DEFINING, SCREENING, AND TESTING CRASH SURROGATES USING NATURALISTIC DRIVING DATA

A Dissertation in
Civil Engineering
by
Kun-Feng Wu

© 2011 Kun-Feng Wu

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 2011
The thesis of Kun-Feng Wu was reviewed and approved* by the following:

Paul P. Jovanis  
Professor of Civil and Environmental Engineering  
Dissertation Advisor  
Chair of Committee

Martin T. Pietrucha  
Professor of Civil and Environmental Engineering

Eric T. Donnell  
Associate Professor of Civil and Environmental Engineering

Aleksandra B. Slavković  
Associate Professor of Statistics and Public Health Sciences

Peggy A. Johnson  
Professor and Head of Civil and Environmental Engineering

*Signatures are on file in the Graduate School
ABSTRACT

Abundant resources have been devoted to better understand crashes and how they may be prevented from happening. A way to better understand crash occurrence and identify potential countermeasures to improve traffic safety is to learn and use crash surrogates, referred to as surrogate events, events in which drivers need to take an evasive maneuver to recover to normative driving. This dissertation proposes a multi-stage procedure to identify well-defined surrogate events, and an event-based approach to connect crashes and surrogate events. The Virginia Tech 100-Car Naturalistic Driving Study dataset on the NHTSA website is used to demonstrate the event-based approach. The first phase of this dissertation discusses the essence of surrogate events, identifies the desirable criteria for a surrogate event from a diverse literature, and demonstrates the use of an event-based approach for assessing safety by using naturalistic driving data to analyze surrogate events. An analytical procedure to identify surrogate events satisfying the desirable criteria from raw naturalistic driving data is proposed in phase two of this dissertation. Phase three of this dissertation tests the sensitivity of different sets of surrogate events identified. Overall, this framework as a whole is promising, since it not only connects surrogate events and crashes well and has sufficient flexibility to accommodate as much information as possible, but also provides a path for a feedback loop, which ensures robustness of the analysis results.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES ................................................................. vii</td>
</tr>
<tr>
<td>LIST OF TABLES ................................................................. ix</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS ............................................................... x</td>
</tr>
</tbody>
</table>

Chapter 1 Introduction ........................................................................... 1

1.1 Surrogate Events ........................................................................... 2
1.2 Naturalistic Driving Study ............................................................... 4
1.3 An Event-Based Approach to Connect Crashes and Surrogate Events ....... 5
1.4 A Multi-Stage Procedure for Identifying Well-Defined Surrogate Events ...... 5
1.5 Relevant Issues when Analyzing and Screening Naturalistic Driving Data ...... 7
1.6 The Objective of Research ............................................................. 8
1.7 References .................................................................................. 9

Chapter 2 Crashes and Surrogate Events: Exploratory Modeling with Naturalistic Driving Data (*Accident Analysis & Prevention, in press*) ................................................. 11

2.1 Background ................................................................................. 11
- 2.1.1 Conflict Studies- The Foundation for Crash Surrogate Analysis ......... 13
- 2.1.2 The Desirable Criteria for Crash Surrogate ....................................... 16
- 2.1.3 The Objective of Research .......................................................... 18
2.2 Formulation of Crash Surrogate and Conditional Crash Probabilities ....... 19
- 2.2.1 Analytical Foundation of an Event-based Model Approach ............... 19
- 2.2.2 Estimating the Conditional Crash Probability Given a Surrogate Event ... 21
- 2.2.3 Using Conditional Crash Probability to Estimate Expected Number of Crashes ... 25
2.3 The Data ................................................................................... 26
2.4 Application of Method to Road Departure Event Analyses with Naturalistic Driving Data ............................................................................................................. 29
- 2.4.1 Model Estimation ...................................................................... 29
- 2.4.2 Estimating the Expected Number of Crashes using Surrogate Events ... 34
2.5 Summary and Discussion ................................................................ 37
2.6 References .................................................................................. 38

Chapter 3 Defining, Screening, and Validating Crash Surrogate Events Using Naturalistic Driving Data (*Accident Analysis and Prevention, in review, forwarded by the 3rd International Conference on Road Safety and Simulation*) ............................................. 42

3.1 Introduction ................................................................................. 42
3.2 Study Goal ................................................................................. 47
3.3 Methodology .............................................................................. 48
- 3.3.1 The Analytic Procedure ............................................................ 50
3.3.2 First Screening ....................................................................... 52
- 3.3.2.1 Receiver Operating Characteristic (ROC) Curve ..................... 52
Appendix C Factors affecting TPF and FPF at Second screening ..................................................... 122
Appendix D Surrogate events for each sample .................................................................................. 123
Appendix E Conditional Crash Probabilities for Each Sample ......................................................... 124
LIST OF FIGURES

Figure 1-1. Conceptualization of the relationship between crashes and safety-relevant events. .......................................................... 3

Figure 1-2. Conceptualization of the relationship between crashes and surrogate events. ...... 4

Figure 1-3. Procedure for screening and validating crash surrogates from naturalistic data... 7

Figure 2-1. Conceptualization of the relationship between crashes and safety-relevant events. .......................................................... 20

Figure 2-2. Conditional crash probability for a surrogate event. .................................. 30

Figure 3-1. Conceptualization of the relationship between crashes and safety-relevant events. .......................................................... 48

Figure 3-2. Conceptualization of the relationship between crashes and safety-relevant events. .......................................................... 51

Figure 3-3. Illustrative Example for Data at Second Screening.................................... 55

Figure 3-4. (Left) Event of Interest vs. (Right) Event Not of Interest: The Impact of The Selection of the Threshold. ...................................... 62

Figure 3-5. ROC Curves for LATD, LATM, and YAWD................................................ 62

