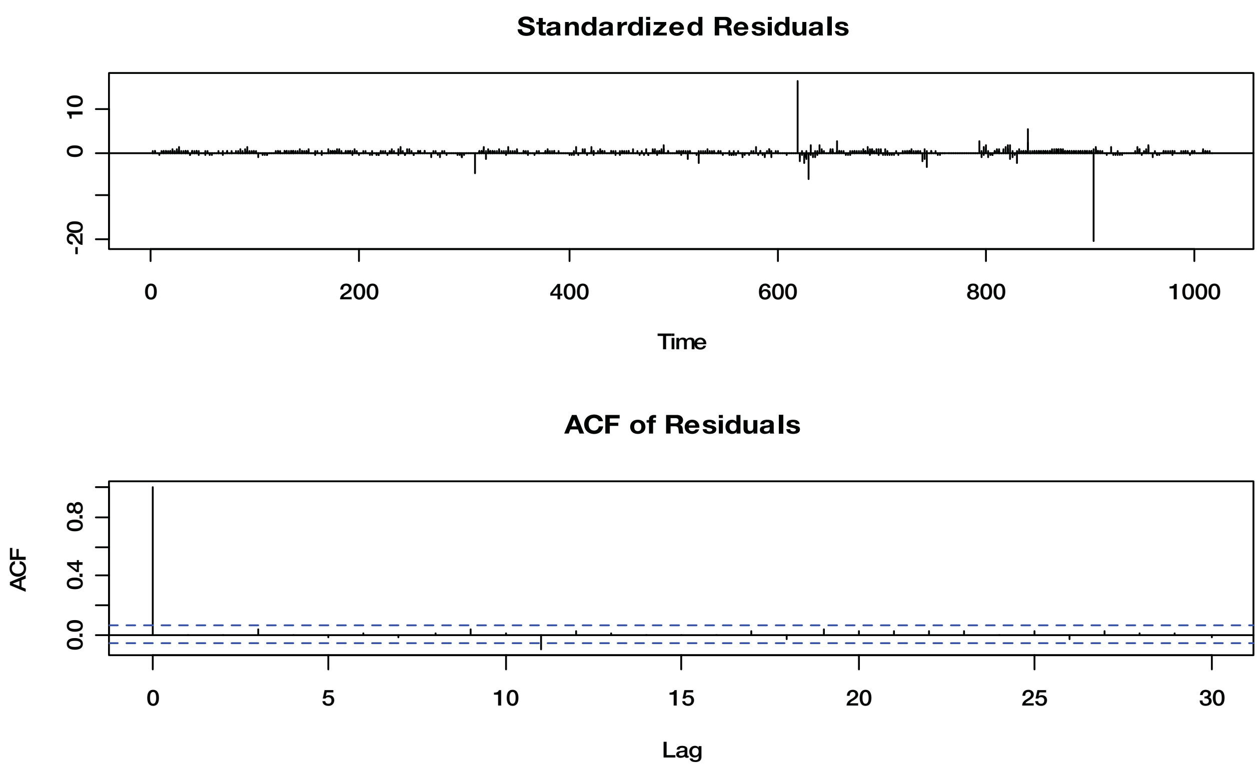



*Figure 6.12. Diagnostic plots for the ARMA(3,2) model fit to Driver 51.*

negligible, and the autocovariance function in the bottom panel suggests that the autoregressive and the moving average structures in the residual account for the residual time dependence.

### Sample Size

Several methods were considered to estimate the sample size needs for conducting a time series analysis in the full-scale study. The methodology is rather complicated, and it was decided that it is beyond the scope of this report to describe the methodology.

### Application to Full-Scale Study

For continuously driving online forecasting in the full-scale study, the research team proposes fitting a normal dynamic linear model (DLM) that permits continuous updating of the forecast distributions when new observations become available. Each update can be made by optimizing some function of the observed data and the previous forecasts. Two such optimization approaches include the minimum mean square and the Bayesian (posterior distribution) criterion. If the variance is known, the Bayesian forecasting for the DLM is essentially equivalent to the Kalman filter used extensively in engineering control processes.

The univariate normal DLM is sometimes known as a state-space model and includes the following:

- Observation equation $Y_t = F_t\theta_t + \nu_t$, with $\nu_t \sim N_1(0, V_t)$
- State evolution equation $\theta_t = G_t\theta_{t-1} + \omega_t$, with $\omega_t \sim N_p(0, W_t)$
- Initial prior $(\theta_0|D_0) \sim N(m_0, C_0)$, where $(m_0, C_0)$ fixed and $D_t = \{Y_t, D_{t-1}\}$

The model states that the underlying "state" $\theta_t$ evolves smoothly over time as an autoregressive process and that the observation at time $t$ is a smooth function of the state. Coefficients $F_t$ and $G_t$ are often assumed to be constant over time, but they can also be allowed to be time dependent.

When the state-space model is linear and when the two random drivers $\nu$ and $\omega$ are normally and independently distributed, forecasting consists essentially of the estimation of normal conditional means at each step. The one-step forecast at each $t$ is then obtained as follows:

- Posterior at $t - 1$: $(\theta_{t-1}|D_{t-1}) \sim N(m_{t-1}, C_{t-1})$.
- Prior at t: $(\theta_t|D_{t-1}) \sim N(G_t m_{t-1}, R_t)$ where $R_t = G_t C_{t-1} G_t' + W_t$.
- Forecast: $(Y_t|D_{t-1}) \sim N(F_t' G_t m_{t-1}, Q_t)$ where $Q_t = F_t' R_t F_t + V_t$.
- Posterior at $t$: $(\theta_t|D_t) \sim N(m_t, C_t)$ where $m_t = G_t m_{t-1} + A_t(Y_t - F_t' G_t m_{t-1})$ and $C_t = R_t - A_t A_t' Q_t$, $A_t = R_t F_t Q_t^{-1}$

The following is an example problem of the Kalman filter from Bar-Shalom et al. 2001. This example involves trying to estimate the distance (range) between two vehicles and their relative speed (range rate).

Consider $X(t) = (\text{range}(t), \text{range rate}(t))'$. Assuming a constant range rate, we have the following:

- Original state equation $x(k) = Fx(k-1)$, where $F$ is the system matrix.
- Original measurement $y(k) = Hx(k)$, where $H$ is the measurement matrix.
- True state equation $x(k) = Fx(k-1) + Gu(k-1)$, where $u(k-1)$ is acceleration.
- Observed measurement $y(k) = Hx(k) + w(k)$, where $w(k)$ is the measurement error.

In previous continuous-time models, the explanatory variables in the additive model explained the vehicle shifts (left ← or right →) related to the variables. Another option is to connect the explanatory variables to vehicle recovery time after model prediction. Let $t_0$ = the starting time where the 3-s prediction confidence (credible) region covers either edge of the lane. If the driver adjusts the vehicle before the ROR event happens, then $t = 0$. Let $t_1$ = the time it takes the whole vehicle to return to the lane. Assume the failure time $t = t_1 - t_0$ is exponentially distributed with the density function $f(t) = \lambda e^{-\lambda t}$. The log hazard function should be modeled as $h(t) = X\beta$. This model answers questions like "recovery time versus road condition" or "recovery time versus driver's record."

## Summary and Conclusions

Several exploratory analysis methods were applied to data extracted from existing naturalistic driving studies to demonstrate ways in which lane departure research questions could be answered in the SHRP 2 full-scale study. The intent of the analyses was to demonstrate different methods that could be used to analyze the data that will result from the full-scale study.

A data sampling approach developed by the SHRP 2 Safety Project S02 researchers was described, and four analysis methods were presented. The four approaches included (1) a data mining approach using classification and regression tree analysis, (2) simple odds ratio and logistic regression, (3) logistic regression for correlated data that accounts for repeated sampling among observations (e.g., repeated sampling for the same driver, trip), and (4) a time series analysis.

Three of these methods were used to evaluate existing naturalistic driving study data, and one method expanded on a varied logistic regression approach that may be better suited to the data from the full-scale study. Data were available from the UMTRI road departure crash warning (RDCW) field operation test (FOT) that contained a number of nonconflict lane departures and samples of normal driving. Methods 1 and 2 ([1] classification and regression tree and [2] simple odds ratio and logistic regression) evaluated the likelihood of a left- or right-side lane departure. A sample-based approach was used in the classification and regression tree analysis, and an event-based approach was used for the logistic regression. Although available sample sizes were limited, both methods resulted in similar results. Both indicated that curve radius, driver age, and type of shoulder were relevant in explaining lane departures. The logistic regression also indicated that both left- and right-side lane departures were more likely to occur at night and were less likely to occur as lane width increased. The model for left-side lane departures indicated that male drivers were more likely than female drivers to be involved in a lane departure, and the model for right-side lane departures indicated that lane departures are more likely on roadway sections with a higher density of lane departure crashes and for drivers who spend more time traveling 10 mph or more over the posted speed limit.

The fourth method, time series analysis, used continuous data to develop a model to predict offset as a function of several vehicle kinematic variables. The method was developed and explained in such a way that it could be adapted to the full-scale study to include various explanatory variables, including driver behavior. This approach allows information, such as driver distraction in previous time periods, to be incorporated into the model.

As indicated, the analyses presented in this chapter were exploratory, with the intent to demonstrate different analysis methods that could be used to analyze the data that will result from the full-scale study. Because the amount of data was limited, the analyses in most cases yielded only preliminary results.

Selecting an appropriate model for the full-scale naturalistic study will depend on the research questions posed and the resources that can be used to reduce data. Each approach has its advantages and limitations in terms of the full-scale naturalistic study. The main advantage of the classification and regression tree analysis is that it can be used to uncover patterns in the data that other methods may mask. The results may indicate that a variable is only relevant at a certain point (splitting value). For instance, there may only be a correlation between lane departures and curves with a radius of 500 ft or less, while no relation exists with larger curve radii. It is difficult to uncover this sort of structure using other models. Tree models are also adept at revealing complex interactions between variables. Each branch may have different combinations of variables, and the same variable can be present in more than one part of the tree. This complexity reveals dependencies between variables and the point at which the dependency exists (Hosmer and Lemeshow, 1986). However, several disadvantages exist for this method. A classification and regression tree may result in unstable decision trees if improper modifications are made. If data have a complex structure, a classification and regression tree may not correctly model the data structure (Timofeev, 2004). A classification and regression tree can also result in an overly complex tree structure and in models that are better for prediction than estimation (Hosmer and Lemeshow, 1986). Additionally, practitioners may not be as familiar with regression tree analysis as other methods, and incorporating the resulting information into decision making may be difficult.

Logistic regression analysis is ideal for the naturalistic driving study because normal driving data can be used to account for exposure. This is important because, historically, it has been difficult to account for driver activity under a range of situations in order to determine if one situation is overrepresented. While naturalistic data provides the volume of data needed to assess the representation of various types of situations, researchers face the challenge of constructing meaningful equivalence classes to define these situations. Results of logistic regression can be expressed as odds ratios, which can easily be explained to lay persons and used by transportation agencies.

One disadvantage is that researchers, when applying logistic regression appropriately to the full naturalistic driving study, need to specify and identify events of interest. As a result, some relationships may not be uncovered.

Time series models are highly appropriate for naturalistic driving study data because they can account for dependencies between driver behaviors and other factors in time intervals. The main advantage of applying a time series analysis to naturalistic driving study data is that the analysis allows relationships between variables across time to be incorporated into the model. As a result, relationships such as driver distraction in previous time periods and probability of a red-light-running crash in a subsequent time period can be established. The biggest drawback to time series models is that they require the use of continuous (raw) data. Reducing variables not already included in the data sets at this level of data segmentation can be tremendously resource-intensive. Additionally, in the case of the example model presented above, only a few variables and a small data set were used and the model was still rather complicated. The results of time series analyses are also not common to highway agencies, and consequently it will be difficult to present results in a manner than can easily be used in decision making.

CHAPTER 7

# Summary

Lane departure crashes make up a substantial number of motor vehicle crashes and account for a disproportionate number of fatalities. Single-vehicle ROR crashes account for almost 39% of traffic fatalities. Two-vehicle head-on crashes result in 18% of noninterchange, nonintersection fatal crashes, with 75% occurring on undivided two-lane roadways. Hence, addressing lane departure crashes is a major safety goal in the United States.

Lane departure is a serious safety concern, yet the relationship between the factors that influence whether a vehicle departs its lane in the first place and the series of actions and events that determine the outcome are complex and not well understood. SHRP 2 is in the process of implementing a large-scale field study to collect naturalistic driving data at various locations throughout the United States. This study will result in a rich database that can be used to evaluate lane departures and better determine how the integration of driver behavior and roadway, environmental, and vehicle factors lead to different outcomes.

This research investigated which lane departure research questions can be answered when data from SHRP 2's full-scale naturalistic driving study data become available. The focus was to determine the necessary and available data factors that could fully answer the lane departure research questions and to conduct initial analyses of existing naturalistic driving study data to develop methods that can be used for the full-scale study. Analytical methods that could be used to answer those research questions were also explored. The focus of this research was on rural, two-lane, paved roadways.

The following paragraphs summarize the information provided in each chapter.

Chapter 1 provided background information and outlined the scope of the research.

Chapter 2 summarized the final research questions and provided the results of a literature review conducted to identify driver, roadway, environmental, and vehicle factors that have some correlation to lane departure crashes. Factors identified included horizontal and vertical curvature, roadway cross section, driveway density, illumination, weather, presence of rumble strips, roadway delineation and signing, pavement edge drop-off, vehicle type, speeding, influence of alcohol or drugs, driver age, and distraction.

Information about the factors known to influence the likelihood and outcome of lane departures was used to formulate a set of lane departure research questions that would be desirable to answer if the appropriate data were available. Data sets from existing naturalistic driving studies were explored, and the most current information about the data likely to be available in the full-scale study was reviewed (i.e., driver, roadway, environmental, and vehicle variables). Research questions were then categorized to distinguish those that were likely to be answered using data from the full-scale study and those that were not likely to be answered because of data limitations. Research questions addressed during the scope of this research were also identified.

Lane departure research questions that are not likely to be feasible in the full-scale study include those that require the following factors: alcohol or drug use by the driver (alcohol sensor will be present but will not record individual use), pavement surface friction measurements, pavement edge drop-off, and quantitative measures of rain, snow, or ice on the road. Identifying the feasible research questions in Chapter 2 required information provided throughout this report, but the information was summarized in Chapter 2 for the purpose of clarity.

Chapter 3 summarized the various data sets used in the research. A description of common data terms was also provided.

Chapter 4 identified data elements that are expected to be necessary for answering the lane departure research questions, based on a survey of the available literature, as well as on the team's expertise on lane departure issues. The accuracy, frequency, and resolution of each data element was determined and described. The availability of the data in the UMTRI and

VTTI databases was reviewed and the limitations described. The team also reviewed the available documentation of the work for SHRP 2 Safety Projects S03 (Roadway Measurement System Evaluation), S04A (Roadway Information Database Developer, Technical Coordination, and Quality Assurance for Mobile Data Collection), and S04B (Mobile Data Collection). Based on these sources, the accuracy, frequency, and resolution of data that are expected to be available to answer the lane departure research questions in the full-scale study were evaluated. The team identified limitations and provided feedback to SHRP 2, as described in Chapter 4. Data elements were also prioritized, because resource limitations in the full-scale study will constrain data collection.

Chapter 5 discussed potential lane departure surrogates that can be obtained from naturalistic driving study data. Literature regarding crash surrogates was summarized, and a methodological approach for selecting and applying crash surrogates was outlined. Existing naturalistic driving study data were also evaluated to determine starting points for setting triggers that would identify lane departure events. The ways normal driving data may be partitioned were also evaluated using existing data. Lateral offset was compared for several driving situations. Differences were noted between driving on a tangent and on a left- or right-hand curve, between nighttime and daytime driving, and between individual drivers. As indicated in the chapter, differences are expected in what constitutes normal driving behavior. This analysis provided some guidance on stratifying normal driving by relevant variables.

Chapter 6 described four analytical approaches that can be used to evaluate naturalistic driving study data and answer lane departure research questions. Several exploratory analysis methods were applied to data extracted from existing naturalistic driving studies to demonstrate ways in which lane departure research questions could be answered in the full-scale study. The intent of the analyses was to demonstrate different analysis methods that could be used to analyze the data that will result from the full-scale study.

A data sampling approach (developed by the SHRP 2 Safety Project S02 researchers) was described, and four analysis methods were presented. The four approaches included (1) a data mining approach using classification and regression tree analysis, (2) simple odds ratio and logistic regression, (3) logistic regression for correlated data that accounts for repeated sampling among observations (e.g., repeated sampling for the same driver, trip), and (4) a time series analysis.

Three of these methods were used to evaluate existing naturalistic driving study data, and one method expanded on a varied logistic regression approach that may be better suited to the data from the full-scale study. Data were available from the UMTRI RDCW FOT that contained a number of nonconflict lane departures and samples of normal driving. Methods 1 and 2 ([1] classification and regression tree and [2] simple odds ratio and logistic regression) evaluated the likelihood of a left- or right-side lane departure. A sample-based approach was used in the classification and regression tree analysis, and an event-based approach was used for the logistic regression. Although available sample sizes were limited, both methods resulted in similar results. Both indicated that curve radius, driver age, and type of shoulder were relevant in explaining lane departures. The logistic regression also indicated that both left- and right-side lane departures were more likely to occur at night and were less likely to occur as lane width increased. The model for left-side lane departures indicated that male drivers were more likely than female drivers to be involved in a lane departure, and the model for right-side lane departures indicated that lane departures are more likely on roadway sections with a higher density of lane departure crashes and for drivers who spend more time traveling 10 mph or more over the posted speed limit.

The fourth method, time series analysis, used continuous data to develop a model to predict offset as a function of several vehicle kinematic variables. The method was developed and explained in such a way that it could be adapted to the full-scale study to include various explanatory variables, including driver behavior. This approach allows information, such as driver distraction in previous time periods, to be incorporated into the model.

Appendices A and B describe the protocols, methods, and variable descriptions used to extract data from the UMTRI RDCW FOT and VTTI 100-car naturalistic driving study data. The method used to extract the data provided a framework that can be used by other researchers working with the full-scale study. Data were extracted manually, which consumed a large amount of resources. However, the framework can be used to automate the extraction of some data. Appendix A also provides a discussion about how lane departures were identified in the UMTRI data set.

# References


American Association of State Highway and Transportation Officials (AASHTO). 1996. *Roadside Design Guide.* American Association of State Highway and Transportation Officials, Washington, D.C.

American Association of State Highway and Transportation Officials (AASHTO). 2001. *A Policy on Geometric Design of Highways and Streets.* American Association of State Highway and Transportation Officials, Washington, D.C.

American Association of State Highway and Transportation Officials (AASHTO). 2004. *A Policy on Geometric Design of Highways and Streets.* American Association of State Highway and Transportation Officials, Washington, D.C.

Abboud, N. K., and B. L. Bowman. 2001. Evaluation of Two- and Four-Foot Shoulders on Two-Lane State Routes. *ITE Journal,* Vol. 71, No. 6, pp. 34–39.

Agresti, A. 2002. *Categorical Data Analysis,* 2nd ed. John Wiley & Sons, Inc., Hoboken, N.J.

Allen, B. L., B. T. Shin, and P. J. Cooper. 1978. Analysis of Traffic Conflicts and Collisions. *Transportation Research Record 667,* TRB, National Research Council, Washington, D.C., pp. 67–76.