Figure 3-6. Structural Difference on Crash Probabilities between Events with Different Context .......................................................... 64

Figure 3-7. ROC Curve for LATD and YAWD at Second Screening.............................. 66

Figure 4-1. Left: Case 1; Right: Case 2. ............................................................. 80

Figure 4-2. ROC Curves at First Screening .......................................................... 92

Figure 4-3. Illustration of Detected Event of Interest After First Screening for Each Measure............................................................. 96

Figure 4-4. Lat10D sample after validation; Left: Crashes; Right: Near Crashes............. 101

Figure 4-5. Yaw30D sample after validation; Left: Crashes; Right: Near Crashes......... 101

Figure 4-6. Average conditional crash probabilities in terms of event scenario for each sample. ........................................................................ 103

Figure 4-7. Expected vs. observed number of crashes.............................................. 104

Figure 4-8. Cross-validation: Lat10D vs. Yaw30D................................................. 105
Figure 5-1. Analytical Procedure for Crash Surrogate Study. ........................................ 115
LIST OF TABLES

Table 2-1. Hydén(1987) conversion factor table. ................................................................. 14
Table 2-2. Conversion factors between conflicts and accidents (from Hydén (1987), Table 5.12). .................................................................................................................. 15
Table 2-3. I independent binomials...................................................................................... 26
Table 2-4. Summary of kinematic search criteria for events in VTTI study (Dingus, et al., 2005). .......................................................................................................................... 28
Table 2-5. Variable definitions. ........................................................................................... 29
Table 2-6. The estimated event-based models. ................................................................. 33
Table 2-7. Conditional crash probabilities based on VTTI 100-car single vehicle run-off-road events. ..................................................................................................................... 36
Table 3-1. Summary of kinematic search criteria for events in VTTI study (Dingus, et al., 2005). .......................................................................................................................... 46
Table 3-2. Variable Definitions. ........................................................................................... 60
Table 3-3. ROC Curves Analysis for LATM and LATD .......................................................... 63
Table 3-4. Chow Test for Intersection v.s. Non-intersection Related Events.......................... 65
Table 3-5. Survival Analysis in Second Screening. ................................................................. 66
Table 3-6. ROC Curves Analysis for LATD at Second Screening ........................................... 66
Table 3-7. Bivariate Probit Model for Crash Occurrence and Events with LATD Greater Than or Equal to 0.7g............................................................ 67
Table 3-8. The Event-based Model Using Valid Surrogate Events. ........................................ 68
Table 3-9. Conditional Crash Probabilities Using Valid Surrogate Events .............................. 69
Table 4-1. Summary of kinematic search criteria for events in VTTI study. ......................... 83
Table 4-2. Factors affecting ROC Performance...................................................................... 94
Table 4-3. Factors affecting TPF and FPF at first screening.................................................. 94
Table 4-4. The incremental value of Lat30D on Yaw30D.......................................................... 95
Table 4-5. Summary of Outputs from All Stages. ................................................................. 99
ACKNOWLEDGEMENTS

To my advisor, Dr. Jovanis for his support, guidance, and friendship I will be eternally grateful.

To the members of my dissertation committee, Dr. Pietrucha, Dr. Donnell, and Dr. Slavković for their valuable advice and review.
Chapter 1

Introduction

"The attraction of using surrogates is twofold. First, in a situation in which it is not known what are the crash frequency and severity consequences of a decision, and when a decision has to be made, the use of judgment to define sensible surrogates is fully justified. Second, if a clear causal link has been established between crashes and some surrogate, it is often best to observe or predict changes in the surrogate as a stepping stone for estimating the change in crash frequency and severity." (Hauer, 1999).

"The most valuable application of crash surrogates is not only to predict future crashes, but also to identify inappropriate driving behaviors, roadway design and operational deficiencies, or to find some other countermeasures that may help reduce crash risk." Grayson and Hakkert (1987).

Abundant resources have been devoted to understanding traffic crashes and how they may be prevented from happening. A better way to understand crash occurrence and identify potential countermeasures to improve traffic safety is to learn and use crash surrogates, referred to as surrogate events, events in which drivers need to take an evasive maneuver to recover to normative driving. This dissertation first establishes the connection between surrogate events and crashes. This dissertation then develops a multi-stage procedure and utilizes naturalistic driving data to search for well-defined surrogate events.
1.1 Surrogate Events

There has been considerable research conducted over the last 40 or more years concerning the development of surrogate events for assessing traffic safety (e.g. Perkins and Harris, 1967; Datta, 1979; Hauer, 1982; Hydén, 1987; Chin and Quek, 1997; Shankar, et.al., 2008; Tarko et.al., 2009; Jovanis, et.al., 2010; McGehee et al., 2010; Guo et al., 2010). The goal of surrogate research is driven by the perceived need to conduct safety analyses (e.g. identification of sites with promise of improvement or evaluation of safety countermeasure effectiveness) more quickly (before a large number of crashes occur) and with more data than are typically available from law enforcement-reported crash records (Datta, 1979; Grayson and Hakkert, 1987; Archer, 2004). Only recently has a consensus emerged concerning the definition of a crash surrogate. The definition is based on the following relationship (Hauer, 1982; Hauer and Gårder, 1986; Davis et.al., 2008; and Tarko et.al., 2009): Number of crashes expected to occur on an entity during a certain period of time ($\lambda$) = number of surrogate events occurring on an entity in that time ($c$) * crash-to-surrogate ratio for that entity ($\pi$). Mathematically,

$$\lambda = \pi \times c$$

Equation (1) provides a definitional link between the expected number of crashes, $\lambda$; the number of observed surrogate events, $c$; and the crash-to-surrogate ratio, $\pi$. This relationship will be further elaborated using Figure 1-1 and Figure 1-2 in the following discussion.

Figure 1-1 is a conceptualization of the relationship between crashes and surrogate events. Conceptually, all traffic events consist of normal driving and safety-relevant events. Safety-relevant events can be thought as events that have potential to become crashes. Many criteria for defining safety-relevant events have been proposed. For example, events in which two or more road users would collide in less than 1.5 seconds if their movement remained unchanged. Nevertheless, it was found that some safety-relevant events that have been proposed are not
directly related to actual crash risk. This dissertation refers to the subset of safety-relevant events that are strongly associated with crash occurrences as well-defined surrogate events. And hence, crashes are a subset of these well-defined surrogate events. Except for this mathematical formulation, the relationship between crashes and surrogate events can also be presented using a time progression.

Figure 1-1. Conceptualization of the relationship between crashes and safety-relevant events.

Figure 1-2 shows a surrogate-to-crash evolution process with a time progression from stage zero to stage three. The progression of crash occurrence can be described as normal driving (stage zero) runs into a safety-relevant event, an event that have potential to become a crash (stage one). Any safety-relevant event may be classified as a surrogate event in terms of some specific conditions (stage two), and finally a surrogate event can only result in either a crash or near crash (stage three), as shown in Figure 1-2. The screening criteria would include kinematic or vehicle movement-related measures (e.g. time-to-collision less than 1.5 seconds) and additional event attributes (e.g. daytime or nighttime condition). As shown in either Equation (1), Figure 1-1, or Figure 1-2, to utilize surrogate events to enhance our knowledge about traffic safety, there is a need to develop a procedure to identify well-defined surrogate events and a
method to connect the relationship between crashes and surrogate events, which is central to this dissertation.

![Diagram](image)

Figure 1-2. Conceptualization of the relationship between crashes and surrogate events.

### 1.2 Naturalistic Driving Study

Although surrogate events could be observed through a variety of methods such as street cameras, they are best measured using naturalistic driving data. The Virginia Tech Transportation Institute (VTTI) 100-Car Naturalistic Driving Study dataset is used for empirical testing (Dingus et al., 2005); it includes 241 drivers, and 12 to 13 months of data collection for each vehicle. A data acquisition system (DAS), consisting of cameras for video recording, kinematic sensors, radar, lane tracking devices, and a hard drive for data storage, was installed in each vehicle. Two unique features of naturalistic data are important for surrogate analysis. First, vehicles are instrumented with video camera technologies that observe the driver and the road ahead of the vehicle continuously during driving. In addition to the video, other onboard sensors continuously record vehicle accelerations in three dimensions as well as rotational motion along the same axes. Radars are also present to record proximity to other vehicles and potential obstacles on the roadway or roadside. Second, drivers are asked to drive as they normally would (i.e. without specific experimental or operational protocols and not in a simulator or on a test track). The period of observation can vary from several weeks to a year or more. Rather than relying on law enforcement officer judgment or witness recollection, the DAS can record virtually all of the
actions of the subject driver before, during, and after each event. Because events are recorded using video and vehicle sensors, individual events can generally be described with greater accuracy and reliability than using reports assembled after the fact (Dingus et al., 2005; Jovanis et al., 2011).

1.3 An Event-Based Approach to Connect Crashes and Surrogate Events

There is a need to extend and refine the use of surrogate events to enhance safety analyses. This is particularly true given the opportunities for data collection presented by naturalistic driving studies. Chapter two discusses general criteria for defining surrogate events and an approach for modeling the crash-to-surrogate ratio. Much of the historical surrogate research has been focused on traffic conflicts at intersections, and there is a need to expand the crash-to-surrogate concept beyond traffic conflicts to many contexts and crash types. This dissertation summarizes desirable criteria for defining surrogate events from a diverse literature. In addition, this dissertation connects crashes and surrogate events using the crash-to-surrogate ratio, $\pi$, the proportion of surrogate events that end up as crashes. The basic framework is borrowed from the discrete-choice model developed by McFadden (1974). A conceptual structure is developed in which the ratio can be estimated using either a Logit or Probit formulation which can provide sufficient flexibility to capture kinematic variables, event variables, driver attributes, and driving environment as predictors in the model specification.

1.4 A Multi-Stage Procedure for Identifying Well-Defined Surrogate Events

Another major component of this dissertation is to develop a procedure that can be used to screen, identify, and validate well-defined surrogate events. Several steps have been added
prior to the estimation of $\pi$, all of which are directed toward the use of a rigorous modeling structure to identify well-defined surrogate events, a set of events that have common etiology. As such, this dissertation seeks to apply statistical methods as part of the methodology. The process begins with identification of safety-relevant events, referred to as events of interest in this dissertation, using kinematic screening criteria (First Screening). This step is also indicated as the first step in the bottom of Figure 1-3. The process in Figure 1-3 ends with the estimation of the conditional crash probabilities, $\pi$. In addition to the basic identification of candidate events from kinematic signatures, Figure 1-3 proposes several additional steps to continue to refine the crashes and near-crashes to be compared. This continual refinement is motivated by the need to identify and group crashes and near-crashes with similar etiologies. The Classification step in Figure 1-3 develops separate event-based models of crashes and near-crashes after grouping the events identified during the first screening. Rather than relying on a single kinematic predictor or set of predictors, the models estimate the probability of an event occurring as a function of a set of kinematic, context and event predictors. A Chow test is used to identify events with similar etiologies and remove events with similar kinematics but with different contexts. The events that are similar after the classification are then screened again by testing a range of kinematic variables to describe the time-dependencies during the event. Here the focus is on misclassification and the tradeoff in correct prediction and false rejection of events; the Receiver Operating Characteristic (ROC) curve method is one of several that can be used in this step. Some additional events will be discarded at this step as being not similar to those continuing in the algorithm. These refined events are validated as similar using a bivariate Probit or similar method which tests the comparability of the entire model of near-crashes against an entire model estimated from crashes alone. It is expected that the models will not be statistically different at this step, but if they are, then there is an ability to enter the feedback loop and revise the criteria for first screening, developing a new set of events for comparison.
1.5 Relevant Issues when Analyzing and Screening Naturalistic Driving Data