Anderson, I. B., K. M. Bauer, D. W. Harwood, and K. Fitzpatrick. 1999. Relationship to Safety of Geometric Design Consistency Measures for Rural Two-Lane Highways. *Transportation Research Record: Journal of the Transportation Research Board, No. 1658,* TRB, National Research Council, Washington, D.C., pp. 43–51.

Archer, J. 2001. *Traffic Conflict Technique: Historical to Current State-of-the-Art.* Institutionen főr Infrastruktur KTH, Stockholm, Sweden.

Ayers, G., B. Wilson, and J. LeBlanc. 2004. Methodology for Identifying Vehicle Movement for Analysis of Field Operational Tests. *Transportation Research Record: Journal of the Transportation Research Board, No. 1886,* Transportation Research Board of the National Academies, Washington, D.C., pp. 92–100.

Bar-Shalom, Y., X.-R. Li, and T. Kirubarajan. 2001. *Estimation with Applications to Tracking and Navigation: Theory, Algorithms, and Software.* John Wiley and Sons, Inc., New York.

Billion, C. E., and N. C. Parson. 1962. Median Accident Study—Long Island, N.Y. *Highway Research Board Bulletin 308,* HRB, National Research Council, Washington, D.C., pp. 64–79.

Bonneson, J. A., K. H. Zimmerman, and M. P. Pratt. 2004. *Red-Light-Running Handbook: An Engineer's Guide to Reducing Red-Light-Related Crashes.* Texas Transportation Institute, College Station, Tex.

Box, P. C. 1972. Comparison of Accidents and Illumination. *Highway Research Record 416,* HRB, National Research Council, Washington, D.C., pp. 1–9.

Box, P. C. 1989. Major Road Accident Reduction by Modern Illumination. *Transportation Research Record 1247,* TRB, National Research Council, Washington, D.C., pp. 32–38.

Boyle, L. B., S. Hallmark, J. D. Lee, D. V. McGehee, D. M. Neyens, and N. J. Ward. 2010. *SHRP 2 Phase II Report: Integration of Analysis Methods and Development of Analysis Plan.* Transportation Research Board of the National Academies, Washington, D.C.

Burgett, A., and K. Gunderson. 2001. Crash Prevention Boundary for Road Departure Crashes: Derivation. Report HS 809 399. *NHTSA Research Note,* Washington, D.C.

Burgett, A., and R. J. Miller, Jr. 2001. *A New Paradigm for Rear-End Crash Prevention Driving Performance.* National Highway Traffic Safety Administration.

Campbell, K. 2009. SHRP 2 Safety: Making a Significant Improvement in Highway Safety. Presented at 4th Safety Research Symposium of the Transportation Research Board, Washington, D.C. http://onlinepubs.trb.org/onlinepubs/shrp2/2009SRS_CampbellThur.PDF. Accessed March 28, 2011.

Campbell, K. L., M. Lepofsky, and A. Bittner. 2003. *F-SHRP Web Document 2: Detailed Planning for Research on Making a Significant Improvement in Highway Safety. Study 2—Safety.* Transportation Research Board, Washington, D.C. http://onlinepubs.trb.org/onlinepubs/f-shrp/f-shrp_webdoc_2.pdf. Accessed July 5, 2011.

Chein, J., D. Albert, L. O'Brien, K. Uckert, and L. Steinberg. 2011. Peers Increase Adolescent Risk Taking by Enhancing Activity in the Brain's Reward Circuitry. *Developmental Science,* Vol. 14, No. 2, pp. F1–F10.

Chin, H., and S. Quek. 1997. Measurement of Traffic Conflicts. *Safety Science,* Vol. 26, No. 3, pp. 169–185.

Chin, H., S. Quek, and R. Chou. 1992. Quantitative Examination of Traffic Conflicts. *Transportation Research Record 1376,* TRB, National Research Council, Washington, D.C., pp. 87–91.

Collier, K., B. Carey, E. Grusy, C. Marjaniemi, and D. Sautter. 1998. *A Perspective on Data Mining.* Center for Data Insight, Flagstaff, Ariz. http://insight.nau.edu/downloads/DM%20Perspective%20v2.pdf. Accessed March 27, 2011.

Council, F. M. 1998. Safety Benefits of Spiral Transitions on Horizontal Curves on Two-Lane Rural Roads. *Transportation Research Record 1635,* TRB, National Research Council, Washington, D.C., pp. 10–17.

Corkle, J., M. Marti, and D. Montebello. 2001. *Synthesis on the Effectiveness of Rumble Strips.* Minnesota Department of Transportation, St. Paul, Minn.

Dang, Q., S. Mazumdar, and P. R. Houck. 2008. Sample Size and Power Calculations Based on Generalized Linear Mixed Models with Correlated Binary Outcomes. *Computer Methods and Programs in Biomedicine,* Vol. 91, No. 2, pp. 122–127.

Deng, Z., J. N. Ivan, and P. E. Garder. 2006. Analysis of Factors Affecting the Severity of Head-On Crashes: Two-Lane Rural Highways in Connecticut. *Transportation Research Record: Journal of the Transportation Research Board, No. 1953,* Transportation Research Board of the National Academies, Washington, D.C., pp. 137–146.

Dingus, T., J. Hankey, J. Antin, S. Lee, E. Eichelberger, S. Klauer, J. Sudweeks, C. Gaylord, D. McGraw, S. McLaughlin, M. Mollenhauer, M. Perez, A. Petersen, A. Alden, P. Madison, T. Gordon, D. LeBlanc, H. Sardar, D. Blower, J. Sayer, M. Barnes, A. Thurtell, J. Patel, J. Campbell, C. Richard, and J. Devonshire. 2008a. *SHRP 2 Project S05: Design of the In-Vehicle Driving Behavior and Crash Risk Study. Task 9: Data System Interim Report (Task 6: Driver Face and Other Video Recording And Processing).* Virginia Tech Transportation Institute, Blacksburg, Va.

Dingus, T., J. Hankey, J. Antin, S. Lee, E. Eichelberger, S. Klauer, J. Sudweeks, C. Gaylord, D. McGraw, S. McLaughlin, M. Mollenhauer, M. Perez, A. Petersen, A. Alden, P. Madison, T. Gordon, D. LeBlanc, H. Sardar, D. Blower, J. Sayer, M. Barnes, A. Thurtell, J. Patel, J. Campbell, C. Richard, and J. Devonshire. 2008b. *SHRP 2 S05: Design of The In-Vehicle Driving Behavior And Crash Risk Study. Task 9: Data System Interim Report (Task 7: Data Items And Instrumentation Package Specifications—Appendices A And B).* Virginia Tech Transportation Institute, Blacksburg, Va.

Dingus, T. A., S. G. Klauer, V. L. Neale, A. Petersen, S. E. Lee, J. Sudweeks, M. A. Perez, J. Hankey, D. Ramsey, S. Gupta, C. Bucher, Z. R. Doerzaph, J. Jermeland, and R. R. Knipling. 2006. *The 100-Car Naturalistic Driving Study: Phase II—Results of the 100-Car Field Experiment.* NHTSA, U.S. Department of Transportation.

Dissanayake, S. 2003. Young Drivers and Run-Off-The-Road Crashes. *Proc., 2003 Mid-Continent Transportation Research Symposium,* Ames, Iowa. www.ctre.iastate.edu/pubs/midcon2003/dissanayakeyoung.pdf. Accessed March 27, 2011.

Dixon, K. 2005. *Southeastern United States Fatal Crash Study.* Georgia Department of Transportation, Forest Park, Ga. www.dot.state.ga.us/doingbusiness/research/Documents/reports/SESummaryFatalCrash.pdf. Accessed March 27, 2011.

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

Donnell, E. T., and J. M. Mason, Jr. 2006. Predicting the Frequency of Median Barrier Crashes on Pennsylvania Interstate Highways. *Accident Analysis and Prevention,* Vol. 38, pp. 590–599.

Elvik, R. 1995. Meta-Analysis of Evaluations of Public Lighting as Accident Countermeasure. *Transportation Research Record 1485,* TRB, National Research Council, Washington, D.C., pp. 112–123.

Ferguson, S. A. 2003. Other High-Risk Factors for Young Drivers: How Graduated Licensing Does, Doesn't, or Could Address Them. *Journal of Safety Research,* Vol. 34, No. 1, pp. 71–77.

FHWA. 1982. *Synthesis of Safety Research Related to Traffic Control and Roadway Elements.* Report FHWA TS-82-232. FHWA, U.S. Department of Transportation.

FHWA. 2004. *The Safety Edge: Preventing Crashes Caused by Unsafe Pavement Edge Drop-offs.* FHWA, U.S. Department of Transportation. www.fhwa.dot.gov/resourcecenter/teams/safety/saf_12TSE.pdf. Accessed July 6, 2011.

FHWA. 2007. *Manual on Uniform Traffic Control Devices,* 2003 ed. with revisions 1 and 2. FHWA, U.S. Department of Transportation.

FHWA, U.S. Department of Transportation. 2009. Red-Light Running. http://safety.fhwa.dot.gov/intersection/redlight/. Accessed July 6, 2011.

Fink, K. L., and R. A. Krammes. 1995. Tangent Length and Sight Distance Effects on Accident Rates at Horizontal Curves on Rural Two-Lane Highways. *Transportation Research Record 1500,* TRB, National Research Council, Washington, D.C., pp. 162–168.

Forbes, G., M. Eng, and P. Eng. 2003. *Synthesis of Safety for Traffic Operations.* Transport Canada, Ottawa, Ontario, Canada.

Garber, N. J., and A. A. Ehrhart. 2000. Effect of Speed, Flow, and Geometric Characteristics on Crash Frequency for Two-Lane Highways. *Transportation Research Record: Journal of the Transportation Research Board, No. 1717,* TRB, National Research Council, Washington, D.C., pp. 76–83.

Garber, N. J., J. S. Miller, S. Eslambolchi, R. Khandelwal, K. M. Mattingly, K. M. Sprinkle, and P. L. Wachendorf. 2005. *An Evaluation of Red Light Camera (Photo-Red) Enforcement Programs in Virginia: A Report in Response to a Request by Virginia's Secretary of Transportation.* Virginia Transportation Research Council, Charlottesville, Va.

Garder, P., and M. Davies. 2006. Safety Effect of Continuous Shoulder Rumble Strips on Rural Interstates in Maine. *Proc., 85th Annual Meeting of the Transportation Research Board,* Washington, D.C.

Gettman, D., and L. Head. 2003. *Surrogate Safety Measures from Traffic Simulation Models.* Report FHWA-RD-03-050. FHWA, U.S. Department of Transportation.

Gillespie, T. D. 1992. *Fundamentals of Vehicle Dynamics.* Society of Automotive Engineers, Warrendale Pa.

Hallmark, S. L., R. Guensler, and I. Fomunung. 2002. Characterizing On-Road Variables that Affect Passenger Vehicle Modal Operation. *Transport and Environment,* Vol. 7 (March), pp. 81–98.

Hallmark, S. L., D. Veneziano, T. McDonald, J. Graham, R. Patel, and F. Council. 2006. *Safety Impacts of Pavement Edge Drop-off.* AAA Foundation for Traffic Safety, Washington. D.C.

Hallmark, S. L., L. Boyle, A. Carriquiry, T. Maze, N. Hawkins, O. Smadi, R. Souleyrette, T. McDonald, Z. Hans, Y. Y. Hsu, A. Mudgal, and E. Fitzsimmons. 2008. *Development of Analytical Tools to Evaluate Road Departure Crashes Using Naturalistic Driving Study Data: Final Report, Phase I.* Center for Transportation Research and Education, Iowa State University, Ames, Iowa.

Hanley, K. E., A. R. Gibby, and T. C. Ferrara. 2000. Analysis of Accident-Reduction Factors on California State Highways. *Transportation Research Record: Journal of the Transportation Research Board, No. 1717,* TRB, National Research Council, Washington, D.C., pp. 36–45.

Hauer, E., F. M. Council, and Y. Mohammedshah. 2004. Safety Models for Urban Four-Lane Undivided Road Segments. *Transportation Research Record: Journal of the Transportation Research Board, No. 1897,* Transportation Research Board of the National Academies, Washington, D.C., pp. 96–105.

Hayward, J. 1972. Near-Miss Determination Through Use of a Scale of Danger. *Highway Research Record 384,* HRB, National Research Council, Washington, D.C., pp. 24–34.

Heimbach, C. L., W. W. Hunter, and G. C. Chao. 1974. Paved Highway Shoulders and Accident Experience. *Journal of the Transportation Engineering Division,* Vol. 100, No. 4, pp. 889–907.

Hickey, J. J. 1997. Shoulder Rumble Strip Effectiveness: Drift-off-Road Accident Reductions on the Pennsylvania Turnpike. *Transportation Research Record 1573,* TRB, National Research Council, Washington, D.C., pp. 105–109.

Highway Safety Research Center. 2000. *Preliminary Evaluation of the North Carolina Graduated Driver Licensing System: Effects on Young*

*Driver Crashes.* University of North Carolina, Chapel Hill, N.C. www.hsrc.unc.edu/pdf/2000/gdlpdf.pdf. Accessed March 27, 2011.

Hosmer, D. W., and S. Lemeshow. 1986. Classification Trees. http://wekadocs.com/node/2. Accessed February 12, 2010.

Hosmer, D. W., and S. Lemeshow. 2000. *Applied Logistic Regression,* 2nd ed. John Wiley and Sons, Inc., New York.

Hsieh, F. Y., D. A. Bloch, and M. D. Larsen. 1998. A Simple Method of Sample Size Calculation for Linear and Logistic Regression. *Statistics in Medicine,* Vol. 17, pp. 1623–1634.

Humphreys, J. B., and J. A. Parham. 1994. *The Elimination or Mitigation of Hazards Associated with Pavement Edge Drop-offs During Roadway Resurfacing.* AAA Foundation for Traffic Safety, Washington, D.C.

Insightful Corporation. 2007. *S-Plus 8: Guide to Statistics, Volume 2.* Seattle, Wash. www.msi.co.jp/splus/support/download/V8/statman2.pdf. Accessed March 27, 2011.

Insurance Institute for Highway Safety. 2009. Q & As: Rollover and Roof Crush. www.iihs.org/research/qanda/rollover.html. Accessed March 27, 2011.

International Commission on Illumination. 1992. *CIE Technical Report No. 93: Road Lighting as an Accident Countermeasure.* Commission Internationale de l'Eclairage, Vienna, Austria.

Khattak, A. J., and J. E. Hummer. 1998. Analysis of Run-off-the-Road Collisions on Two-Lane Rural Highways in North Carolina. *ITE Journal,* Vol. 68, No. 10, p. 18.

Klauer, S. G., T. A. Dingus, V. L. Neale, J. D. Sudweeks, and D. J. Ramsey. 2006. *The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data.* Report DOT HS 810 594. NHTSA, U.S. Department of Transportation.

Kohinoor, K., and S. Weeks. 2009. The Sound of Safety. *Public Roads,* Vol. 72, No. 4. www.fhwa.dot.gov/publications/publicroads/09janfeb/02.cfm. Accessed March 27, 2011.

Krull, K. A., A. J. Khattak, and F. M. Council. 2000. Injury Effects of Rollovers and Events Sequences in Single-Vehicle Crashes. *Transportation Research Record: Journal of the Transportation Research Board, No. 1717,* TRB, National Research Council, Washington, D.C., pp. 36–54.

Lamm, R., E. M. Choueiri, J. C. Hayward, and A. Paluri. 1988. Possible Design Procedure to Promote Design Consistency in Highway Geometric Design on Two-Lane Rural Roads. *Transportation Research Record 1195,* TRB, National Research Council, Washington, D.C., pp. 111–122.

LeBlanc, D., J. Sayer, C. Winkler, R. Ervin, S. Bogard, J. Devonshire, M. Mefford, M. Hagan, Z. Bareket, R. Goodsell, and T. Gordon. 2006. *Road Departure Crash Warning System Field Operational Test: Methodology and Results.* University of Michigan Transportation Research Institute, Ann Arbor, Mich.

Lewin, I., P. Box, and R. E. Stark. 2003. *Roadway Lighting: An Investigation and Evaluation of Three Different Light Sources.* Arizona Department of Transportation, Phoenix, Ariz. www.azdot.gov/TPD/ATRC/publications/project_reports/PDF/AZ522.pdf. Accessed February 15, 2010.

Liu, G., and K. Y. Liang. 1997. Sample Size Calculations for Studies with Correlated Observations. *Biometrics,* Vol. 53, No. 3, pp. 937–947.

Maas, C. J. M., and J. J. Hox. 2005. Sufficient Sample Sizes for Multilevel Modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences,* Vol. 1, No. 3, pp. 86–92.

Mammar, S., S. Glaser, and M. Netto. 2006. Time to Line Crossing for Lane Departure Avoidance: A Theoretical Study and an Experimental Setting. *IEEE Transactions on Intelligent Transportation Systems,* Vol. 7, No. 2, pp. 226–241.

Mathsoft. 1997. *S-Plus 4 Guide to Statistics.* Mathsoft, Inc., Cambridge, Mass.

Maze, T. H., and G. Burchett. 2004. *Safety Performance of Rural Expressway Intersections.* Center for Transportation Research, Iowa State University, Ames, Iowa.

Maze, T. H., Z. Hans, M. Agarwal, and G. D. Burchett. 2006. Weather Impacts on Traffic Safety and Operations. *Proc., 2006 Annual Meeting and Exhibit Compendium of Technical Papers.* Institute of Transportation Engineers, Washington, D.C.

McGehee, D. V., C. Carney, M. Raby, J. D. Lee, M. Reyes. 2007. The Impact of an Event-Triggered Video Intervention on Rural Teenage Driving. *Proc., 4th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design,* Stevenson, Wash.

McGinnis, R. G., M. J. Davis, and E. A. Hathaway. 2001. Longitudinal Analysis of Fatal Run-off-Road Crashes, 1975 to 1997. *Transportation Research Record: Journal of the Transportation Research Board, No. 1746,* TRB, National Research Council, Washington, D.C., pp. 47–58.

McLaughlin, S. B., J. M. Hankey, S. G. Klauer, and T. A. Dingus. 2009. *Contributing Factors to Run-off-Road Crashes and Near-Crashes.* Report DOT HS 811 079. NHTSA, U.S. Department of Transportation.

Miaou, S. P., and H. Lum. 1993. Statistical Evaluation of the Effects of Highway Geometric Design on Truck Accident Involvements. *Transportation Research Record 1407,* TRB, National Research Council, Washington, D.C., pp. 11–24.

Miles, J. D., P. J. Carlson, M. P. Pratt, and T. D. Thompson. 2005. *Traffic Operational Impacts of Transverse, Centerline, and Edgeline Rumble Strips.* Texas Transportation Institute, College Station. http://tti.tamu.edu/documents/0-4472-2.pdf. Accessed March 27, 2011.

Mounce, J. M. 1981. Driver Compliance with Stop-Sign Control at Low-Volume Intersections. *Transportation Research Record 808,* TRB, National Research Council, Washington, D.C., pp. 30–37.

Mohamedshah, Y. M., F. F. Paniatie, and A. G. Hobeika. 1993. Truck Accident Models for Interstates and Two-Lane Rural Roads. *Transportation Research Record 1407,* TRB, National Research Council, Washington, D.C., pp. 35–41.

Morgan, J., R. Dougherty, A. Hilchie, and B. Carey. 2009. *Sample Size and Modeling Accuracy with Decision Tree Based Data Mining Tools.* Center for Data Insight, Northern Arizona University, Flagstaff, Ariz. http://insight.nau.edu/downloads/Sample%20Size%20and%20Modeling%20Accuracy.pdf. Accessed March 27, 2011.