While this dissertation develops an event-based approach to estimate $\pi$ using well-defined surrogate events, screened and validated by the multi-stage procedure, one question that can be posed is: how sure are we that the data set obtained by screening on kinematic variables is the best data set for surrogate analysis? Development of a refined set of surrogate events has received only cursory treatment in past naturalistic analyses, and hence it is unclear how sensitive these surrogate events are to different screening criteria. One direction for this question would be the exploration of algorithms to more thoroughly screen events as potential surrogate events. This dissertation not only tests how different sets of surrogate events would affect analysis results, but also explores and discusses relevant issues when screening and analyzing naturalistic driving data. The issues include: (1) how to more precisely assess the performance of the screening criteria to detect events of interest and identify surrogate events; (2) what would be the impacts of different sets of surrogate events identified from the same raw data on safety analysis; and (3) how to determine when a set of surrogate events is preferred to others. By taking these issues into
consideration, it is anticipated that surrogate events that are comparable, tractable, and transferable will be identified.

1.6 The Objective of Research

The objective of this research is to formulate and test an analytical paradigm for surrogate analysis that integrates desirable attributes from a diverse literature. The components of the paradigm include:

- The development of a framework for a crash surrogate that is estimable using statistical methods,
- Testing of the paradigm using existing naturalistic driving data, and
- Development of recommendations for constructive future research in the area.

This dissertation proposes a multi-stage procedure to identify well-defined surrogate events, and an event-based approach to connect crashes and surrogate events. The VTTI 100-Car Naturalistic Driving Study dataset on the NHTSA website is used to demonstrate the event-based approach. The first phase of this dissertation discusses the essence of crash surrogates, identifies the desirable criteria for a surrogate event from a diverse literature, and demonstrates the use of an event-based approach for assessing safety using naturalistic driving data to analyze surrogate events. An analytical procedure to identify surrogate events satisfying the desirable criteria from raw naturalistic driving data is proposed in phase two of this dissertation. Phase three of this dissertation tests the sensitivity of different sets of surrogate events identified. Overall, this framework as a whole is promising, since it not only connects surrogate events and crashes well and has sufficient flexibility to accommodate as much information as possible, but also provides a path for a feedback loop, which ensures robustness of analysis results.
The organization of this study is as follows. Chapter two describes the development of the event-based approach, which connect the relationship between well-defined surrogate events and crashes. Chapter three presents a multi-stage framework for screening, identifying, and validating well-defined surrogate events. Chapter four discusses all other relevant issues as screening and analyzing naturalistic driving data in searching for well-defined surrogate events. Lastly, chapter five summarizes the findings and provides recommendations for future research.

1.7 References


Chapter 2

Crashes and Surrogate Events: Exploratory Modeling with Naturalistic Driving Data (Accident Analysis & Prevention, in press)

There is a need to extend and refine the use of crash surrogates to enhance safety analyses. This is particularly true given opportunities for data collection presented by naturalistic driving studies. This dissertation connects original research on traffic conflicts to the contemporary literature concerning crash surrogates using the crash-to-surrogate ratio, \( \pi \). A conceptual structure is developed in which the ratio can be estimated using either a Logit or Probit formulation which captures context and event variables as predictors in the model specification. This allows the expansion of the crash-to-surrogate concept beyond traffic conflicts to many contexts and crash types.

The structure is tested using naturalistic driving data from a study conducted in the United States (Dingus, et al. 2005). While the sample size is limited (13 crashes and 38 near crashes), there is reasonable correspondence between predicted and observed crash frequencies using a Logit model formulation. This chapter concludes with a summary of empirical results and suggestions for future research.

2.1 Background

There has been considerable research conducted over the last 40 or more years concerning the development of crash surrogates for assessing traffic safety (e.g. Perkins and Harris, 1967; Datta, 1979; Hauer, 1982; Hydén, 1987; Chin and Quek, 1997; Shankar, et.al., 2008; Tarko et al., 2009; Jovanis, et.al., 2010); McGehee et al., 2010; Guo et al., 2010). The goal
of surrogate research is driven by the perceived need to conduct safety analyses (e.g. identification of sites with promise of improvement or evaluation of safety countermeasure effectiveness) more quickly (before a large number of crashes occur) and with more data than are typically available from law-enforcement-reported crash records (Datta, 1979; Grayson and Hakkert, 1987; Archer, 2004).

Only recently has a consensus emerged concerning the definition of a crash surrogate. The definition is based on the following relationship (Hauer, 1982; Hauer and Gårder, 1986; Davis et al., 2008; and Tarko et al., 2009): Number of crashes expected to occur on an entity during a certain period of time \( \lambda \) = number of crash surrogates occurring on an entity in that time \( c \) * crash-to-surrogate ratio for that entity \( \pi \). Mathematically,

\[
\lambda = \pi c
\]

Here we take the liberty of using the more current term, “surrogate”, instead of the original term “conflict”. Equation (1) must hold for any meaningful crash surrogate. And hence, Equation (1) is the conceptual foundation for this crash surrogate analysis. Equation (1) provides a definitional link between the expected number of crashes, \( \lambda \); the number of observed surrogate events, \( c \); and the crash-conflict conversion factor, \( \pi \). If one can develop a method to estimate \( \pi \) and observe the number of conflicts, \( c \), then one may be able to use conflicts to estimate expected crash frequency.