Nagpual, P. S. 2009. Chapter 10: Classification and Regression Trees. In *Guide to Advanced Data Analysis.* www.unesco.org/webworld/idams/advguide/Chapt10.htm. Accessed March 27, 2011.

NHTSA. 2005a. NHTSA Vehicle Safety Rulemaking and Supporting Research Priorities: Calendar Years 2005–2009. www.nhtsa.gov/cars/rules/rulings/priorityplan-2005.html. Accessed July 6, 2011.

NHTSA. 2005b. Traffic Safety Facts 2003: A Compilation of Motor Vehicle Crash Data from the Fatality Analysis Reporting System and the General Estimates System. www-nrd.nhtsa.dot.gov/Pubs/809767.PDF. Accessed March 27, 2011.

Neuman, T. R., R. Pfefer, K. L. Slack, K. K. Hardy, H. McGee, L. Prothe, K. Eccles, and F. Council. 2003. *NCHRP Report 500: Guidance for Implementation of the AASHTO Strategic Highway Safety Plan. Volume 4: A Guide for Addressing Head-On Collisions.* Transportation Research Board of the National Academies, Washington, D.C.

Noyce, D., and V. V. Elango. 2004. Safety Evaluation of Centerline Rumble Strips: Crash and Driver Behavior Analysis. *Transportation Research Record: Journal of the Transportation Research Board, No. 1862,* Transportation Research Board of the National Academies, Washington, D.C., pp. 44–53.

Outcalt, W. 2001. *Centerline Rumble Strips.* Report CDOT-DTD-R-2001-8. Colorado Department of Transportation, Denver, Colo.

Oxley, J., B. Corben, S. Koppel, B. Fildes, N. Jacques, M. Symmons, and I. Johnston. 2004. *Cost-Effective Infrastructure Measures on Rural Roads.* Report 217. Monash University Accident Research Centre, Victoria, Australia.

Pomerleau, D., T. Jochem, C. Thorpe, P. Batavia, D. Pape, J. Hadden, N. McMillan, N. Brown, and J. Everson. 1999. *Run-off-Road Collision Avoidance Using IVHS Countermeasures.* Report DOT HS 809 170. NHTSA, U.S. Department of Transportation.

Perrillo, K. 1998. *The Effectiveness and Use of Continuous Shoulder Rumble Strips.* FHWA, U.S. Department of Transportation.

Persaud, B., G. Bahar, C. J. Mollett, and C. Lyon. 2004. Safety Evaluation of Permanent Raised Snow-Plowable Pavement Markers. *Transportation Research Record: Journal of the Transportation Research Board, No. 1897,* Transportation Research Board of the National Academies, Washington, D.C., pp. 148–155.

Porter, R. J., E. T. Donnel, and K. Mahoney. 2004. Evaluation of Effects of Centerline Rumble Strips on Lateral Vehicle Placement and Speed. *Transportation Research Record: Journal of the Transportation Research Board, No. 1862,* Transportation Research Board of the National Academies, Washington, D.C., pp. 10–16.

Pratt, M., J. Miles, and P. Carlson. 2006. Evaluation of Operational Impacts of Installation of Centerline and Edge Line Rumble Strips. *Transportation Research Record: Journal of the Transportation Research Board, No. 1973,* Transportation Research Board of the National Academies, Washington, D.C., pp. 80–88.

Preston, H. 2009. Horizontal Curves: A New Method for Identifying At-Risk Locations for Safety Investment. *Proc., 2009 Mid-Continent Transportation Research Symposium,* Ames, Iowa. www.intrans.iastate.edu/pubs/midcon2009/PrestonHorizontal.pdf. Accessed March 27, 2011.

Retting, R. A., S. A. Ferguson, and C. M. Farmer. 2007. *Reducing Red Light Running Through Longer Yellow Signal Timing and Red Light Camera Enforcement: Results of a Field Investigation.* Insurance Institute for Highway Safety, Arlington, Va.

Rice, T. M., C. Peek-Asa, and J. F. Kraus. 2004. Effects of the California Graduated Driver Licensing Program. *Journal of Safety Research,* Vol. 35, No. 4, pp. 375–381.

Roberts, C. A., S. Washington, and J. Leonard. 1999. Forecasting Dynamic Vehicular Activity on Freeways: Bridging the Gap Between Travel Demand and Emerging Emissions Models. *Proc., 78th Annual Meeting of the Transportation Research Board,* Washington, D.C.

Russell, E. R., and M. J. Rys. 2005. *NCHRP Synthesis 339: Centerline Rumble Strips.* Transportation Research Board of the National Academies, Washington, D.C.

Sabey, B. E., and H. D. Johnson. 1973. *Road Lighting and Accidents: Before and After Studies on Trunk Road Sites.* TRRL Report LR 586. Transport and Road Research Laboratory, Department of the Environment, Crowthorne, Berkshire, United Kingdom.

Salman, N., and K. Al-Maita. 1995. Safety Evaluation at Three-Leg, Unsignalized Intersections by Traffic Conflict Technique. *Transportation Research Record 1485,* TRB, National Research Council, Washington, D.C., pp. 177–185.

SearchSQLServer.com. 2009. Data Mining. http://searchsqlserver.techtarget.com/sDefinition/0,,sid87_gci211901,00.html. Accessed July 2009.

Shankar, V. N., R. B. Albin, J. C. Milton, and F. L. Mannering. 1998. Evaluating Median Crossover Likelihoods with Clustered Accident Counts: An Empirical Inquiry Using the Random Effects Negative Binomial Model. *Transportation Research Record 1635,* TRB, National Research Council, Washington, D.C., pp. 44–57.

Shankar, V. N., S. Chayanan, S. Sittikariya, M.-B. Shyu, N. K. Juvva, and J. C. Milton. 2004. Marginal Impacts of Design, Traffic, Weather, and Related Interactions on Roadside Crashes. *Transportation Research Record: Journal of the Transportation Research Board, No. 1897,* Transportation Research Board of the National Academies, Washington, D.C., pp. 156–163.

Shankar, V., P. P. Jovanis, J. Aguero-Valverde, and F. Gross. 2008. Analysis of Naturalistic Driving Data: Prospective View on Methodological Paradigms. *Transportation Research Record: Journal of the Transportation Research Board, No. 2061,* Transportation Research Board of the National Academies, Washington, D.C., pp. 1–8.

Smith, E. B., and J. N. Ivan. 2005. Evaluation of Safety Benefits and Potential Crash Migration Due to Shoulder Rumble Strip Installation on Connecticut Freeways. *Transportation Research Record: Journal of the Transportation Research Board, No. 1908,* Transportation Research Board of the National Academies, Washington, D.C., pp. 104–113.

Songchitruksa, P., and A. Tarko. 2004. *Using Imaging Technology to Evaluate Highway Safety: Right-angle Collisions.* Joint Transportation Research Program, Purdue University, West Lafayette, Ind.

Songchitruksa, P., and A. Tarko. 2006. The Extreme Value Theory Approach to Safety Estimation. *Accident Analysis and Prevention,* Vol. 38, pp. 811–822.

Sosslau, A. B., A. B. Hassam, M. M. Carter, and G. V. Wickstrom. 1978. *NCHRP Report 187: Quick-Response Urban Travel Estimation Techniques and Transferable Parameters—User's Guide.* TRB, National Research Council, Washington, D.C.

Souleyrette, R. 2001. *Paved Shoulders on Primary Highways in Iowa: An Analysis of Shoulder Surfacing Criteria, Costs, and Benefits.* Center for Transportation Research and Education, Ames, Iowa.

Spainhour, L. K., D. Brill, J. O. Sobanjo, J. Wekezer, and P. V. Mtenga. 2005. *Evaluation of Traffic Crash Fatality Causes and Effects: A Study of Fatal Traffic Crashes in Florida from 1998–2000 Focusing on Heavy Truck Crashes.* Department of Civil Engineering, Florida A&M University and Florida State University, Tallahassee, Fla.

StatSoft. 2008. Classification and Regression Trees. www.statsoft.com/TEXTBOOK/stcart.html. Accessed February 15, 2010.

Steyer, R., A. Sossoumihen, and G. Weise. 2000. Traffic Safety on Two-Lane Rural Roads: New Concepts and Findings. *Proc., 2nd International Symposium on Highway Geometric Design,* Mainz, Germany, pp. 299–312.

Sun, X., J. Park, V. O. Tekell, Jr., and N. S. Ludington. 2007. Evaluation of Pavement Edge Lines on Narrow Highways with Low Traffic Volume in Louisiana. *Transportation Research Record: Journal of the Transportation Research Board, No. 1989,* Transportation Research Board of the National Academies, Washington, D.C., pp. 115–122.

Szabo, S., and B. Wilson. 2004. Application of a Crash Prevention Boundary Metric to a Road Departure Warning System. *Proc., 2004 Performance Metrics for Intelligent Systems Workshop,* Gaithersburg, Md.

Tan, P. N., M. Steinbach, and V. Kumar. 2006. *Introduction to Data Mining.* Addison-Wesley, Upper Saddle River, N.J.

Taylor, W. C., G. Abu-Lebdeh, and S. Rai. 2005. Effect of Continuous Shoulder Rumble Strips and Pavement Marking on Lateral Placement of Vehicles. *Transportation Research Record: Journal of the Transportation Research Board, No. 1911,* Transportation Research Board of the National Academies, Washington, D.C., pp. 105–112.

Timofeev, R. 2004. Classification and Regression Trees (CART): Theory and Applications. Master's Thesis. Center for Applied Statistics and Economics, Humboldt University, Berlin, Germany.

Torbic, D. J., D. W. Harwood, D. K. Gilmore, R. Pfefer, T. R. Neuman, K. L. Slack, and K. K. Hardy. 2004. *NCHRP Report 500: Guidance for Implementation of the AASHTO Strategic Highway Safety Plan.*

*Volume 7: A Guide for Reducing Collisions on Horizontal Curves.* Transportation Research Board of the National Academies, Washington, D.C. http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_rpt_ 500v7.pdf. Accessed March 27, 2011.

Transportation Research Board. 2007. Strategic Highway Research Program. www.trb.org/SHRP2. Accessed February 15, 2010.

Tsyganov, A. R., R. B. Machemehl, N. M. Warrenchuk, and Y. Wang. 2006. *Before-After Comparison of Edgeline Effects on Rural Two-Lane Highways.* Texas Department of Transportation, Austin, Tex.

Ulmer, R. G., A. F. Williams, and D. F. Preusser. 1997. Crash Involvements of 16-Year-Old Drivers. *Journal of Safety Research,* Vol. 28, No. 2, pp. 97–103.

Vogt, A., and J. Bared. 1998. Accident Models for Two-Lane Rural Segments and Intersections. *Transportation Research Record 1635,* TRB, National Research Council, Washington, D.C., pp. 18–29.

Williams, A. F. 2001. *Teenage Passengers in Motor Vehicle Crashes: A Summary of Current Research.* Insurance Institute for Highway Safety, Arlington, Va.

Williams, A. F. 2003. Graduated Licensing in the United States. *ITE Journal,* Vol. 73, No. 9, pp. 28–30.

Williams, A. F., A. K. Lund, D. F. Preusser, and R. D. Bloomberg. 1997. Results of a Seat Belt Use Law Enforcement and Publicity Campaign in Elmira, New York. *Accident Analysis and Prevention,* Vol. 19, pp. 243–249.

Wolf, J., R. Guensler, S. Washington, and W. Bachman. 1998. High-Emitting Vehicle Characterization Using Regression Tree Analysis. *Transportation Research Record 1641,* TRB, National Research Council, Washington, D.C., pp. 58–65.

Zegeer, C. H., R. C. Deen, and J. G. Mayes. 1981. Effect of Lane and Shoulder Widths on Accident Reduction on Rural Two-Lane Roads. *Transportation Research Record 806,* TRB, National Research Council, Washington, D.C., pp. 33–43.

Zegeer, C. V., R. Stewart, F. M. Council, D. W. Reinfurt, and E. Hamilton. 1992. Safety Effects of Geometric Improvements on Horizontal Curves. *Transportation Research Record 1356,* TRB, National Research Council, Washington, D.C., pp. 11–19.

Zhang, C., and J. N. Ivan. 2005. Effects of Geometric Characteristics on Head-On Crash Incidence on Two-Lane Roads in Connecticut. *Transportation Research Record: Journal of the Transportation Research Board, No. 1908,* Transportation Research Board of the National Academies, Washington, D.C., pp. 159–171.

Zimmer, R. A., and D. L. Ivey. 1982. *Pavement Edges and Vehicle Stability: A Basis for Maintenance Guidelines.* Report 328-1. Texas Transportation Institute, Texas A&M University, College Station, Tex.

APPENDIX A

# Methodology for Extraction of Data Elements from the UMTRI Naturalistic Driving Study Data Set

This appendix includes description of the methodology to extract variables necessary to answer lane departure research questions using the UMTRI naturalistic driving study data set. The focus of the analysis is lane departures on rural, two-lane roadways.

All the data were extracted and reduced manually, since the team was exploring and evaluating what was available with the data. This took a tremendous amount of resources. Once researchers spend some time familiarizing themselves with the data in the full-scale study, it is expected that methods can be developed to automate some of the data extraction.

The data sets used are described in detail in Chapter 3.

## Data Preparation

The UMTRI data were provided in the form of continuous vehicle data. Each row of data represented 0.1 s of driving. Vehicle data were provided by vehicle alert type as indicated in the section "University of Michigan Transportation Research Institute Field Operational Test In-Vehicle Data" at the beginning of Chapter 3. The road departure crash warning (RDCW) system included six levels of alerts to indicate to the driver that he or she was about to leave his or her lane or was traveling too fast on a curve. The alerts included a right- and left-lane departure cautionary alert, right- and left-lane departure imminent alerts, cautionary curve speed warning, and imminent curve speed warning. A seventh designation was used to indicate that a vehicle was negotiating a curve, but no alert has been included for this.

Data were provided on 44 drivers. Data were divided by alert type. Over 2,000 alerts/curves were received. Continuous data were provided, with a resolution of 10 Hz (one row represents 0.1 s). A total of 1,506,525 rows of data were received. Data were provided for 30 s before an alert was recorded to approximately 30 s after (called a trace). Instances of a vehicle negotiating a curve were also provided as samples of regular driving and also included approximately 60 s of data (approximately 600 rows per alert/curve). The database contained a number of data fields (columns) with data from the instrumentation system, such as lateral acceleration and forward speed. The data corresponding to each alert is referred to as a "vehicle trace." The data also contained a number of columns that represent data from sources other than the instrumentation system, such as static driver characteristics (age, gender) and roadway variables (lane width, road class).

A large amount of time was spent reviewing the vehicle trace data set to determine what was included and how to link the UMTRI database and forward video data. Latitude and longitude were provided for each row of data, and geographic files were created for each vehicle trace using ArcMap. Each UMTRI vehicle trace was overlaid with aerial imagery, and the roadway type identified from the aerial image was matched to individual vehicle traces. Vehicle location (e.g., tangent, curve, freeway ramp) was determined and mapped to the corresponding vehicle rows. Vehicle data in the vicinity of a major intersection where the vehicle would have stopped or slowed significantly were identified and tagged. Data around intersections where the vehicle stopped or slowed significantly were not included in the final analysis.

Overlap in vehicle traces was also identified and removed. The team found significant overlap in the UMTRI vehicle traces. This is due to the manner in which the data were queried. For instance, UMTRI extracted data on curves, so when several curves were in a row, a vehicle trace was extracted for each curve. This resulted in several vehicle traces with overlapping data (see Figure A.1). The data for the overlapping areas were for the same vehicle, trip, and times. It was necessary to identify overlaps so that the same data would not be used twice in the analysis.

The rural designation used by UMTRI was for data collected in locations where the population was less than 50,000. This resulted in vehicle traces located in developed

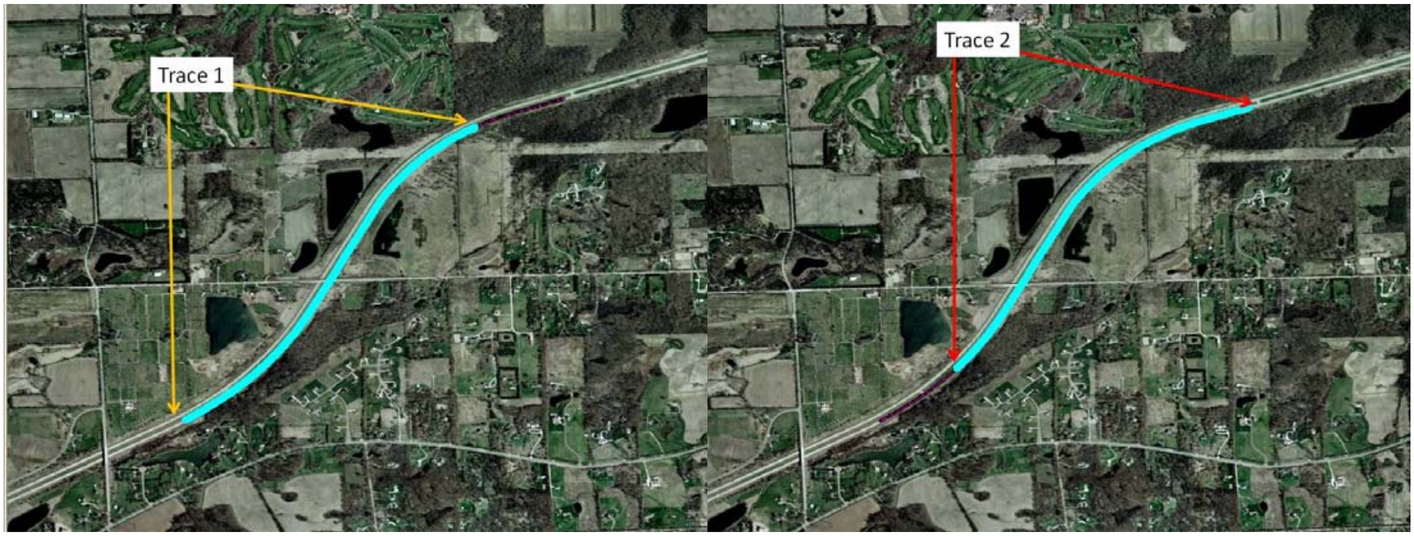



Image source: Esri. © 2010 i-cubed. Vehicle trace source: UMTRI.

***Figure A.1. Overlapping vehicle traces.***

areas. The team identified and removed vehicle traces where the activity occurred within an incorporated area or in an area with significant development along the roadway. Only regular rural driving was desired, and areas with a large amount of development would result in different driving patterns.

A large number of vehicle traces that indicated a driver was negotiating a curve turned out to be either a vehicle turning at an intersection or a vehicle turning off of or onto a roadway. Occurrences of these types were indicated. If the majority of the vehicle trace was regular, uninterrupted rural driving, the nonintersection or nonturning portion of the trace was retained. When the majority of the vehicle trace was turning activity, the trace was not included.

Preparation of the data for further analysis required a significant amount of manual data reduction. Additionally, the lane tracking system used in the RDCW did not perform well on unpaved surfaces. As a result, the team decided to focus on rural, two-lane, paved roadways. All resulting analyses were for this type of roadway.

## Available List of Variables Included in Analysis and Methodology to Extract Additional Variables

The following is a list of variables available in the UMTRI data or extracted to be included in the analyses of lane departures using the naturalistic driving study data. The variables are summarized by category: general, driver, vehicle, roadway, environmental, and exposure.