While there is some agreement concerning a surrogate definition, several other issues remain relatively understudied including methods to identify crash surrogates, tests for the validity of crash surrogates, and the use of crash surrogates to assess road safety. While many crash surrogates have been proposed and studied, much of the crash surrogate research has historically been focused on the Traffic Conflict Technique (TCT) applied at intersections. The exploration of crash surrogates thus begins with a review of this early research followed by a
description of our conceptualization of crash surrogate analysis. The dissertation then tests the conceptualization empirically, followed by conclusions and lessons learned for future research.

### 2.1.1 Conflict Studies- The Foundation for Crash Surrogate Analysis

The most well-known and studied crash surrogate is the traffic conflict. In the first conflict study (Perkins and Harris, 1967), conflicts were defined based on evasive actions taken by drivers such as the appearance of brake lights or sudden lane changes. A more specific definition proposed in the first workshop on traffic conflicts is that a traffic conflict is "an observable situation in which two or more road users approach each other in space and time to such an extent that there is a risk of collision if their movement remained unchanged" (Amundsen and Hydén, 1977). Hydén (1987) suggested that "distance in time" is a better method to rate crash severity than "distance in space" and "deceleration power", and suggested a conflict with time-to-accident less than 1.5 seconds as a serious conflict.

The general approach of a traffic conflict study was to collect crash and conflict data from a number of intersections, and estimate the "conversion factor" (the \( \pi \) in Eq.(1)), which represents the relationship between conflicts and crashes (Hydén, 1987). First, crashes and conflicts were split with regard to the variables "traffic class" and "kind of road users", as shown in Table 2-1. “Traffic class” is a driving context variable, and “kind of road user” is an event type variable. Oppe (1986) suggested a similar classification using maneuver type and severity level.
Table 2-1. Hydén(1987) conversion factor table.

<table>
<thead>
<tr>
<th>Traffic class 1: All situations in low speed intersections and situations in high speed intersections with only turning cars involved</th>
<th>Car-Car Crash/Conflict</th>
<th>Car-Bicycle Crash/Conflict</th>
<th>Car-Pedestrian Crash/Conflict</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x₁, y₁)</td>
<td>(x₂, y₂)</td>
<td>(x₃, y₃)</td>
<td></td>
</tr>
</tbody>
</table>

| Traffic class 2: Situations in signalized intersections with only turning cars involved | (x₄, y₄) | (x₅, y₅) | (x₆, y₆) |

| Traffic class 3: Situations in high speed intersections with at least one straight forward going car involved | (x₇, y₇) | (x₈, y₈) | (x₉, y₉) |

| Traffic class 4: Situations in signalized intersections with at least one straight forward going car involved | (x₁₀, y₁₀) | (x₁₁, y₁₁) | (x₁₂, y₁₂) |

There are 12 observation elements defined in Table 2-1. Each element consists of a set of measures \((x_k, y_k)\) where \(x_k\) is the number of recorded crashes for element \(k\), and \(y_k\) is the number of observed conflicts. \(x_k\) and \(y_k\) are assumed to belong to Poisson processes. Formally,

\[
X_i \sim \text{Poisson}(\pi_k \lambda_k B_k) \tag{1}
\]

\[
Y_i \sim \text{Poisson}(\lambda_k A_k) \tag{2}
\]

Thus \(x_k\) and \(y_k\) belong to Poisson processes with the mean intensity of \(\lambda_k A_k\) and \(\pi_k \lambda_k B_k\). \(A_{ki}\) and \(B_{ki}\) indicate exposure, which is the product of the crossing traffic flows for element \(k\) at intersection \(i\) and time length of study period (Hydén, 1987). \(\lambda_k\) is an intensity that specifies the frequency of conflicts for element \(k\). Let \(x_{ki}\) indicates the number of crashes for element \(k\) observed at intersection \(i\); and \(y_{ki}\) indicates the number of conflicts for element \(k\) observed at intersection \(i\); The conversion factor \(\pi_k\), which specifies the probability that a conflict ends up with a crash in element \(k\), can be obtained as follows:

\[
\pi_k = \left( \frac{\sum x_{ki} A_{ki}}{\sum y_{ki} B_{ki}} \right) \tag{3}
\]
In Hydén's study, the number of conflicts or crashes in some of the 12 elements was very small, leading to uncertainty in estimating the conversion factors for these elements. Hence, Hydén suggested that elements be merged when there is limited data available.

Table 2-2 (taken from Hydén, 1987) illustrates how these conversion factors can be applied. The estimation is based on the data collected at 50 intersections in Malmo, Sweden, during 1974-1975. Values in brackets are the 90 percent confidence intervals. The expected number of crashes for each element is estimated according to the number of observed conflicts and the conversion factor. For traffic class 1+2, and the car-car scenario, the conversion factor is $3.2 \times 10^{-5}$. This means that if we observe 100,000 conflicts a year at an intersection, we would expect to see 3.2 crashes in a year at that intersection. Or, if we observe 300 conflicts in one day, then we expect to see $300 \times 365 \times 3.2 \times 10^{-5} = 3.5$ crashes in a year.

<table>
<thead>
<tr>
<th>Traffic Class</th>
<th>Car-Car</th>
<th>Car-Unprotected road user</th>
</tr>
</thead>
</table>
| 1+2 (CI)      | $3.2 \times 10^{-5}$  
(2.0$\times 10^{-5}$, 6.9$\times 10^{-5}$) | $15.3 \times 10^{-5}$  
(12.2$\times 10^{-5}$, 19.6$\times 10^{-5}$) |
| 3+4 (CI)      | $11.1 \times 10^{-5}$  
(8.2$\times 10^{-5}$, 16.1$\times 10^{-5}$) | $3.2 \times 10^{-5}$  
(70.5$\times 10^{-5}$, 113.3$\times 10^{-5}$) |

There have been several additional studies of traffic conflicts. Risser (1985) found that the sum of all errors in driving behavior shows correlation with the subjects' accidents in the past five years as well as with the subjects' traffic conflicts during a one-hour driving test. Cooper (1984), using a post-encroachment time rather than traditional conflict definition, and Migletz et al. (1985), using conflict numbers without taking the severity of a conflict into account (FHWA method), reported instances where crash experience appeared to increase, have no relation, or decrease as observed conflicts increased. However, Sayed and Zein (1999) found a statistically
significant relationship between 3-year crash frequency and observed conflicts with an $R^2$ of 0.77 at signalized intersections, but no significant relationship at unsignalized intersections.

Some argue that the inconsistency in conflict studies results from problems with the definition, validity, and reliability of conflict measurement (Chin and Quek, 1997). Chin and Quek (1997) discuss the evolution of several definitions for a conflict. They view the main issue as the objective evaluation of a conflict, since it is difficult to judge when drivers start their evasive actions. Hauer and Gårder (1986) suggest that the "quality" of a conflict event definition is associated with the magnitude of the variance of $\pi$. Saunier et al. (2008) propose using an automated system to address issues that hinder the wide use of the traffic conflict technique. Tarko and Songchitruksa (2005) proposed an extreme-value-theory approach to estimate the conversion factor (the $\pi$ in Eq. (1)) based on measured post-encroachment time instead of crash and conflict counts. Davis et al. (2008), focusing on rear-end crashes, proposed a counterfactual approach to estimate the $\pi$ in Eq. (1) by obtaining the probability that an event could have been a crash. These more recent studies suggest a consensus emerging on the importance of $\pi$ in dealing with crash surrogates. Another area of discussion is the identification of desirable criteria for a surrogate, which integrates a number of contemporary issues raised in the literature.

2.1.2 The Desirable Criteria for Crash Surrogate

While much of the historical crash surrogate research has been focused on traffic conflict applications, the recent literature has discussed broader issues concerning surrogates. One area of study has been the desirable attributes of surrogates. Among the criteria suggested for surrogates are:
1. The surrogate should have a short period of data collection (Tarko, 2005) and be more frequent than accidents (Svensson, 1998). This criterion is fundamental to the earliest traffic conflict studies.

2. A surrogate should be correlated with a clinically meaningful outcome (Tarko et al., 2009). One can infer from this definition that a surrogate is an event with attributes similar to crashes (Davis and Swenson, 2006; Davis et al., 2008; Shankar et al., 2008; Jovanis et al., 2010; McGehee et al., 2010; Guo et al., 2010) and useful as a supplement to crashes, especially in understanding crash frequency and severity (Hauer, 1999; Tarko et al., 2009). This criterion supports the notion that both crashes and surrogates are events that are described by multiple dimensions such as context, driver and vehicle, and event attributes (Shankar et al., 2008; Jovanis et al., 2010).

3. A surrogate should have a statistical and causal relationship to crashes (Svensson, 1998; Guo et al., 2010). Closely associated with this concept is the idea that surrogates should have the characteristics of near-crashes in a hierarchical continuum; crashes are at the highest level, and passes with a minimum of interaction are at the lowest level (Svensson, 1998; Guo et al., 2010).

4. Surrogates should fully capture the effect of a treatment in a way similar to how the treatment would affect crashes (Hauer, 1999; Shankar et al., 2008; Tarko et al., 2009). In order for this criterion to be met, the surrogate would have to have contributing factors similar to a crash.

5. Surrogates are "markers" correlated to a crash, with a time scale underpinning (e.g. the crash event is viewed as a time endpoint) (Shankar et al., 2008). This endpoint concept is readily observed for crashes but may be more difficult for surrogates. Despite the difficulty, this criterion argues for strong representation of time-dependencies within the analysis framework of the crashes and surrogates.
Most of the crash surrogates proposed in the literature are surrogate measures with only one metric. Examples include: time-to-collision (e.g. Hyden, 1987; Chin and Quek, 1997); deceleration rate (e.g. Hyden, 1987); post-encroachment time (e.g. Hyden 1987; 1996; Topp, 1998); deceleration-to-safety time (e.g. Topp, 1998); gap time, encroachment time, time-to-zebra (Varhelyi, 1996); proportion of stopping distance (FHWA, 2003); shock-wave frequency (Van Aren and DeVos, 1997); "Jerks" (composite g-force and speed) (Gully et al., 1995; Wahlberg, 2000); standard deviation of lateral position (Vogel, 2003); design consistency (IHSDM, 2008); time-line crossing (Vogel, 2003; Gordon et al., 2009); right-lane departure warning (Gordon et al., 2009); and time-to-right-edge crossing (Gordon et al., 2009). Virtually all of these metrics involve vehicle kinematics; one should recognize that there may be events in which no kinematic trigger is apparent (see Hydén (1987) for an early discussion of this issue). Some distraction or fatigue-related events are examples of such events. These individual metrics may be useful crash surrogates if specifically defined with associated events and placed in the proper context. It is the context that helps to provide the desirably positive attributes represented in surrogate criteria two to five. Davis et al. (2008), Shankar et al. (2008), Jovanis et al. (2009), and McGehee et al. (2010) are among the few who have recognized this explicit connection between context and surrogate metrics.

2.1.3 The Objective of Research

This chapter seeks to provide a framework that will facilitate the use of surrogates in road safety analyses. There is a need to place the early research on traffic conflicts in a more general structure that allows for the use of contemporary data unavailable when traffic conflict studies were undertaken (specifically naturalistic driving data sets). The research develops a conceptual structure for the estimation of $\pi$, the crash-to-surrogate ratio, which forms the foundation for the
approach. Data from a previous naturalistic driving study are used to assess the proposed structure. The chapter concludes with an assessment of the implications of the research, particularly for analyses with naturalistic driving data.