### General Variables Included in the UMTRI Data Set

The following general data fields were included in the UMTRI database. They were not included as variables in any of the analyses but were used in the extraction of other data.

- **Time:** In centiseconds (cs) since DAS started. Indicates time into the trip. It was used to identify alert times, identify overlap between vehicle traces, flag the beginning and end of events, determine how far into the trip an event occurred, and was used as an identifying feature. It was also used to correlate forward video images to vehicle traces.
- **Starttime:** Indicates the time an alert identified by UMTRI started. Time in cs since DAS started.
- **Endtime:** Indicates time an event identified by UMTRI ended. Time in cs since DAS started.
- **EventID:** Indicates the type of alert that was flagged for the vehicle trace. The alert was defined by the UMTRI FOT. This information was used to flag potential lane departures as defined by the UMTRI researchers. Events were defined differently for the research described in this report. This information was only used as a starting point to define events. The data field EventID used the following convention:
  - 1: LDW Cautionary Left;
  - 2: LDW Cautionary Right;
  - 3: LDW Imminent Left;
  - 4: LDW Imminent Right;
  - 5: Curve Speed Cautionary;
  - 6: Curve Speed Imminent; and
  - 7: Negotiating Curve (this indicated the presence of a curve but in most cases represented normal driving).

- **LDWBoundaryRight** and **LDWBoundaryLeft:** Represent the type of lane line to the left or right of the vehicle's present lane. This information was used by the UMTRI system to determine position and offset within a lane. Lane lines were indicated as follows:
  - 0: missing;
  - 1: dashed;
  - 2: solid; and
  - 3: virtual.
- **Latitude** and **Longitude:** In degrees. Both were used in identifying vehicle position for each time interval and in creating vehicle traces in Esri's geographic information system package, ArcMap, as shown in Figure A.1.
- **Heading:** In degrees. Indicates GPS heading for each time interval.
- **Radius:** In meters. The radius of curvature was calculated from the lane departure warning system. This data field was subsequently determined to be inaccurate, so the radius of curve was then calculated from aerial imagery by the CTRE team.
- **ThruLanes:** Indicates the number of through lanes. This variable was used to identify vehicle traces on two-lane roadways.
- **ShoulderLeft** and **ShoulderRight:** In meters. These variables indicate shoulder width. In all cases, their values were recorded as 5 m. These values were subsequently determined to be inaccurate, so actual shoulder widths were then determined and extracted using the forward images.

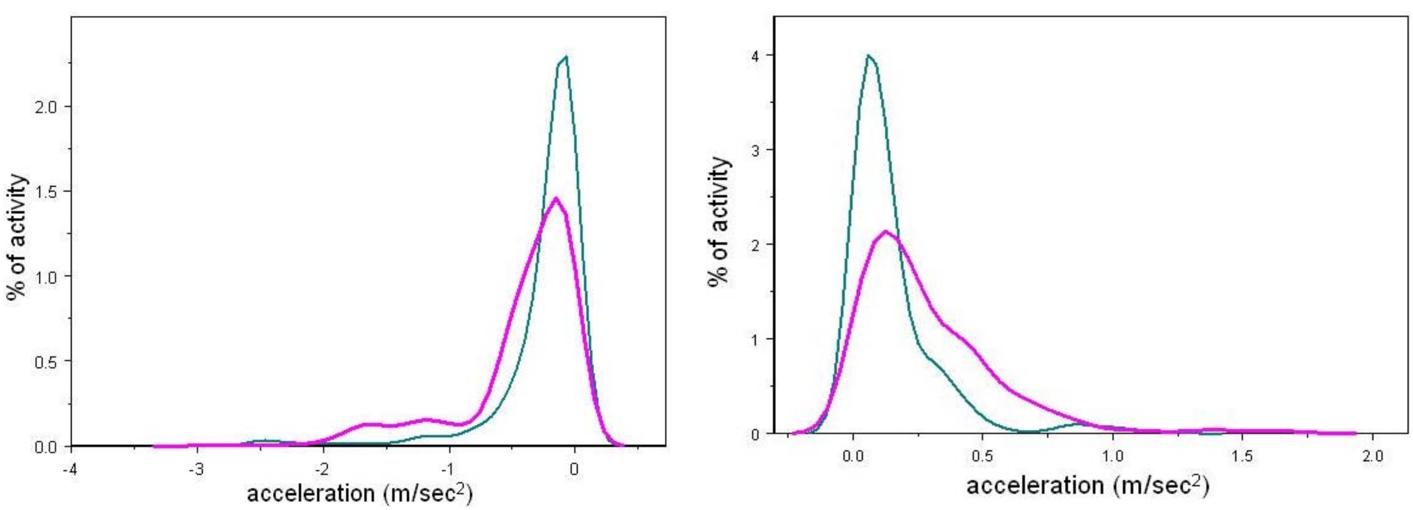
## Driver Variables

The following driver variables were available with the original data set or were extracted from the various data sets.

- **Driver age:** In years. Provided with data set.
- **Driver gender:** Provided with data set; 1 = male, 2 = female.
- **Trip number:** Provided with data set. The number of trips previous to and including the current trip.
- **Aggression_Accel:** Percentage. This variable reflects aggressive driving. It is defined as the percentage of time a specific driver exceeds a set acceleration level. The acceleration level is specific to the facility (e.g., two lane, freeway, intersection). The acceleration level can be determined by developing a distribution of accelerations (in $m/s^2$) for all drivers in the data set where vehicle activity is on two-lane rural roadways and the data did not include stopping or starting at intersections. The threshold will be set once all data have been reduced for all drivers. Acceleration for each time interval is available from the variable $A_x$. Figure A.2 shows the acceleration distribution for Drivers 6 and 12. Acceleration should be separated by positive and negative acceleration.
- **OvrSpd5** and **OvrSpd10:** Percentage. These variables reflect aggressive driving. They are defined as the percentage of time a specific driver exceeds the posted speed limit by 5 or 10 mph, respectively. The percentage of time driver $k$ exceeds the speed limit by $i$ mph was calculated using Equation A.1:

$$\text{PerOverSpeed}_{ik} = \frac{\text{number of records where speed for driver } k \geq i \text{ mph over speed limit}}{\text{Total number of records for driver } k} \quad \text{(A.1)}$$

This reflects driving traces included in the analysis of two-lane roadways. Records with missing data were not included in the analysis.




(a) (b)

***Figure A.2. (a) Distribution of deceleration and (b) acceleration ($m/s^2$) for Drivers 6 and 12.***

*Table A.1. Acceleration Statistics for Individual Drivers for Two-Lane Rural Paved Roads*

| | Dr6 | Dr8 | Dr12 | Dr14 | Dr16 | Dr17 | Dr18 | Dr24 |
|---|---|---|---|---|---|---|---|---|
| Min: | –2.59 | –2.96 | –1.20 | –0.47 | –0.93 | –2.11 | –2.27 | –2.23 |
| Mean: | –0.07 | –0.09 | –0.02 | 0.14 | 0.01 | –0.03 | –0.02 | –0.02 |
| Max: | 1.25 | 1.71 | 1.17 | 0.94 | 2.09 | 1.96 | 1.52 | 1.29 |
| Std Dev. | 0.37 | 0.53 | 0.27 | 0.31 | 0.28 | 0.34 | 0.42 | 0.38 |
| | **Dr28** | **Dr35** | **Dr48** | **Dr51** | **Dr59** | **Dr60** | **Dr64** | **Dr85** |
| Min: | –2.74 | –2.36 | –2.44 | –2.61 | –3.08 | –2.78 | –2.32 | –2.67 |
| Mean: | 0.00 | –0.03 | –0.07 | –0.11 | –0.09 | –0.07 | –0.03 | –0.03 |
| Max: | 1.76 | 2.09 | 1.19 | 1.22 | 0.74 | 1.88 | 1.26 | 2.15 |
| Std Dev. | 0.35 | 0.40 | 0.33 | 0.40 | 0.49 | 0.54 | 0.52 | 0.43 |

- **OvrCurveSpd5** and **OvrCurveSpd10:** Percentage. These variables reflect aggressive driving. They are defined as the percentage of time a specific driver exceeds the advisory curve speed limit by 5 or 10 mph (when there is a curve advisory). The percentage of time driver *k* exceeds the curve advisory speed limit by *i* mph was calculated using Equation A.2:

$$\text{PerOverCurveSpeed}_{ik} = \frac{\text{number of records where speed for driver } k \geq i \text{ mph over curve advisory speed limit}}{\text{Total number of records for driver } k \text{ where there are curve speed limit advisories}} \quad \text{(A.2)}$$

## Summary Statistics for Aggressive Driving Variables

Several variables were created to assess a measure of aggressive driving. Acceleration distributions were created for each driver for nonevent (normal) data. Table A.1 shows the acceleration summary statistics for individual drivers. The data represent driver accelerations during nonevents (normal driving). As shown, individual driver characteristics vary. Figures A.3 and A.4 provide box plots showing ranges of driver acceleration characteristics.

Table A.2 provides a summary of the percentage of time each driver travels over the posted speed limit by 5 or 10 mph, or the curve advisory speed limit by 10 or 15 mph. Cells shown as NA indicate curves with no advisory speed limits for the

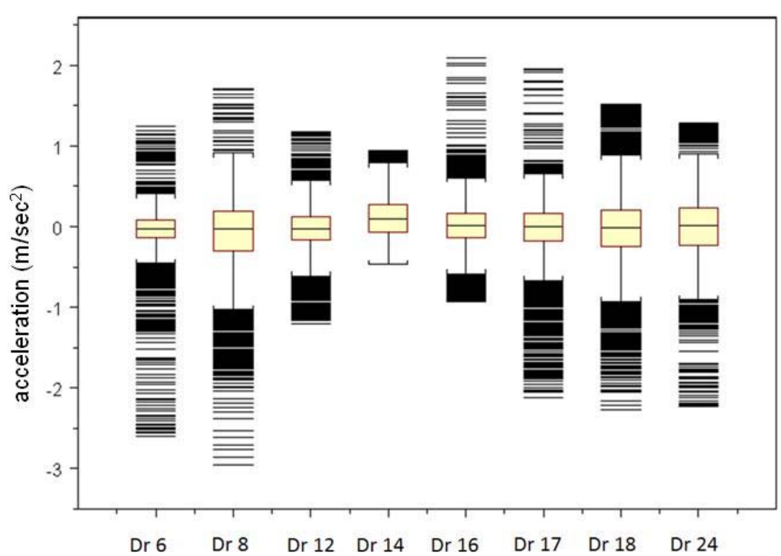



***Figure A.3. Box plots showing acceleration ranges for Drivers 6, 8, 12, 14, 16, 17, 18, 24.***

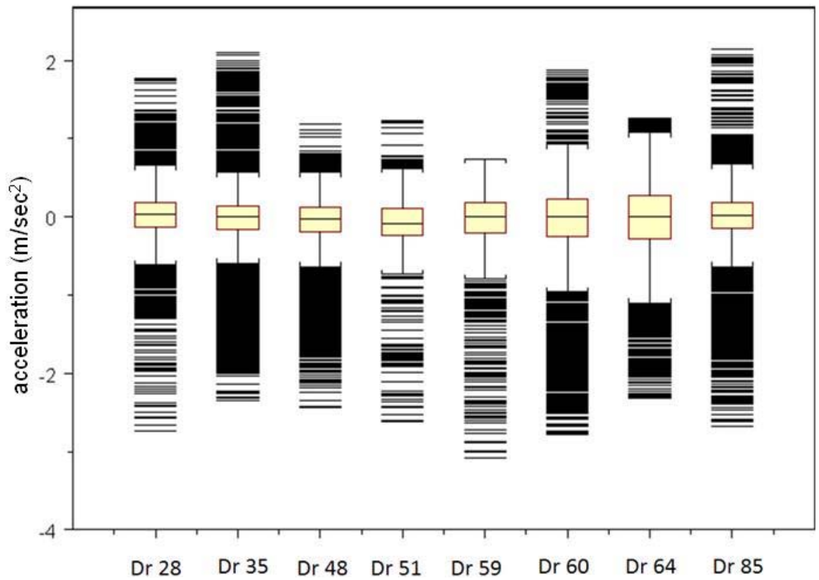



***Figure A.4. Box plots showing acceleration ranges for Drivers 28, 35, 48, 51, 59, 60, 64, 85.***

***Table A.2. Percentage of Time Driver Spends Over Posted or Advisory Speed Limit***

| | Posted Speed Limit | | Advisory Speed | |
|---|---|---|---|---|
| Driver | Over 5 mph | Over 10 mph | Over 10 mph | OvrAdv15 |
| 6 | 0.0% | 0.0% | NA | NA |
| 8 | 49.0% | 7.0% | NA | NA |
| 12 | 8.0% | 0.0% | 83.0% | 11.0% |
| 14 | 36.0% | 5.3% | NA | NA |
| 16 | 1.2% | 0.0% | 31.4% | 14.1% |
| 17 | 0.9% | 0.0% | 2.3% | 0.0% |
| 18 | 12.1% | 6.0% | 41.7% | 27.1% |
| 24 | 36.0% | 3.0% | 78.0% | 12.0% |
| 28 | 90.2% | 81.8% | 100.0% | 45.4% |
| 35 | 21.9% | 2.1% | 41.1% | 16.0% |
| 48 | 25.0% | 3.0% | 75.0% | 47.0% |
| 51 | 13.2% | 0.0% | NA | NA |
| 59 | 8.5% | 0.0% | 89.0% | 7.5% |
| 60 | 41.8% | 15.0% | 63.7% | 25.0% |
| 64 | 55.5% | 11.1% | 61.1% | 28.4% |
| 85 | 25.0% | 3.0% | 75.0% | 47.0% |

driver or the advisory speed was unknown. As indicated, drivers regularly travel over the posted and advisory speed limits but do so at different frequencies. Driver 6 rarely traveled over the posted speed limit and Driver 17 rarely traveled over either the posted or advisory speed limit. In contrast, Driver 28 traveled over the posted speed limit by 5 or 10 mph most of the time, exceeded the advisory speed on curves by 10 mph all of the time, and exceeded the advisory speed by 15 mph 45% of the time.

## Vehicle Variables

All vehicles in the UMTRI data set were of the same type. Thus, no individual vehicle characteristics, such as vehicle height, were considered. The following vehicle variables were included in the data set or were extracted:

- **Speed:** In meters per second (m/s). Provided with the data set, this variable indicates forward (longitudinal) vehicle velocity for each time interval.
- **LateralSpeed:** In m/s. Provided with the data set, this variable indicates lateral vehicle velocity for each time interval.
- **$A_x$:** In meters per second squared (m/s$^2$). Provided with the data set, this variable indicates forward (longitudinal) vehicle acceleration for each time interval. Deceleration is defined as negative acceleration.
- **$A_y$:** In m/s$^2$. Provided with the data set, this variable indicates lateral vehicle acceleration for each time interval. Deceleration is defined as negative acceleration.

- **Brake:** Provided with the data set, this is a categorical variable that indicates whether the brake was engaged, where 0 = not engaged and 1 = engaged.
- **Engaged:** Provided with the data set, this is a categorical variable that indicates whether cruise control was engaged, where 0 = off and 1 = on.
- **Roll:** In degrees. Provided with the data set, this variable indicates degrees of roll for each time interval.
- **RollRate:** In degrees per second (deg/s). Provided with the data set, this variable indicates roll rate for each time interval.
- **PitchRate:** In deg/s. Provided with the data set, this variable indicates pitch rate for each time interval.
- **YawRate:** In deg/s. Provided with the data set, this variable indicates pitch rate for each time interval.
- **AmrRight** and **AmrLeft:** In meters. Provided with the data set, these variables indicate distance to the nearest object to the right or left, respectively, for each time interval.
- **TravelDirection:** Determined from the time stamp and the aerial image (categorical). This variable indicates the primary direction of travel, where NB = northbound, SB = southbound, NEB = northeast bound, and so forth.

## Roadway Variables

The following variables are those that relate to the roadway and were either available with the UMTRI data set or were extracted from variables or other data sets.

- **LaneWidth:** In meters. Provided with the data set, this variable records the lane width for every time interval and was calculated within the instrumentation package based on the presence of left and right lane lines. Since lane width can vary from time interval to time interval, lane width was averaged across all intervals for each vehicle trace according to Equation A.3.

$$\text{LaneWidthAvg}_k = \Sigma \frac{\text{LaneWidth}_i}{N_k} \quad (A.3)$$

where

$\text{LaneWidthAvg}_k$ = average lane width for the travel lane calculated for each vehicle trace,
$\text{LaneWidth}_i$ = lane width for time interval $i$, and
$N_k$ = number of time intervals in vehicle trace $k$.

- **RoadType:** Determined from aerial imagery (categorical). Vehicle traces were overlaid with aerial imagery and time intervals were coded according to the corresponding roadway type. A single trace could consist of vehicle activity on several different roadway types. Areas around intersections were only designated as "intersection" when the vehicle would have to stop or slow down to yield right-of-way. Time intervals for intersections where the vehicle was not presented with any traffic control were coded as the regular roadway type (e.g., two-lane undivided). The variable RoadType was compared against the variable RoadClass included with the data, but was more descriptive. RoadType was designated using the following conventions:
  - 1: Two-lane undivided (one lane each direction);
  - 2: Four-lane undivided;
  - 3: Six-lane undivided;
  - 4: Four-lane divided (two lanes each direction);
  - 5: Six-lane divided (usually three lanes each direction);
  - 6: Freeway ramp (further indicated as diamond or cloverleaf ramp);
  - 7: At or near intersection;
  - 8: Other;
  - 9: Eight-lane divided; and
  - 9999: unknown.
- **CurveType:** Determined from aerial imagery (categorical). This variable indicates whether the curve is to the left or right from the driver's perspective (inside or outside of curve) during a particular time interval. Direction was confirmed by the forward video. CurveType was designated using the following conventions:
  - 0: No curve;
  - 1: Curve right; and
  - 2: Curve left.
- **PvmMarking:** Determined from the forward video (categorical). This variable is a subjective assessment of the visibility of pavement markings. A driver's ability to lane keep depends to some extent on having positive guidance as to the location of the traveled lane. The team reviewed the forward video and assigned pavement marking condition value according to the variable's categories. It should be noted that pavement markings for the same stretch of roadway would appear differently at night or under wet conditions than during the day or dry conditions. Examples are shown in Figures A.5 and A.6. This variable was an attempt to determine the marking visibility from the perspective of the driver and was defined with the following categories:
  - 0: Highly visible;
  - 1: Visible;
  - 2: Partially obscured;
  - 3: Obscured; and
  - 4: Nonexistent.
- **CurveSign:** Determined from the forward video (categorical). This variable indicates whether some type of curve signing can be seen in the time intervals corresponding to a particular curve. Curve signing includes chevrons, curve warning signs, and curve advisory speed signs. A sign was indicated as being a curve warning when it simply provided additional information about the curve similar to those

Source: UMTRI.

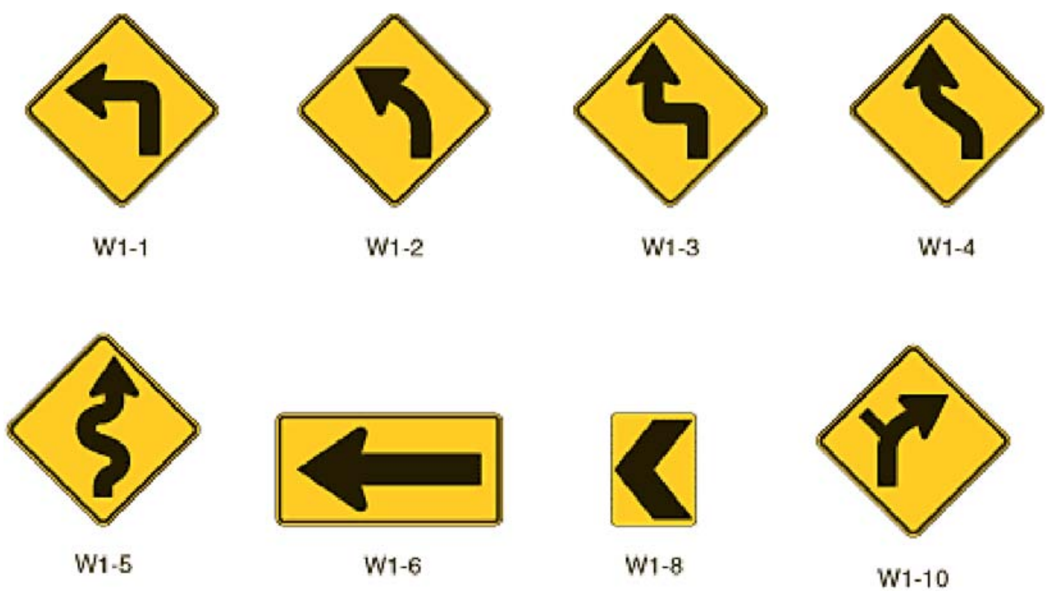
*Figure A.5. Pavement markings indicated as "highly visible" under nighttime conditions.*

Source: UMTRI.

*Figure A.6. Pavement markings indicated as "visible" under nighttime conditions.*

shown in Figure A.7. A sign with an advisory speed was also indicated as a curve advisory sign.

- **RoadwayLength:** In meters. Determined from the aerial imagery. This variable measures the length of each vehicle trace using a distance measuring tool in ArcMap. RoadwayLength is used in estimating density, crash rate, and driveways per mile.
- **ShoulderType:** Determined from the forward video (categorical). This variable classifies from visual observation the type of shoulder along a roadway section of a vehicle trace. Shoulder type is classified according to the following conventions:
  - 1: Paved;
  - 2: Paved/gravel;
  - 3: Gravel;
  - 4: Earth;
  - 5: Earth/paved;
  - 6: No shoulder; and
  - 7: Partially paved (includes 2 and 5 when used).
- **ShoulderWidth:** In meters. Measured from the forward image by calibrating a known distance (lane width) in the image.
- **PavedShldrWidth:** In meters. Measures the part of a shoulder's width that is paved. PavedShldrWidth is measured from the forward image by calibrating a known distance (lane width) in the image.
- **Radius:** In meters. Measures the radius of a curve using aerial images.
- **PostedSpeed** and **AdvisorySpeed:** In miles per hour. Curve advisory speed limit and posted speed limit were included in some of the data set's vehicle traces. When they were not included, they were obtained from the forward imagery when available. When posted speed limit was not included and could not be obtained from the forward imagery, it was obtained where possible from the crash data of a roadway segment. Posted speed limit was included as a variable in the Michigan crash data. In most cases there were multiple crashes along a roadway segment; all these crashes were confirmed to have consistently occurred in places with posted speed limits.
- **DwyDensity:** In driveways/m. Measured from aerial imagery and verified with the forward video when necessary. This variable counts the number of driveways along a vehicle trace. Driveway density is the number of driveways




Source: FHWA 2007.

*Figure A.7. Common curve warning signs.*

to the right of the vehicle in the direction of travel divided by vehicle trace length.

Driveway density was calculated as indicated above but not included as a variable in the analysis. In retrospect, if driveway density were included, it would be more appropriate to estimate it for a set distance in the immediate vicinity of the respective data points.

## Environmental Variables

The following describes environmental variables that were either available with the UMTRI data set or were extracted.

- **Wiper:** Provided with the data set (categorical). This variable indicates wiper blade status, which is an indicator of ambient precipitation in a time interval. Wiper status was designated using the following conventions:
  - 0: Off;
  - 1: Low;
  - 2: High;
  - 3: Invalid; and
  - 4: Intermittent.
- **Headlamp:** Provided with the data set (categorical). This variable indicates headlamp status, which is an indicator of ambient lighting conditions in a time interval. Headlamp status was designated using the following conventions:
  - 0: Off;
  - 1: Parking;
  - 2: Low; and
  - 3: High.
- **SolarZenithAngle:** In degrees. Provided with the data set. SolarZenithAngle can be used to determine time of day.
- **RoadSurf:** Determined from the forward video (categorical). This variable specifies pavement surface condition according to the following conventions:
  - 0: Bare (no evidence of precipitation);
  - 1: Wet;
  - 2: Snow cover along edge of roadway but travel lane is bare or mostly bare;
  - 3: Snow cover along edges and within roadway but bare vehicle tracks;
  - 4: Light snow cover over entire roadway surface; and
  - 5: Medium or greater snow cover over entire traveled way.
- **TimeOfDay:** In most cases, determined from the time stamp and forward video. This variable indicates the period when a vehicle trace occurred (categorical). It was recorded according to the following conventions:
  0: Daytime;
  1: Dawn/dusk; and
  2: Nighttime.
- **EnvCondition:** Obtained from the forward video. This variable indicates the prevailing atmospheric conditions when the driving trace occurred. Windshield wiper state can also be used to determine precipitation. EnvCondition indicates atmospheric conditions and may not correlate to pavement surface conditions. For instance, the prevailing environmental condition may be clear but there may be snow on the roadway surface. Environmental condition was designated using the following conventions:
  0: Clear (no precipitation);
  1: Light to moderate rain;
  2: Heavy rain;
  3: Light to moderate snow;
  4: Heavy snow; and
  5: Fog.
- **Lighting:** Determined from the forward imagery. This variable indicates the presence of street lighting (categorical). Most nonintersection, noninterchange sections of rural roadways are unlit. Street lighting conditions were categorized according to the following conventions:
  0: No overhead street lighting;
  1: Continuous lighting along roadway segment; and
  2: Intersection or interchange lighting but no continuous lighting on segment.
- **SegmentLength:** Segment length was measured from aerial imagery and reported in meters. It was used in calculating crash density, vehicle density, and so forth.

## Measure of Exposure Variables

The following variables were used as measures of exposure. These were either available with the UMTRI data set or were extracted as described.

- **AADT:** In vehicles per day (vpd). Annual average daily traffic for each roadway was provided with the UMTRI data set.
- **TimeDriving:** In seconds. The amount of time that a driver had been driving before the start of the vehicle trace was determined from the vehicle trace time stamp. Drivers who have been on the road for a significant time may be more likely to become drowsy or inattentive.
- **OnVehDensity:** In vehicles per meter (v/m). The number of oncoming vehicles that passed the subject vehicle during the driving trace was determined from the forward video. Oncoming traffic density was calculated by dividing the total number of oncoming vehicles by segment length.
- **PassDensity:** In v/m. Determined from the forward video, this variable provides the number of vehicles the subject vehicle passes. Passed traffic density was calculated by dividing total number of vehicles passed by the segment length.

*Figure A.8. Subject vehicle considered to be following lead vehicle.*

- **OtherPassDensity:** The number of vehicles traveling in the same direction that passed the subject vehicle was determined from the forward video. The density of vehicles passing the subject vehicle (v/m) was calculated by dividing the number of passing vehicles by the segment length.
- **Following:** Determined from the forward video. This variable indicates whether the subject vehicle was following another vehicle (categorical). A vehicle following very closely could also be detected by the forward radar. A vehicle is considered to be following another vehicle if it is close enough that the lead vehicle could influence its behavior. Figure A.8 shows an example of this. Figure A.9 shows an example of a situation in which a vehicle would not be considered as following another. Following was designated by the following conventions:
  0: Not following;
  1: Following; and
  2: Following closely.

### Other Exposure Variables

Other information available about the UMTRI data may also be used to determine exposure. It was reported that 80% of vehicle trips in the UMTRI FOT data were during the day and 20% were at night (LeBlanc et al., 2006). This information can be used to determine nighttime exposure. Trip length, travel by location (rural versus urban), and average trip distance by age and gender are available in the report by LeBlanc et al. (2006).

- **LDCrashes:** The number of lane departure crashes that had occurred along the roadway where the vehicle trace was located was also determined. Crash data were available from the Michigan DOT as described in Chapter 3. The Michigan crash data contains four sequences of events for up to three vehicles. Lane departure crashes were identified by reviewing sequence of events and crash type. Data were available for 2000 to 2006 (7 years of crash data). If any of the following were indicated for any vehicle in a crash, the crash was identified as being a lane departure crash:
  - Crossed centerline or median.
  - Ran-off-road left.
  - Ran-off-road right.
  - Re-entered road.
  - Collision with fixed object:
    - Bridge, pier, or abutment;
    - Bridge parapet end;
    - Bridge rail;
    - Guardrail face;
    - Guardrail end;
    - Median barrier;
    - Traffic sign post;
    - Traffic signal post;
    - Luminaire support;
    - Utility pole;
    - Other pole;
    - Culvert;
    - Curb;
    - Ditch;
    - Embankment;
    - Fence;
    - Mailbox;
    - Tree;
    - Railroad crossing signal;
    - Building;
    - Traffic island;
    - Fire hydrant;
    - Impact attenuator; and
    - Other fixed object.

  Crashes that were identified as being lane departure crashes were extracted into a separate database and were plotted along with vehicle traces in ArcMap. Crashes falling along the vehicle trace were selected. The crash information was reviewed and any crashes which were not indicated as occurring on the roadway where the vehicle trace was located were discarded.
- **CrashDensity:** Crash density in crashes per mile was calculated by dividing total number of lane departure crashes along the vehicle trace by the length of the trace.

*Figure A.9. Subject vehicle not considered to be following lead vehicle.*

## Identifying and Extracting Lane Departure Incidents

One of the research questions addressed by the team is how to define lane departure crash surrogate events and to develop thresholds between those events on the basis of vehicle kinematics, such as lateral acceleration. In order to answer the two research questions, it was necessary to identify actual lane departures within the UMTRI data set so that this information could be used to begin identifying thresholds. This section describes how vehicle lane departure events were identified and extracted from the UMTRI data set. Discussion of thresholds is provided in Chapter 5.

As indicated in the section on data preparation at the beginning of this appendix, some of the vehicle traces provided by UMTRI had been flagged as a lane departure or curve warning alert. Alerts were identified according to the road departure crash warning (RDCW) system field operational testing protocol. The thresholds set for the RDCW system alerts are discussed in Chapter 3. Their identification of alerts was used as a starting to point to identify lane departures. Other instances of lane departures were also found in the vehicle traces as described in the following sections.

## Identifying Lane Departure Incidents

The main method to identify lane departures was to evaluate vehicle wheel path. Lane departures were also confirmed by a review of the forward imagery. Wheel paths were determined by calculating a vehicle position within its lane for each record of data (one record per 0.1 s of vehicle activity). The UMTRI data set had the following variables that were used to calculate wheel path.

- **TrackWidth:** In meters. Width of vehicle wheelbase was used to calculate offset from the left and right lane lines. Since all vehicles in the data set were of the same type, track width was consistent between vehicles.
- **LaneWidth:** In meters. Provided with the data set, this variable recorded the lane width for each time interval and was calculated within the instrumentation package based on presence of left and right lane lines. Since lane width is calculated, it will vary from time interval to time interval even though in reality the lane width would not vary in this manner.

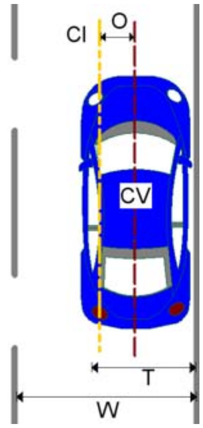



***Figure A.10. Schematic of variables used to calculate lane edge and wheel path locations.***

- **Offset:** In meters. Provided with the data set, this variable indicates the vehicle offset from the center of the lane as calculated by the lane departure warning system. Offset is shown as the variable O in Figure A.10.
  - Right and left lane edge and right and left wheel paths were calculated for each time interval for each vehicle path. Figure A.10 shows a schematic of the variables used to calculate lane edge and wheel path.
  - Lane line and wheel path locations are referenced from the right lane edge (RLE) which is set as the reference point (0). Position is positive moving to the left as shown in Figure A.11. Left lane edge is calculated using Equation A.4.

$$\mathrm{LLE}_k = \mathrm{RLE}_k + W_k \quad \text{(A.4)}$$

where

$\mathrm{LLE}_k$ = position of left lane edge for time interval $k$,
$\mathrm{RLE}_k$ = position of right lane edge for time interval $k$, and
$W_k$ = lane width measured at time interval $k$.

Since RLE is always referenced as 0, $\mathrm{LLE}_k = W_k$.

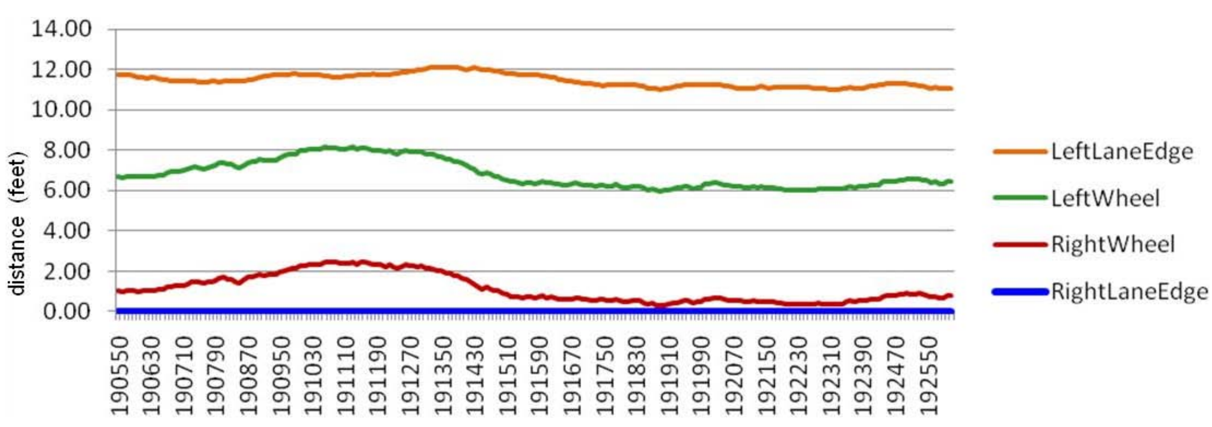



***Figure A.11. Lane edge and wheel path plotted by time interval.***

Right wheel path is calculated using lane width, lane offset, and track width according to Equation A.5:

$$RWP_k = \frac{W_k - T - O_k}{2} \quad (A.5)$$

where

$RWP_k$ = right wheel path (in meters) is the location of the right wheel path relative to the right and left lane lines for time interval $k$,
$W_k$ = lane width (in m) measured by RCWS for time interval $k$,
$T$ = vehicle track width, given as 1.73 m, and
$O_k$ = vehicle offset (in meters) is the distance between the centerline of the vehicle and the centerline of the lane for time interval $k$.

Left wheel path was calculated using Equation A.6:

$$LWP_k = RWP_k + T \quad (A.6)$$

where

$RWP_k$ = right wheel path (in meters) for time interval $k$ and
$T$ = vehicle track width, given as 1.73 m.

Once wheel paths and lane edge were calculated, they were converted to feet since it was more intuitive to view wheel path traces in familiar units.

A plot of lane edge and wheel path position was created in Excel for each vehicle trace as shown in Figure A.12. Each plot was evaluated to determine where lane departures occurred.

A lane departure was defined as a vehicle wheel path crossing over the right or left lane line and encroaching upon either the shoulder or adjacent lane as shown in Figure A.13. Lane encroachment to the right was determined when the right wheel path ($RWP_k$) had a negative value since the right lane edge was defined as 0. An encroachment to the left was determined when the value for the left wheel path ($LWP_k$) was greater than the lane width ($W_k$). Because there is some uncertainty in estimation of where the lane edges are, UMTRI used a buffer and only included encroachments that were greater than 0.1 m past the lane edge (LeBlanc et al., 2006). The research team adopted its convention and only included lane departures when the vehicle was more than 0.1 m (0.328 ft or 3.94 in.) beyond the left or right lane edge.

Each wheel path plot was also inspected in conjunction with the corresponding forward imagery. In some cases, a lane departure had occurred but was intentional, such as a vehicle turning into a driveway or moving over for a parked vehicle. Figure A.13, for instance, shows a subject vehicle moving over for a stopped vehicle. Situations where a lane departure was intentional were not included as lane departures in the analyses.

## Additional Information Extracted for Lane Departures

Once a lane departure was identified, additional information about the lane departure was extracted using the various data sets. The angle of departure (θ) from the roadway was calculated as shown in Figures A.14 and A.15. The approximate linear path of the right wheel for a right-lane departure or left wheel for a left-lane departure was determined using vehicle path data just prior to the lane departure. The approximate linear path the departing tire would have followed had the vehicle not recovered was determined by estimation. The linear paths and geometrical relationships were used to determine the angle of departure as shown in Figure A.15.