2.2 Formulation of Crash Surrogate and Conditional Crash Probabilities

This section describes the analytical foundation for the proposed paradigm. A discrete outcome formulation allows the paradigm to connect crash and surrogate events in a hierarchy.

2.2.1 Analytical Foundation of an Event-based Model Approach

Figure 2-1 is a conceptualization of the approach used to estimate $\pi$ in Equation 1. Normal driving leads to a series of events that may be of interest for further study based upon a set of screening criteria. Once the events are identified, further modeling is needed to refine the event set and test for consistency in event etiology (i.e. all events in which Prob($Y_1 = 1$)). At this refinement stage, events of interest include crash outcomes and those without crash outcomes (referred to here as near-crashes). This terminology has been inconsistently used in the past as surrogate events are normally thought of as only including those outcomes without a crash. If one is interested in the application of Equation 1, one must include as refined events, all crash and near crash events with consistent etiology (e.g. road departure or rear end). Given the refined set of events for which Prob($Y_1 = 1$), one can then assess $\pi$ in Equation (1).

The initial screening for naturalistic data has been traditionally conducted using vehicle kinematic triggers (e.g. Dingus et.al. 2005), but other triggers are also possible as long as they capture the dynamics unfolding during the event. The grouping of events is a recognition that the crashes and near crashes should have similar etiologies to be meaningfully analyzed. At this point
in the analytic structure we recognize that some of the events of interest, while triggered initially, do not have the same etiology; they should be moved to the branch represented by the $\text{Prob}(Y_1 = 0)$; i.e., an event that is not a crash or near crash of interest. Modeling approaches that are useful at this grouping step should represent the dynamic nature of evolving events (e.g. survival analysis) and also the detection of distracted or fatigue-involved events (Wu and Jovanis, 2011).

In this dissertation, we have a data set in which this classification has already occurred through video review of events. The last step is the modeling of $\pi$. This should be conducted with a categorical outcome model such as a Logit or Probit (Amemiya, 1981; Cameron and Trivedi, 2005).

![Figure 2-1. Conceptualization of the relationship between crashes and safety-relevant events.](image)

Notice that the probability $Y_2 = 1$ given $Y_1 = 1$, represents the conditional probability of a crash given an event identified as a potential surrogate event. The conditional crash probability is interpreted as the proportion of surrogate events that evolve into crashes. For instance, a conditional crash probability of 0.05 indicates that if there are 100 surrogate events (i.e. events in
which \( Y_1 = 1 \), the researcher's best prediction of how many events would evolve to crashes is 5. Recalling the definition of \( \pi \) from Equation (1), one can quickly see the conceptual similarities in the two probabilities. The term \( \pi \) is the proportion of events of interest that are crashes (see Equation 1).

What remains is to develop a model structure to support this conceptualization that can be estimated using naturalistic driving data. Part of this formulation leads to a more refined derivation of the relationship between \( \pi \) and the probability of a crash, given an event, \( Y_1 \). It should be noted that this analysis works with events of interest; it is the application to a specific safety management problem (e.g. network screening) that requires explicit use of exposure using one of several exposure measures (e.g. vehicle miles of travel, annual average daily traffic). The modeling of events and the relationship between crashes and near crashes is the focus of this dissertation, not their application in safety management including exposure.

2.2.2 Estimating the Conditional Crash Probability Given a Surrogate Event

With respect to the conceptualization in Figure 2-1, we assume the events of interest have been identified through a series of kinematic or other search techniques and we have a data set of events with a range of attributes including driver, vehicle, roadway, environment and event-based (Jovanis, et.al., 2010; Jovanis, et.al., 2011; Wu and Jovanis, 2011). In this case, a visual search of the naturalistic driving video was conducted after the kinematic screening to verify that an event of interest had occurred, consistent with common practice in today's naturalistic driving data analysis (e.g. Dingus, et al., 2005). In addition, an algorithm was used to further refine the set of events \( Y_1 = 1 \) to have statistically similar etiology (Wu and Jovanis, 2011).

The basic framework is borrowed from the discrete-choice model developed by McFadden (1974). The idea is to replace the random utility function with a crash function. In the
settings of the discrete-choice model, there are two types of covariates: case-specific variables, which do not vary across choices, and alternative-specific variables, which vary across choices. In the present study, the case-specific variables indicate the event attributes that do not vary across event outcomes, i.e., across near crash or crash outcomes. What is required to obtain these data is to observe either type of outcome in a data set (readily accomplished in naturalistic driving studies).

As an initial formulation, let us consider the case of two outcomes: near crash and crash. Given a particular surrogate event in Figure 2-1 \((Y_1 = 1)\), it can evolve to a crash \((Y_2 = 1)\) with a conditional crash probability \(P\). It can be thought of as a latent process with the crash occurring if the underlying crash function (CF) exceeds some value. Hence, the conditional crash probability can be depicted as follows:

\[
P = \text{Prob}(Y_2 = 1|Y_1 = 1,W,T) = \text{Prob}(R_1 > R_0)
\]

\[
CF_1 = V_1 + \epsilon_1 = B_1W + \Gamma T_1 + \epsilon_1
\]

\[
CF_0 = V_0 + \epsilon_0 = B_0W + \Gamma T_0 + \epsilon_0
\]