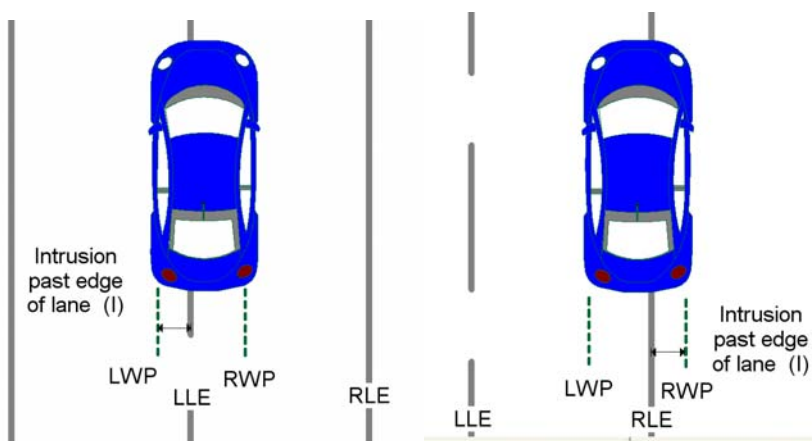
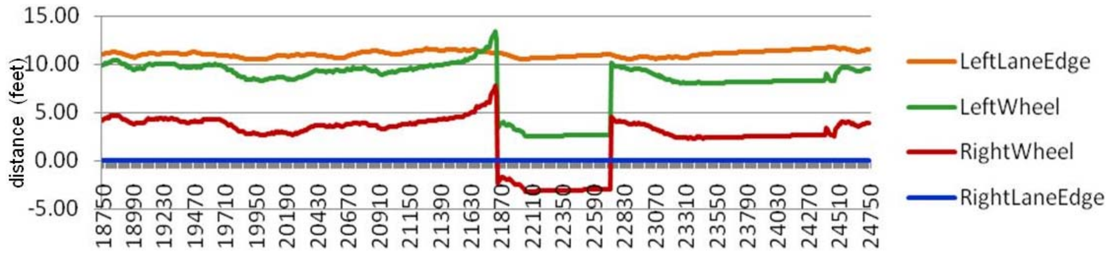



*Figure A.12. Vehicle wheel path intruding on left lane edge (left image) or right lane edge (right image).*




(a)

(b)

***Figure A.13. Identification of intentional lane departure for vehicle parked on shoulder: (a) vehicle wheel path shows lane departure and (b) video indicates lane departure was the result of the subject vehicle moving over for parked vehicle.***

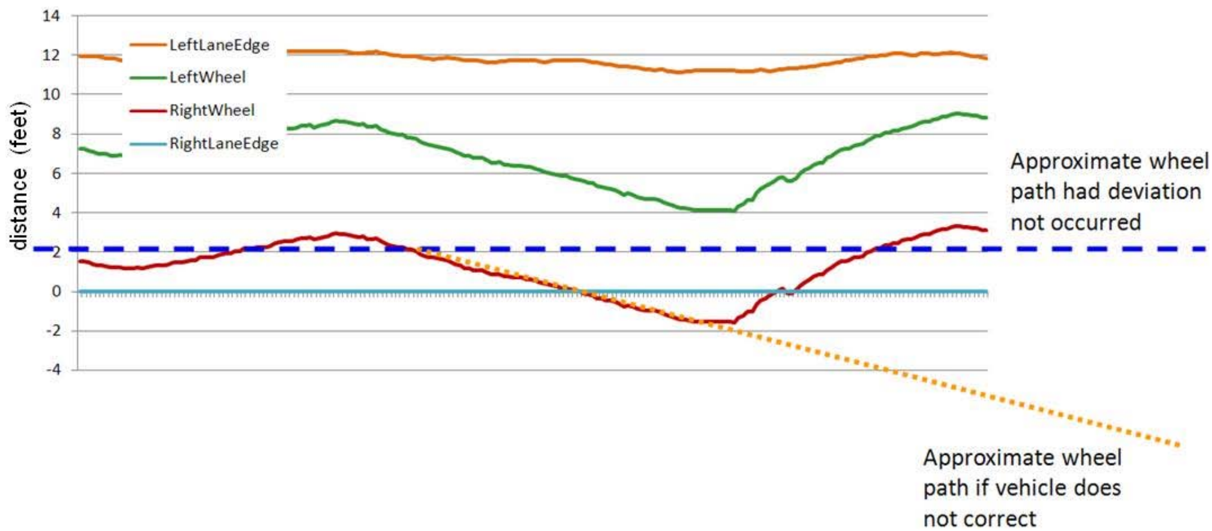



***Figure A.14. Estimation of wheel path for vehicle trace.***

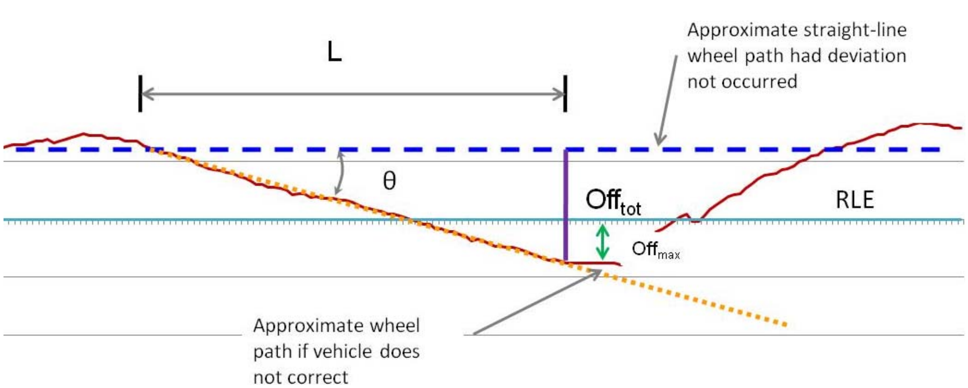



***Figure A.15. Schematic for angle of departure.***

The angle of departure was calculated using Equation A.7:

$$\text{Angle of departure } (\theta) = \arctan\left(\frac{L}{\text{Off}_{tot}}\right) \quad \text{(A.7)}$$

where

Angle of departure ($\theta$) = angle that the vehicle departed from its original straight-line wheel path (in degrees),

$L$ = longitudinal distance vehicle traveled from the point the wheel departed from the straight-line path to the point of maximum offset ($\text{Off}_{max}$), and

$\text{Off}_{tot}$ = total distance from the straight-line path to the point of maximum offset ($\text{Off}_{max}$).

Maximum offset ($\text{Off}_{max}$) was the maximum distance that the wheel encroached beyond the edge of the lane. Maximum offset for right-lane departures was calculated by subtracting the offset value for the maximum point of encroachment to the right from the right-lane edge position. Maximum offset for left-lane departures was calculated by subtracting left-lane edge from offset at the maximum point of encroachment.

APPENDIX B

# Methodology for Extraction of Data Elements from the Virginia Tech Transportation Institute Naturalistic Driving Study Data Set

The following describes the variables selected to evaluate lane departure research questions using the VTTI naturalistic driving study data set (100-car study). It also describes the methodology to extract those variables when relevant.

VTTI had already identified near crashes and crashes as part of the 100-car naturalistic driving study data. These were the only data available; there were no exposure data. Thirty-three near crashes and crashes were provided by VTTI after a data request was made as part of this project. VTTI has since made some data from their 100-car study publically available on a data distribution website (http://forums.vtti.vt.edu/index.php?/files/category/3-100-car-data/). Data on rural areas from this website showed that lane departure was involved. These data from the VTTI website had not been included among the data earlier provided to the research team. The narrative and other data descriptions for data from both sources were reviewed. Several cases that had been identified as near crash or crash have been found to be the result of intentional lane changes or merging by the subject or another driver. These instances were not included in the present study, because it focuses on unintentional lane departures. Excluding these cases yielded a total of 29 crashes and near crashes that were used in various analyses. However, the VTTI data consisted of all rural roadway types (e.g., ramp, divided, two lane), so there were very few samples for each particular roadway type. A description of the data sets is provided in Chapter 3.

VTTI had already identified the start and end times for each near crash or crash to reduce events. Their convention was used to define the start and end points when continuous data were used.

Data from VTTI were already reduced in the sense that crashes and near crashes were identified; the video data were already reduced, providing event narratives, eyeglance information, and so forth. The data were processed into various databases to facilitate analysis.

The VTTI data required very little processing and included data for both divided roadways and two-lane roadways. As a result, both were used in several of the analyses even though only data for two-lane roadways were extracted from the UMTRI data. Only data on rural roadways were included.

Since spatial information was not provided with the VTTI data, additional data could not be extracted from other sources, such as aerial imagery.

Continuous data were provided at 10 Hz (0.1 s) for each crash or near crash.

The following lists the variables that were available in the VTTI data or was extracted and was expected to be used in the analysis.

## Vehicle Variables

Information was obtained by the VTTI data reductionist, unless otherwise indicated.

- **Vehicle type:** No information was provided about vehicle type.
- **Acceleration:** Forward and side acceleration were provided for each record of data, measured in g-force (*g*).
- **Speed:** Forward speed in mph.
- **Available maneuvering room forward and rear:** Range (ft) and position (degrees) of obstructions in range of radar.
- **Brake:** Brake status (off/on).
- **Turn signal state:** Status of turn signal (off, right, left).
- **Vehicle factors:** Vehicle factors that may have contributed to event (e.g., tire defect or malfunction, wiper defect or malfunction).

## Driver Variables

Information was obtained by the VTTI data reductionist, unless otherwise indicated.

- **Age:** Driver age.
- **Gender:** Driver gender.
- **Driver reaction:** Driver reaction in response to the event (e.g., steered to left, steered to right).

- **Driver behavior:** Driver actions that occurred near the event, actions that led to the event, or actions taken to avoid the event (e.g., exceeded speed limit, avoided animal, driving without lights).
- **Driver impairment:** Potential driver factors (e.g., drowsy, angry, drugs).
- **Driver distraction:** Up to three distractions that the driver was engaged in 5 to 6 s prior to the onset of the event (e.g., lost in thought, reading).
- **Hand on wheel:** Describes whether driver had hands on wheel (no hands, left hand, right hand, both hands).
- **Visual obstructions:** Factors that may have interfered with driver's line of sight (e.g., curve, trees, rain).
- **Reaction of other drivers:** Action or maneuvers by other drivers causing the event or in response to the event.

## Event Variables

Information was obtained by the VTTI data reductionist, unless otherwise indicated.

- **Event start and end:** Time stamp that marks approximate start and end of event.
- **Event nature:** Type of conflict that occurred (e.g., conflict with lead vehicle, conflict with merging vehicle).
- **Preincident maneuver:** The action that the vehicle was engaged in just prior to the event (e.g., going straight, stopped in traffic).
- **Maneuver judgment:** Indication of the legality of maneuver leading to the event as determined by the data reductionist.
- **Precipitating event:** Event that started the sequence leading to the crash or near crash (e.g., subject over right-lane line).
- **Post maneuver:** Vehicle action after avoidance of crash or near crash.
- **Vehicles:** Number of vehicles, type of vehicles, maneuver of other vehicles involved in the event.
- **Fault:** Indication of which driver caused the event.
- **Vehicle position:** Position of surrounding vehicles (e.g., in front and to right).

## Roadway Variables

Information was obtained by the VTTI data reductionist, unless otherwise indicated.

- **Infrastructure:** Roadway factors that may have affected a driver's ability to safely navigate the roadway (e.g., roadway alignment, weather).
- **Lanes:** Number of traffic lanes in the direction of travel.
- **Geometry:** Presence of curve or grade (e.g., straight/level, curve/level, straight/grade).
- **Land use:** Land use in area at start of the event (e.g., church, residential).

## Traffic Variables

Information was obtained by the VTTI data reductionist, unless otherwise indicated.

- **Density:** Level of service (A to F) as determined by the data reduction.
- **Traffic control:** Traffic control at start of the event (e.g., stop sign).
- **Intersection:** Position of vehicle relative to intersection or junction at the time of the event (e.g., nonintersection, intersection, driveway).

## Environmental Variables

Information was obtained by the VTTI data reductionist, unless otherwise indicated.

- **Roadway surface condition:** Roadway surface condition that may have caused a reduced coefficient of friction (e.g., wet, dry, ice).
- **Lighting:** Light condition (dawn, daylight, dusk, dark/lighted, dark/not lighted).
- **Weather:** Ambient weather (clear, cloudy, fog, mist, rain, snow, sleet, smoke).

## Data Limitations

A number of variables that were determined to be necessary to answer the lane departure research questions were not available from any source in the VTTI data. A list of data variables necessary to answer the lane departure research questions is summarized in Chapter 4, which also outlines the limitation of the VTTI data. In summary, the primary limitations that affected the research team's ability to fully answer research questions included the following:

- No lane tracking information, such as lane width or vehicle offset, was available that could determine vehicle position relative to the lane. The researchers had to rely on the VTTI data reductionist's interpretation of whether a vehicle had departed its lane.
- Vehicle spatial position (latitude/longitude) was not provided. As a result, the researchers could not overlay the vehicle traces with aerial imagery or other spatial data sets to extract additional information, such as radius of curve.
- Forward video resolution made it difficult to determine a number of factors that could be determined in the UMTRI forward video. For instance, in the UMTRI data, it was possible to tell the distance to an object on the side of the roadway and to determine the pavement surface condition.

APPENDIX C

# Assessment of Availability of Roadway Data Elements in the SHRP 2 Naturalistic Driving Study Data Acquisition System

## Background

A preliminary test run of the SHRP 2 data acquisition system (DAS) was evaluated to determine what roadway data elements could feasibly be extracted in the event they are not available in roadway data sets from SHRP 2 Safety Project S04A (Roadway Information Database Developer, Technical Coordination, and Quality Assurance for Mobile Data Collection) or Safety Project S04B (Mobile Data Collection). A list of roadway factors that are necessary to answer lane departure research questions was developed by the CTRE team as part of their efforts in Safety Project S04A. The research team for this project reviewed the list of roadway data elements and DAS data set and commented on which roadway factors can be extracted from the DAS data set which may be useful to researchers for Safety Project S08 (Analysis of the SHRP 2 Naturalistic Driving Study Data), as well as to other researchers. Ideally, roadway information will be available from the mobile mapping data collection for Safety Project S04B. However, Safety Project S04B data collection will not cover all areas where naturalistic driving study data will be collected, and not all necessary factors will be collected under Safety Project S04B. It is therefore important to comment on whether these factors can be reduced from the naturalistic driving study data when they cannot be obtained from existing data sets or the roadway data set.

Evaluation of the accuracy of the lane tracking system and of the global positioning system (GPS) is beyond the scope of this project.

### Description of DAS

VTTI instrumented a test vehicle with what is expected to be the final version of the data acquisition system that will be used by the pilot study sites in SHRP 2 Safety Project S07 (In-Vehicle Driving Behavior Field Study). The CTRE team received two sets of test runs from VTTI. The data were collected over the test route that had been surveyed for vendors participating in the mobile mapping data collection rodeo for SHRP 2 Safety Project S03 (Roadway Measurement System Evaluation) (Figure C.1). The data received included two video files with four video views (Figure C.2) and a database with raw system data that included GPS coordinates for the following vehicle data:

- Vehicle kinematics (e.g., speed, forward acceleration);
- Lane position;
- Presence of lane lines;
- Steering wheel position;
- Turn signal state;
- Temperature; and
- Light level.

### Methodology to Extract Data Elements from DAS

The team reviewed the GPS traces and the forward and back videos to determine roadway data items that could be extracted. The team assessed the data using three methods as described in the following sections.

#### *Comparing List of Roadway Data Elements to Forward Video*

First, the team used a list of roadway data elements that had been identified as part of SHRP 2 Safety Project S04A. The list identified roadway data elements that had been indicated as being important in addressing either road departure or intersection crashes research questions. Research questions from Safety Projects S01 (Development of Analysis Methods Using Recent Data), S02 (Integration of Analysis Methods and Development of Analysis Plan), S05 (Design of the In-Vehicle Driving Behavior and Crash Risk Study), and S06 (Technical Coordination and Quality Control) were reviewed and data

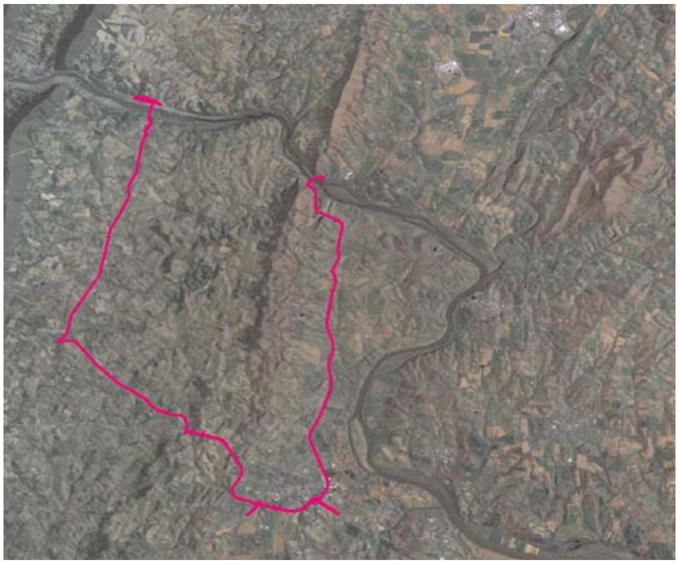


***Figure C.1. Trace of DAS activity data.***

elements identified. Next, the team reviewed data elements included in the Highway Safety Information System (HSIS), Model Inventory of Roadway Elements (MIRE), and the Model Minimum Uniform Crash Criteria (MMUCC). Data elements included in those inventories not already identified were added to the list. A survey was conducted under SHRP 2 Safety Project S04A to solicit additional user input. Roadway data elements not already identified were added to the list and a final list of roadway data elements that would be necessary to comprehensively answer lane departure or intersection safety questions was completed.

The team reviewed the forward video, and as each item was encountered, they noted whether the data element could be seen in the forward video and whether it was likely that data reductionists could identify and extract the data element. The list of data elements and assessment of whether it could be collected using the forward video is provided in Tables C.1 to C.5. The team primarily tested whether a feature was present. Location of an object can be associated with a corresponding roadway feature. For instance, a vehicle trace can be located to a particular curve and forward imagery could indicate presence of several chevrons. As a result, the curve could be coded as having chevrons. However, spacing of the chevrons or exact location of the chevrons could not be determined. An approximate location could be identified if the forward video frame, vehicle's spatial position, and location of a proximate feature on an aerial image could be linked. This would be significantly affected by the accuracy of the GPS.

Length cannot be accurately calculated. However, length of an object, such as length of a horizontal curve, could be roughly estimated if the begin and end points are located in the forward frame of the video, which provides an estimate of time ($t$). Vehicle speed ($v$) can be extracted from vehicle data in the DAS and length ($L$) calculated as:

$$L = v * t$$

As indicated, it is possible to approximate location and length of an object. However, there was no feasible way to test the accuracy of these measurements for this exercise. As a result, the effort focused on identification of features. Elements included in the list of roadway data elements that could not feasibly be extracted are indicated as not applicable (NA). In some cases a feature was included in the list but the data reductionists did not view that object in the DAS database. This may be because the feature was not present in the test run made by the

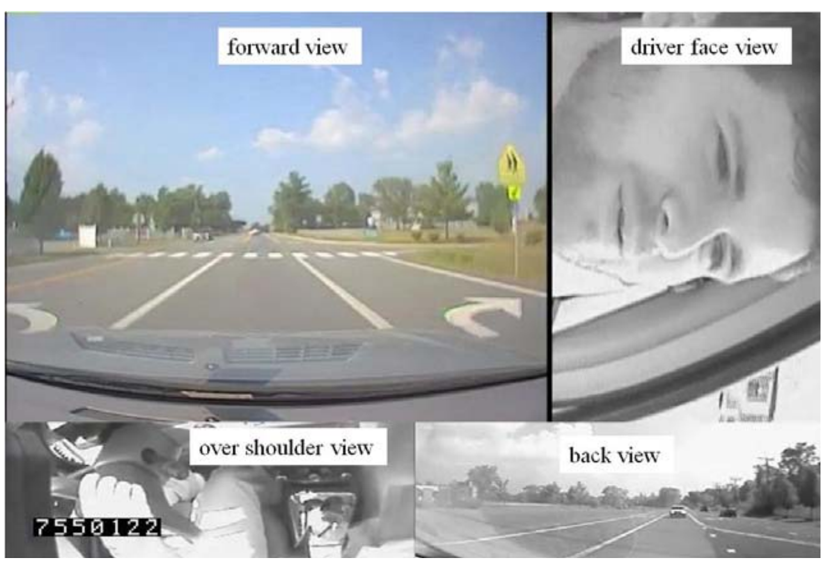



***Figure C.2. Video views provided with DAS.***

*Table C.1. Identification of Curve Data Elements and Roadway Cross-Section Features*

| Data Element | Features | DAS |
|---|---|---|
| **Horizontal curvature** | Presence | Yes |
| | Length | Could be approximated from time between frames and vehicle speed |
| | Location | Begin and end points can be approximately located |
| | Presence and type of spirals | NA |
| | Tangent length between adjacent curves | Could be approximated from time between frames and vehicle speed |
| | Radius or degree of curve | NA |
| | Presence and amount of superelevation | NA |
| | Direction of curve | Can be determined from driver perspective; left-hand vs. right-hand |
| | Curve deflection angle | NA |
| **Vertical curvature** | Vertical curve length | Difficult to establish the begin point of a vertical curve |
| | Tangent length between adjacent curves | NA |
| | Grade (percent) | NA |
| | Grade direction | NA |
| | Terrain type | Can determine flat vs. hilly vs. mountainous |
| | Grade length | Difficult to establish the begin point of a grade |
| **Cross section** | Lane width | NA |
| | Surface width | NA |
| | Number of lanes | Yes |
| | Lane direction (one way, two way, auxiliary, reversible) | Yes |
| | Cross slope | NA |
| | Type (e.g., regular; two-way left-turn lane) | Yes |
| | Turn lane length | NA |
| | Surface type | Yes |
| | Median type | Yes; could differentiate between grass, flush, raised |
| | Median width | NA |
| | Curb type | Yes |
| | Type and characteristics of bicycle facilities | Either was not present or could not be identified |
| | On-street parking type | Yes |
| **Shoulder** | Right/left shoulder type | Yes |
| | Right/left shoulder paved width | NA |
| | Right/left total shoulder width | NA |
| | Right/left shoulder slope | NA |
| | Right/left shoulder condition | NA |

***Table C.2. Identification of Signs by Type***

| Data Element | Features | DAS |
|---|---|---|
| **Regulatory signs** | Speed limit | Yes; in some cases sign text was blurred |
| | Pass/no pass zones | Either was not present or could not be identified |
| | Other (lane end, do not enter, no parking) | Yes; in some cases sign text was blurred |
| | School area | Yes; although text was blurred in some cases, sign shape was distinct |
| | Railroad crossing | Either was not present or could not be identified |
| | Stop | Yes; shape was distinct |
| | Yield | Yes; shape was distinct |
| **Warning signs** | Horizontal alignment signs and location (e.g., chevron, curve advisory speed) | Yes; chevrons can be identified; with curve advisory signs, in some cases sign text/symbol was blurred |
| | Roadway cross-section changes (e.g., lane ends) | Yes; in some cases sign text was blurred |
| | Vehicular warning (e.g., horse and buggy) | Either was not present or could not be identified |
| | Nonvehicular warning (e.g., deer, pedestrian, snowmobiles) | Yes; in some cases sign text was blurred |
| | Object markers | Yes; shape was distinct |
| | Speed reduction | Yes; in some cases sign text was blurred |
| | Slippery when wet | Either was not present or could not be identified |
| **Guide signs** | Guide destination signs (type and location) | Yes; although text was blurred in some cases, sign shape was distinct |
| | Route signs (type and location) | Yes; in some cases sign text was blurred |
| | Route sign auxiliary signs (type and location) | Yes; in some cases sign text was blurred |
| | Advance turn and directional arrow auxiliary signs (type and location) | Yes; in some cases sign text was blurred |
| **Service signs (e.g., camping, food)** | Sign type | Yes; in some cases sign text was blurred |
| **Other** | School crossing | Yes; shape was distinct |

VTTI instrumented car (e.g., no roundabouts were present) or the feature may have been present but data reductionists did not observe an instances of the feature.

It should be noted that both data sets were collected under clear conditions in the daytime. There was some glare, but there were no adverse ambient conditions. Consequently, the ability to identify features does not account for that variation in conditions that will be present in the full-scale study.

Table C.1 indicates which curve and roadway cross-section features could be identified. The items apply to both tangent sections and intersections. The presence of horizontal and vertical curves could be easily determined for pronounced curves. A very flat vertical curve or horizontal curve with a large radius would be difficult to identify. Most of the features of a curve cannot be determined (e.g., radius, superelevation). Width of objects cannot be determined using any of the DAS data elements.

Figure C.3 illustrates some of the cross-section data elements as viewed by the data reductionist.

A list of sign types that were identified as being important to lane departure or intersection research questions is listed in Table C.2, along with an indication if they can be identified in the forward imagery of the DAS. However, the text or symbols on the sign face was often difficult to read. Signs with distinct shapes (e.g., stop, yield) were the easiest to identify. The sign face for speed limit signs was usually legible. It was more difficult to identify signs when there was significant glare or foliage along the roadway. It was also difficult to detect signs when there was on-street parking or the test vehicle was traveling on an inside lane away from the road edge. This was particularly problematic when other vehicles were between the test vehicle and road edge. Table C.2 indicates if the sign could be identified. Figure C.4 illustrates some of the sign data elements as viewed by the data reductionists.

*Table C.3. Identification of Pavement Marking and Lighting*

| Data Element | Features | DAS |
|---|---|---|
| **Pavement markings** | Edge line | Yes |
| | Centerline (e.g., dashed, solid) | Yes |
| | Location of pass/no pass | NA |
| | Lane line | Yes |
| | Center island | Yes |
| | Arrows (e.g., merge, left only) | Yes |
| | Text (e.g., Slow, School Ahead) | Yes |
| | Raised pavement markings | Either was not present or could not be identified |
| | Stop and yield lines | Yes |
| | Crosswalks | Yes |
| | Parking | Yes |
| | Other (e.g., speed hump, HOV, colored pavement, curve ahead) | Yes, found instance of colored pavement |
| **Illumination** | Overhead lighting type | Presence and type of mast arm could be determined |
| | Overhead lighting location | NA |
| | Overhead lighting characteristics (e.g., lumens) | NA |
| | Type of in-pavement lighting | Did not encounter in database |
| | Location of in-pavement lighting | Did not encounter in database |

Pavement markings were evaluated and are listed by type in Table C.3. Pavement markings in almost all cases were easily identified. Pavement markings included lane lines, painted median/gore areas, and on-pavement markings such as stop bars or turn lane designations. In most cases, a qualitative assessment of condition was possible. For instance, markings could be categorized by grouping such as "like new," "good condition," "faded but visible," or "faded barely visible." An assessment of lighting is also included in Table C.3. Examples of pavement markings and lighting from the analysis are shown in Figure C.5.

Evaluation of roadway surface elements and identification of objects in the clear zone are listed in Table C.4. Surface type could be identified between asphalt, concrete, and gravel. No surface condition data elements (e.g., friction, roughness) can be determined. Only obvious roadway defects, such as patching, could be identified in the forward video. Clear zone elements were also included. The type of objects within the clear zone could easily be determined. This includes trees, utility poles, guardrails, and so forth. Examples of roadway surface and clear zone elements from the analysis are shown in Figure C.6.

Table C.5 provides information about how well countermeasures and access management features could be extracted. Only centerline rumble strips were identified in the forward video. Edge line, shoulder line, and advance stop line rumble strips were either not located along any of the roadways or were present but could not be identified. One speed feedback sign and one flashing beacon were identified in the data. Although it was not possible to read the text on the feedback sign, it was possible to determine through the forward video whether the sign was activated or turned off. Presence of driveways and type of median can be identified in the DAS data set. Together they can be used to estimate level of access control. Bridge characteristics were also included in Table C.5. Bridge type and presence of barriers and abutments could be determined from the forward video. Examples of data elements as they appeared in the data set are provided in Figure C.7.

The data elements listed in Tables C.1 to C.5 apply to the entire roadway. Data elements in Table C.6 are specific to intersections. Several cross-section features were identified as important elements that are specific to intersections. Number of lanes for each approach of an intersection could usually be determined if the cross streets were visible within the forward video frame. The number of lanes could always be determined for the approach where the vehicle was located. Left- and right-turn prohibitions can be determined if the corresponding sign can be identified.

*Table C.4. Identification of Roadway Surface and Clear Zone Characteristics*

| Data Element | Features | DAS |
|---|---|---|
| **Road surface** | Surface type (e.g., gravel, asphalt, PCC) | Yes |
| | Surface friction | NA |
| | Macro-texture | NA |
| | Pavement roughness | NA |
| | Pavement condition | NA |
| | Roadway rideability | NA |
| **Roadway defects** | Pavement edge drop-off | NA |
| | Roughness | NA |
| | Surface irregularities | Some irregularities could be determined, but it was difficult to distinguish from shade |
| | Road debris (best source would be forward video) | Yes; observed plastic bag flying down the street |
| **Clear zone** | Type of objects within clear zone (tree, utility pole, sign) | Yes |
| | Clear zone distance | NA |
| | Slope beyond edge of shoulder | NA |
| | Presence, type of guardrail | Yes |
| | Guardrail end | Yes |
| | Guardrail face | Yes |
| | Curb presence | Yes |
| | Curb type | Yes |
| | Right-of-way | NA |
| | Roadside hardware types and location (e.g., barriers, culverts) | Could only identify if hazard marking was present |
| | On-street parking | Yes |
| | Concrete barrier | Yes |
| | Other longitudinal barriers | Yes |

The type of intersection control by approach could be determined in all cases for the approach where the vehicle was located. In some cases, type of control could be determined for adjacent or opposing approaches. When a signal is present, it can be inferred that all approaches are signalized. When a stop sign is present, it can be assumed that the opposing approach is also stop controlled if it cannot be determined from the forward view. When no control is noted for the approach where the test vehicle was traveling and control for adjacent approaches cannot be determined, it may be difficult to determine if the intersection is uncontrolled or whether the adjacent streets have stop control. When stop control is present for the approach where the vehicle is located and the control cannot be identified for adjacent approaches, it may be difficult to determine whether the intersection is two-way or four-way stop controlled. No instances of advance stop line rumble strips and red-light-running cameras were found. It is possible that they were present but were not identified.

The signal phase (red, yellow, green) could be determined for the approach where the vehicle was traveling and can be inferred for other approaches by vehicle movements. Signal phase state could be identified in almost all situations, with the exception of turn arrows. Turn arrows were very difficult to discern in many cases. Signal progression could be inferred based on the number of times the vehicle received the green phase through a series of intersections.

Several intersection features are shown in Figure C.8 as they were identified in the forward video.

Red-light running is a significant cause of many intersection crashes. The team therefore reviewed the forward video view to determine whether a data reductionist could identify

*Table C.5. Identification of Other Countermeasures and Access*

| Data Element | Features | DAS |
|---|---|---|
| **Other countermeasures** | Type of edge line or shoulder rumble strips | Either was not present or could not be identified |
| | Location of edge line or shoulder RS | Either was not present or could not be identified |
| | Type of centerline rumble strips | Either was not present or could not be identified |
| | Location of centerline rumble strips | Either was not present or could not be identified |
| | Advance stop line rumble strips | Either was not present or could not be identified |
| | Type of speed feedback signs | Encountered one speed feedback sign; speed display was blurred, but it was possible to tell if activated or not |
| | Channelizers, delineators | Yes |
| | Presence of safety edge | NA |
| | Automated speed enforcement | Could only be detected if signing was present; either was not present or could not be located |
| | Cable barrier | Yes |
| | Crash attenuators/cushion | Yes |
| | Vertical deflection (e.g., speed tables, raised intersection) | Was not present in database |
| | Application of high friction surfaces | NA |
| **Access** | Driveway density | Driveways could be identified and counted |
| | Roadway facility type (e.g., collector, arterial) | NA |
| | Access control | Could determine presence and type of median as well as driveways; access could be inferred |
| **Bridge structures** | Type (overpass, underpass, water crossing) | Yes |
| | Bridge deck width | NA |
| | Barriers (e.g., railing) | Yes |
| | Abutments | Yes |

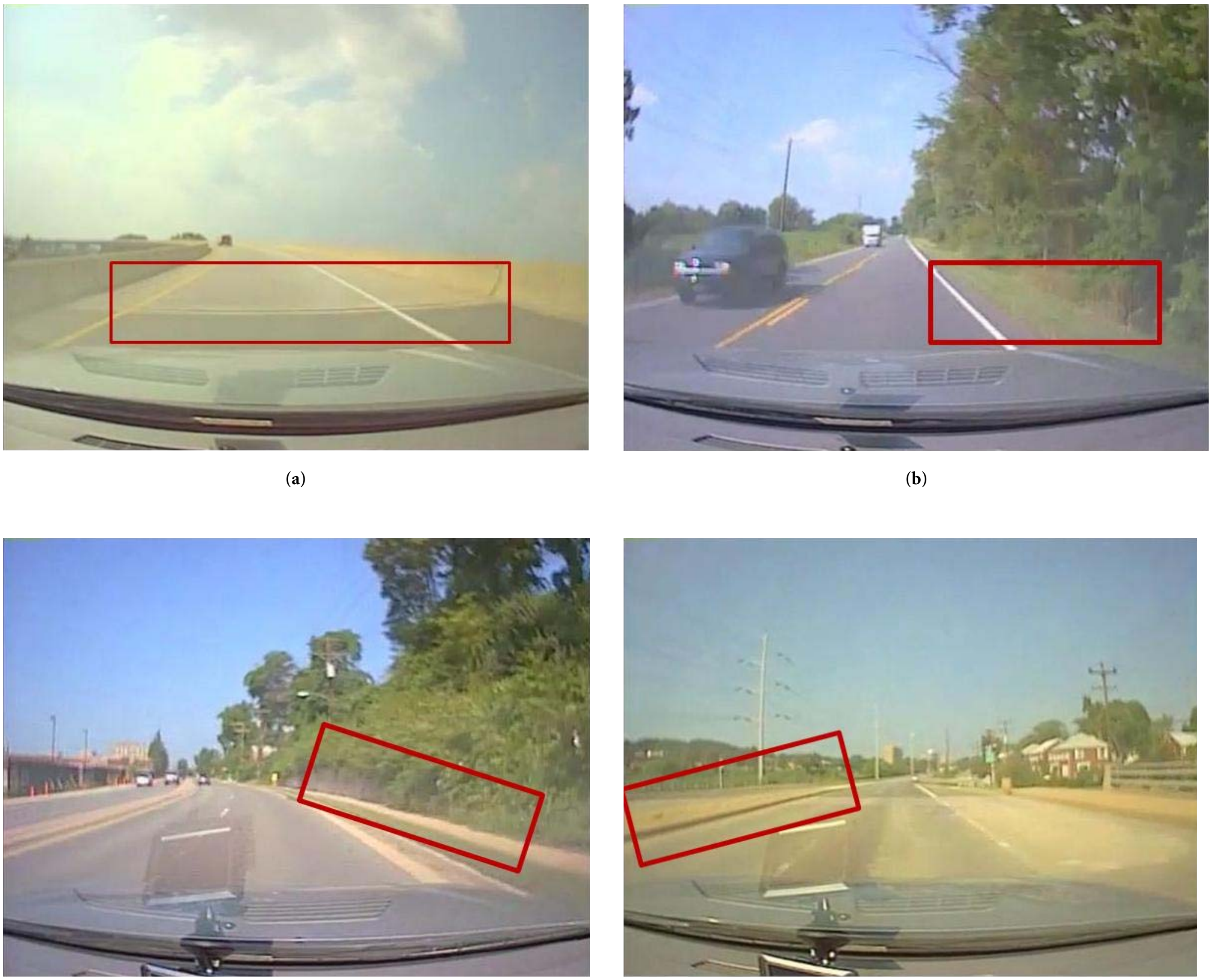
whether the instrumented vehicle ran the red light, although red-light running is technically not a roadway data element. The signal state could be observed in most cases, so the signal change from yellow to red and the position of the instrumented vehicle relative to the stop bar could usually be observed. Figure C.9 shows the signal turn from green to yellow to red. Since signal state can be determined, a situation where the front of the vehicle has crossed the stop bar after the signal turns red could be identified as running the red light. Figure C.10 shows the instrumented test vehicle crossing the stop bar while the signal is yellow.

### *Evaluation of Percentage of Time That Feature Characteristics Can Be Identified*

The second evaluation method assessed the number of times characteristics of a particular type of feature could be identified. The team selected several critical items that would be necessary in evaluating lane departure or intersection crashes. At least 10 of the data items were identified, if present, and a determination made as to whether features could be extracted. For instance, a number of regulatory and advisory signs were identified. Then the number of times the sign message could be detected was recorded. For instance, an advisory sign may be detected by color and shape, but the text or symbol could not be read.

Results are shown in Table C.7 for the majority of features. As the table indicates, chevrons were located three times in the DAS databases. The chevron symbol and number of chevrons could be identified in all three instances. Work zone signs were encountered five times and were assessed to determine whether the sign text or symbol could be interpreted. In all five cases, no text or symbol could be identified. Ten route signs were extracted and 70% of the time the route or guiding

*(text continues on page 137)*



(a) (b) (c) (d)

***Figure C.3. Identification of cross-section elements: (a) illustrates change from asphalt to concrete (surface type); (b) paved/earth shoulder; (c) sidewalk along roadway; and (d) raised median.***

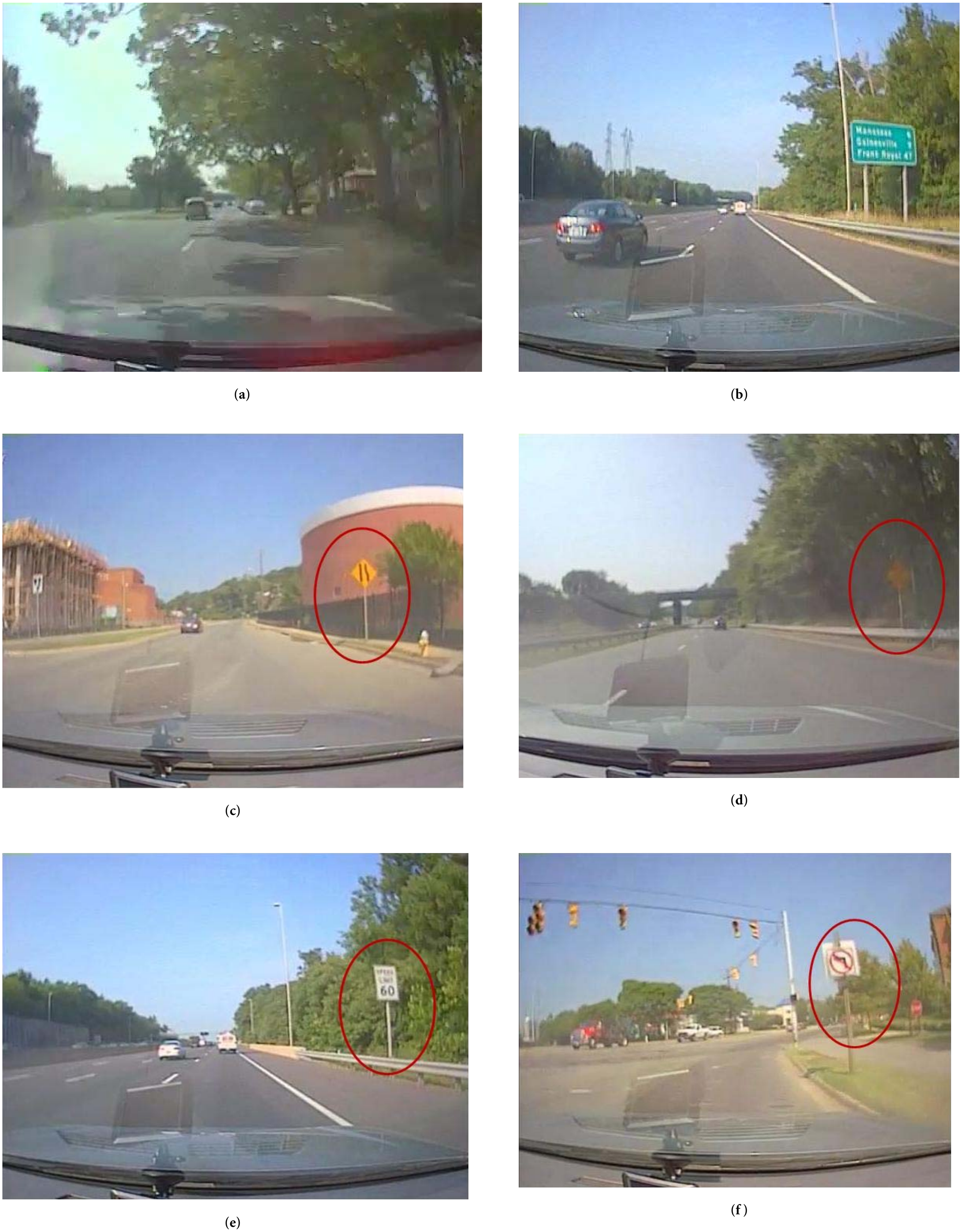



*Figure C.4. Identification of sign types: (a) difficulty identifying signs with glare or vegetation present; (b) guide sign; (c) warning sign face visible; (d) warning sign face not visible; (e) speed limit sign; and (f) turn prohibition.*

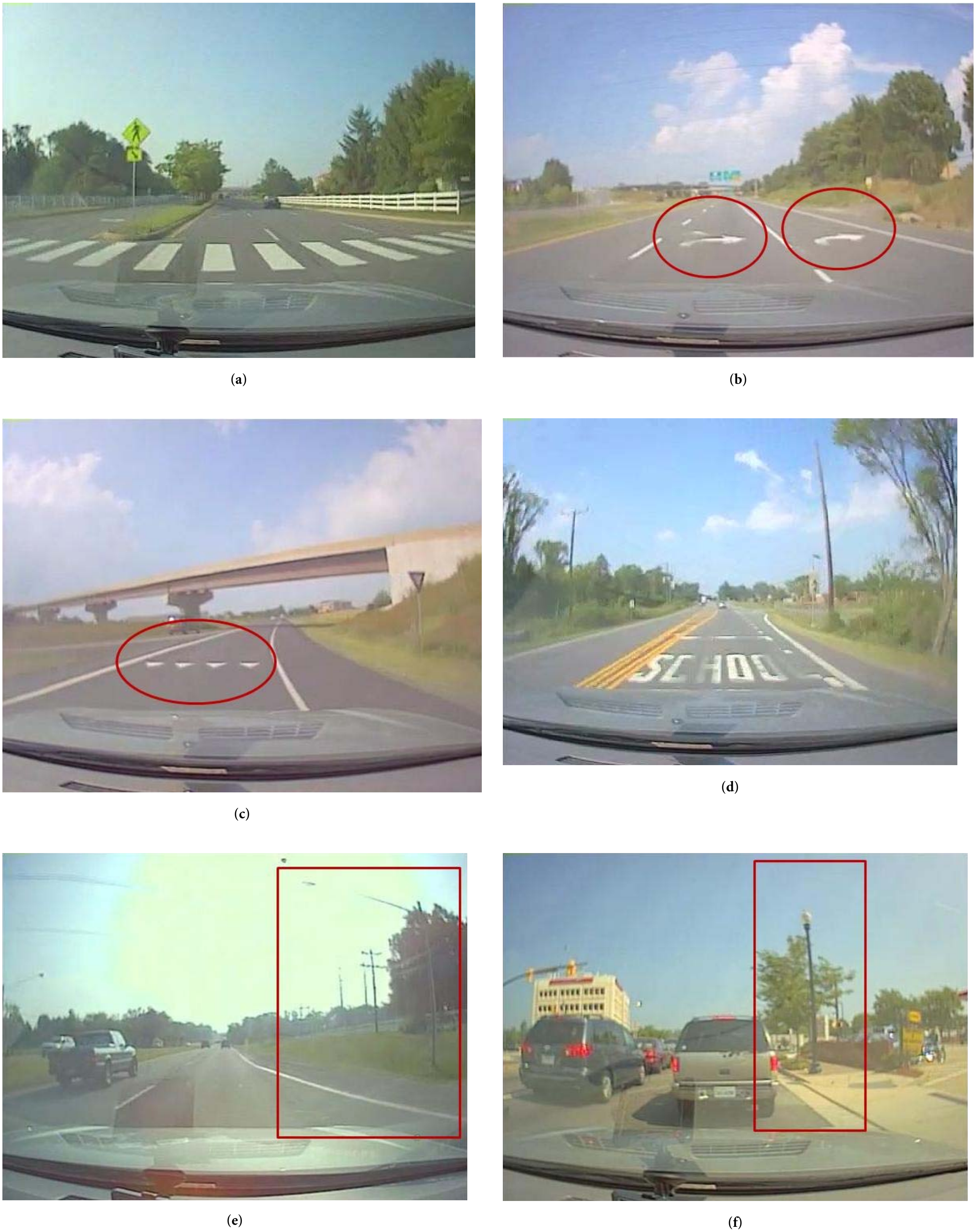



***Figure C.5. Identification of pavement markings and lighting: (a) crosswalk; (b) turn arrows; (c) other on-pavement markings; (d) fading school sign; (e) overhead lighting (cobra head); and (f) decorative street light.***

(a)

(b)

***Figure C.6. Identification of roadway surface and clear zone elements: (a) guardrail within clear zone and (b) pavement damage.***

(a)


(b)

***Figure C.7. Identification of countermeasures: (a) colored pavement marking and (b) speed feedback sign (activated).***

*Table C.6. Identification of Intersection Features*

| Data Element | Features | DAS |
|---|---|---|
| **Cross section** | Intersection/interchange type (includes railroad) | Yes |
| | Number of lanes by approach | Yes; for the approach where the vehicle is located |
| | Number of approaches | Yes |
| | Left- and right-turn prohibitions | If the corresponding sign can be identified |
| | Intersection skew angle | NA |
| | Channelization (islands) | Yes |
| | Intersection offset (whether crossroad approach centerlines are directly opposed or offset by some distance) | NA |
| | Intersection offset distance | NA |
| **Intersection control** | Control by approach (includes railroad crossing signals) | Yes; for the approach where the vehicle is located |
| | Signal characteristic (e.g., signal head configuration, lens size) | Lens head configuration could usually be determined |
| | Type of signalization (e.g., fixed, actuated) | NA |
| | Presence of left-turn arrow (indication of left-turn phasing) | Was very difficult to see turn arrows, but signal head configuration could be determined, so presence of left arrow could be inferred |
| | Detector type | NA |
| | Overhead beacons | Either was not present or could not be identified |
| | Pedestrian signal | Could detect presence if in line of sight when passing through intersection |
| | Pedestrian signal features (e.g., push button) | NA |
| **Other counter-measures** | Advance stop line rumble strips | Either was not present or could not identify |
| | Red-light running countermeasures | NA |
| **Other** | Sight distance | An estimate can be made based on driver's line of sight from forward view |
| | Signal progression | Signal state could be determined in most cases, so progression could be inferred |
| | Traffic signal state | Yes, except for left/right arrow |
| | Red-light running | Yes, if stop line can be identified |

(a)

(b)

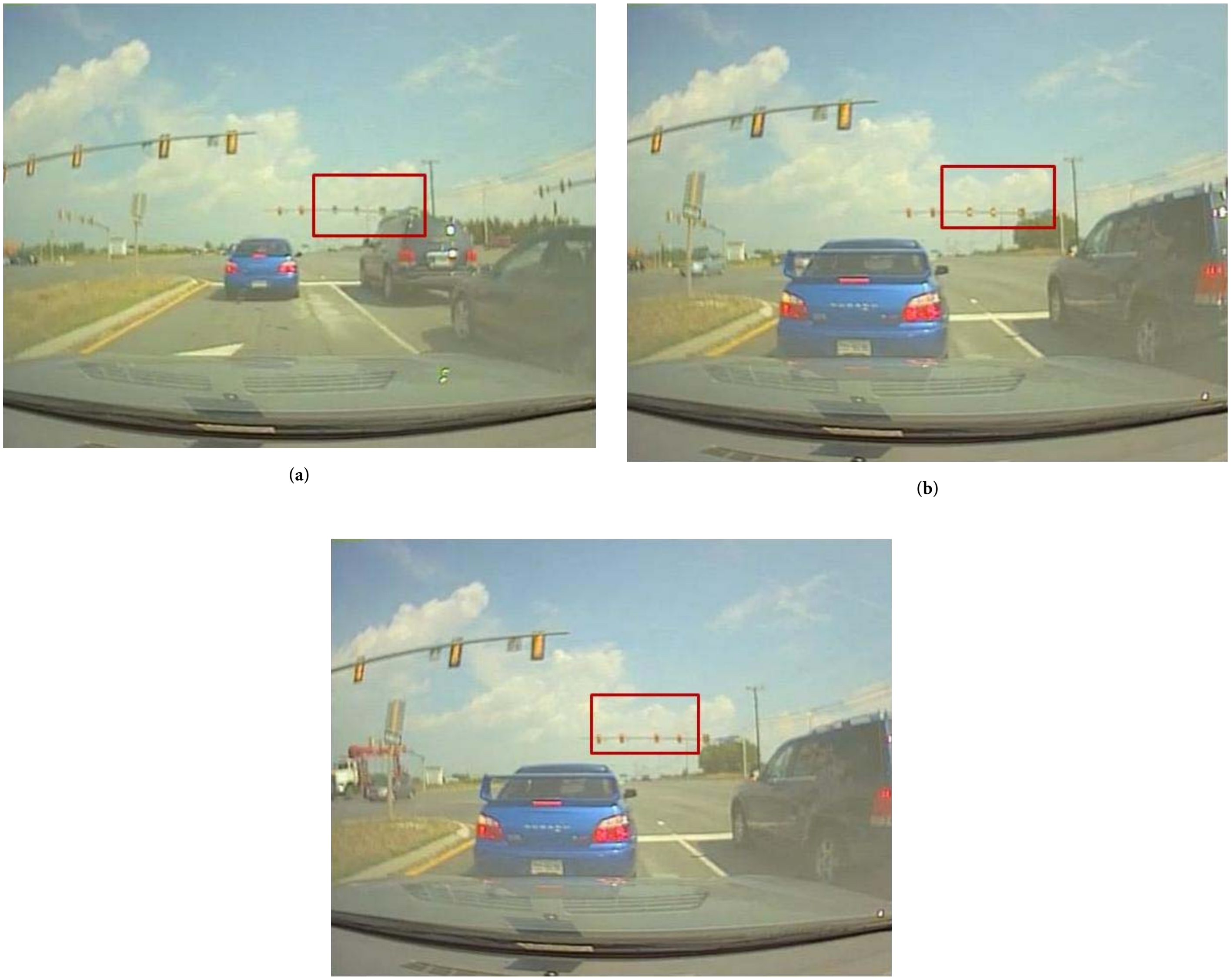
(c)

(d)

(e)

(f)

***Figure C.8. Identification of intersection features: (a) channelization; (b) turn prohibition; (c) signal head configuration; (d) signal phase easily identified (note green left-turn arrow); (e) signal phase difficult to identify for left turn; and (f) signal phase difficult to identify.***



(a)

(b)

(c)

***Figure C.9. Features to identify red-light running: (a) signal identified as green; (b) signal identified as yellow; and (c) signal identified as red.***

*(continued from page 130)*
information could be read. Ten on-pavement markings were also selected and, in all cases, the type of marking could be identified. This included left turn and "SCHOOL." Sixteen speed limit signs were extracted and the numeric speed limit could be identified for all of the signs. In some cases, the text "Speed Limit" may not have been legible, but the sign could be identified since it was distinct. Fifteen regulatory signs other than speed limit, stop, or yield, were identified by shape and color. The sign message could be determined for 73% of the signs. Signals were identified for the approach where the instrumented vehicle was traveling for 26 intersections. The overhead signal phase could be determined 100% for through movements. In some cases it was difficult to determine the state for left-turn arrows.

Thirty-five warning signs were identified and an attempt was made to identify the sign text or symbol. The signs only included those with the traditional diamond shape. Warning signs such as channelizers, which have a distinct shape, were not included in this assessment. When the sign message could be determined, it is indicated in Table C.8. When the message was not legible, it was listed by whether it was text or symbol, since this could be determined even if the exact message could not. The sign message could be determined for around 46% of the signs.


***Figure C.10. Vehicle crossing stop bar during yellow.***

### *Comparison of Data Elements in DAS with Safety Project S03 Rodeo Elements*

Several of the data elements that were collected as part of SHRP 2 Safety Project S03 data collection rodeo with mobile mapping vans were used to determine whether the same data item could be identified in the DAS data set. Data elements from the rodeo data set were overlaid with the GPS vehicle traces from the DAS data. The GPS data points nearest the feature in question were selected and corresponding time stamps noted. The time stamps were located in the forward video and the forward view searched for several frames before and after to locate the object. The process is depicted in Figure C.11. If the object was located, it was noted and compared against the description for the data element in the rodeo data set.

Eighteen signs were extracted from the rodeo data. Two signs could not be located in the DAS in any reasonable proximity. Many of the signs were identified by shape or color, but in many cases the text or symbol was not legible. Results are shown in Table C.9.

Table C.10 shows identification of pavement markings. As noted, seven pavement markings that were selected in the rodeo data could be identified in the forward video.

Miscellaneous other objects were also compared between the two data sets as shown in Table C.11. Two streetlights were located in the mobile mapping data. One could be identified in the forward video in the approximate location. The second streetlight was located in an area of heavy vegetation and significant glare was present in the forward image. As a result, the streetlight could not be distinguished from other background features. One segment was indicated as having centerline rumble strips in the mobile mapping data. The approximate begin

***Table C.8. Assessment of Warning Signs***

| Warning Sign Type | Times Encountered | Times Recognized |
|---|---|---|
| Total warning signs | 35 | 16 |
| End of divided roadway | 3 | 3 |
| Lane ends | 1 | 1 |
| Left-hand curve | 2 | 2 |
| Merge | 2 | 2 |
| Reduced speed ahead | 3 | 3 |
| Right-hand curve | 1 | 1 |
| Signal ahead | 2 | 2 |
| Start divided highway | 1 | 1 |
| Watch for deer | 1 | 1 |
| Unknown symbol | 9 | 0 |
| Unknown text | 10 | 0 |

***Table C.7. Assessment of Ability to Identify Features***

| Data Element | Feature Assessed | Times Encountered | Times Recognized | |
|---|---|---|---|---|
| Chevrons | Chevron symbol and number of chevrons | 3 | 3 | 100.0% |
| Work zone sign | Message | 5 | | 0.0% |
| Guide/route sign | Message | 10 | 7 | 70.0% |
| On-pavement marking | Message | 10 | 10 | 100.0% |
| Speed limit | Speed limit | 16 | 16 | 100.0% |
| Regulatory | | 15 | 11 | 73.3% |
| Signal | Phase | 26 | 26 | 100.0% |
| Stop sign | Message | 3 | 3 | 100.0% |

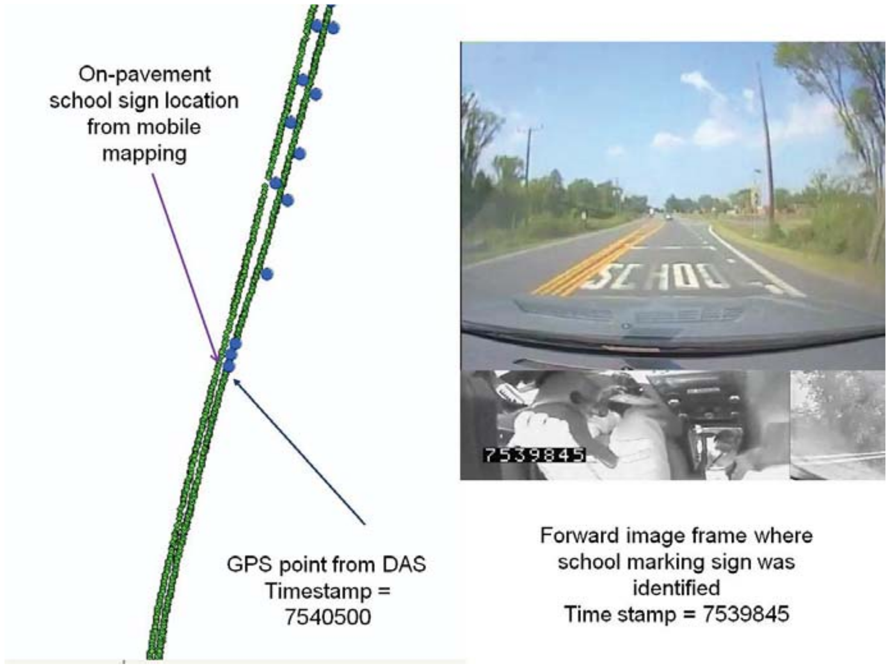



***Figure C.11. Method to compare objects in mobile mapping database with features identifiable in the DAS forward view.***

***Table C.9. Comparison of Signs Between Rodeo Data and DAS***

| Sign Type | How Sign Was Identified |
|---|---|
| Speed limit | Could not locate |
| Exit sign | Could not locate |
| Merge | Merge symbol |
| Food next right | Identified blue sign with text but could not read text |
| Gas next right | Identified blue sign with business symbols |
| Overhead guide | Overhead guide sign |
| Adopt a highway | Identified small blue sign but could not read text |
| School speed limit when flashing | Flashing beacon |
| Object marker | Object marker |
| School sign | School sign |
| State route | Route sign |
| Object marker | Object marker |
| School sign | School sign |
| Right lane must turn | Identified white sign |
| Right lane must turn | Identified sign over right-turn lane |
| Right lane must turn | Identified sign over right-turn lane |
| Stop here on red | By text |

***Table C.10. Comparison of Pavement Markings Between Rodeo Data and DAS***

| Pavement Markings | How Markings Were Identified |
|---|---|
| Stop bar | By shape |
| School | Text “SCHOOL” |
| Stop bar | By shape |
| Right turn only | By shape |
| Right-turn arrow | By shape |
| Stop bar | By shape |
| Stop bar | By shape |

***Table C.11. Comparison of Other Objects Between Rodeo Data and DAS***

| Object | How Object Was Identified |
|---|---|
| Street light | By shape |
| Street light | Could not identify, heavy tree cover and bad sun angle |
| Centerline rumble strips | Identified approximate beginning of rumble strips |
| Guardrail | Identified end of guardrail |

point of the rumble strip section was located. The presence of guardrail in the forward video was easily identified.

## Summary

The team reviewed sample data sets that were collected using what is expected to be the final version of the instrumented vehicle data acquisition system. The data were collected by VTTI using a test vehicle that traversed the same route where data were collected for a demonstration of mobile mapping vendor capabilities. Three methods were used to evaluate how well roadway data elements could be identified using data from the DAS and how feasible data extraction using this method would be.

In the first method, data reductionists reviewed the forward video from the DAS to assess which roadway features could be identified using manual data reduction. The team had developed a list of relevant roadway data elements as part of their work for Safety Project S04A. The team determined which of the data elements could be collected, if applicable. Most features could be recognized to some degree in the DAS forward video. The presence of a sign could be determined in most cases, but text and symbols were frequently illegible.

In the second method, the team extracted a sample of some data elements and examined them to determine what percentage of the time certain features about the element could be identified. For instance the team identified 35 instances of warning signs. The text or symbol could only be recognized slightly less than half of the time.

In the third method, location of several data elements collected in Safety Project S03 for the mobile mapping rodeo were compared with the GPS position and the time stamped from the DAS. The location of the object in the forward video was then identified, if possible.

In summary, a large number of roadway features, including traffic signal state, could be identified in the DAS. The general location of roadway features relative to the roadway could be determined, but actual location could not be established. For instance, it was possible to determine that a school crossing sign was located just before a crosswalk.

## Related SHRP 2 Research

Integration of Analysis Methods and Development of Analysis Plan (S02)
Roadway Measurement System Evaluation (S03)
Roadway Information Database Developer, Technical Coordination,
and Quality Assurance for Mobile Data Collection (S04A)
Mobile Data Collection (S04B)
Design of the In-Vehicle Driving Behavior and Crash Risk Study (S05)
Technical Coordination and Quality Control (S06)
In-Vehicle Driving Behavior Field Study (S07)
Analysis of the SHRP 2 Naturalistic Driving Study Data (S08)

*Membership as of March 2011.