where \(CF_1\) is the crash function value which represents the likelihood that the surrogate event evolves to a crash, and \(CF_0\) is the likelihood that the surrogate event evolves to a near crash. The crash functions are employed to describe the surrogate event. \(V\) indicates observed factors in the crash functions, consisting of \(W\) and \(T\), and \(\epsilon\) indicates unobserved factors. \(W\) indicates case-specific variables (event attributes such as roadway contexts), which are the same for both crash and near-crash functions. \(T\) indicates alternative-specific variables. As an example, as we observe a rear-end crash, \(T_1\) is the actual deceleration rate for \(CF_1\), and \(T_0\) is the required deceleration rate for \(CF_0\) to stop the event from becoming a crash (the required deceleration rate that could have made the crash a near crash). However, for a near crash, \(T_0\) is the actual
deceleration rate for $CF_0$, and the $T_1$ for $CF_1$ is the deceleration rate less than the minimum deceleration rate to stop the event from becoming a crash (the deceleration rate that could have made the near crash a crash). If variables of $T$ are available, the model can be estimated with the conditional Logit model (McFadden, 1974). If the $T$ variables are not available, this Logit formulation simply degenerates back to the regular binary Logit model (Greene, 2003). The $B_1$ and $B_0$ vectors capture the effect of changes in event attributes on the crash and near crash functions, respectively ($B_1 \neq B_0$). The $\Gamma$ vector captures the effects of alternative-specific variables on the crash functions. If the crash function ($CF_1$) is greater than the near-crash function ($CF_0$) for a surrogate event, then this surrogate event will be modeled as evolving to a crash, and is given as:

$$ P = \text{Prob}(Y_2 = 1|Y_1 = 1,W,T) = \text{Prob}(CF_1 > CF_0) = \text{Prob}(\epsilon_0 < \epsilon_1 + V_1 - V_0) \quad (8) $$

The Probit model can also be used to construct the conditional crash probability. For the Logit formulation, the unobserved factor $\epsilon$ follows the Gumbel distribution (or called the type I extreme value distribution), and for Probit, the unobserved factor follows a Normal distribution. For a Logit formulation, the density for each unobserved component of the crash function is:

$$ f(\epsilon) = e^{-\epsilon} e^{-\epsilon} \quad (9) $$

and the cumulative distribution is

$$ F(\epsilon) = e^{-\epsilon} \quad (10) $$

If $\epsilon_1$ is given, then

$$ P = \text{Prob}(\epsilon_0 < \epsilon_1 + V_1 - V_0) = e^{-e^{-(\epsilon_1 + V_1 - V_0)}} \quad (11) $$

Since $\epsilon_1$ is actually not given, the conditional crash probability is the integral of all $P$ over all values of $\epsilon_1$ weighted by its density function, $f(\epsilon)$. 
Due to the identification issue for solving both $B_1$ and $B_0$, $B_0$ is normalized to zero, and hence:

$$CF_1 = B_1W + \Gamma T_1 + \varepsilon_1$$

$$CF_0 = \Gamma T_0 + \varepsilon_0$$

where $B_1' = B_1 - B_0$. And now, the conditional crash probability becomes

$$P = \text{Prob}(Y_2 = 1|Y_1 = 1, W, T) = \frac{e^{V_1}}{e^{V_1} + e^{V_0}} = \frac{e^{B_1'W + \Gamma T_1}}{e^{B_1'W + \Gamma T_1} + e^{\Gamma T_0}} = \frac{e^{B_1'W + \Gamma(T_1 - T_0)}}{e^{B_1'W + \Gamma(T_1 - T_0)} + 1}$$

And a logistic regression can be written as:

$$\log\left(\frac{P}{1-P}\right) = \log\left(\frac{e^{V_1}}{1 - e^{V_1}} \right) = \log\left(\frac{e^{V_0}}{1 - e^{V_0}} \right) = V_1 - V_2 = B_1'W + \Gamma(T_1 - T_0)$$

An advantage of using this formulation of crash probability is that the availability of outcome-specific attributes $T$ implies that one can evaluate the treatment effects of countermeasures. As a hypothetical example, suppose that there is a countermeasure that can reduce the required stopping distance or perception-reaction time by 20 percent, then the treatment effect on reducing crash probabilities can be evaluated.

For a Probit formulation, the density for each unobserved component of risk is distributed with a mean vector of zero and covariance matrix $\Omega$. The density of $\varepsilon_n' = (\varepsilon_0, \varepsilon_1)$ is

$$\phi(\varepsilon_n) = \frac{1}{(2\pi)^{1/2} |\Omega|^{1/2}} e^{-\frac{1}{2} \varepsilon_n^t \Omega^{-1} \varepsilon_n}$$

The conditional crash probability is

$$P = \int I(\varepsilon_0 < \varepsilon_1 + V_1 - V_0) \phi(\varepsilon_n) d\varepsilon_n$$
where \( I(.) \) is an indicator of whether the statement in parentheses holds, and the integral is over all values of \( \varepsilon_n \). The integral does not have a closed form. For more detail, please refer to Train (2009).

### 2.2.3 Using Conditional Crash Probability to Estimate Expected Number of Crashes

As shown in Equation (15), the conditional crash probability would vary in terms of variables \( W \) and \( T \). Let event scenario \( i \) represent each combination of \( W \) and \( T \). Equation (15) can be rewritten as:

\[
P_i = \text{Prob}(Y_2 = 1 \mid Y_1 = 1, I = i)
\]

If \( N_i \) represents surrogate events observed for event scenario \( i \), then the probability distribution of the crash count, \( X \), for event scenario \( i \) would obey a binomial distribution with conditional crash probability \( P_i \):

\[
X_i = x_i \sim \text{Binomial}(N_i, P_i)
\]

where \( x_i \) is the number of crashes observed for event scenario \( i \), \( x_i = 1, 2, ..., N_i \), \( P_i \) is the conditional crash probability given \( Y_i = 1 \) and event scenario \( i \). Therefore, given an event scenario \( i \), if \( N_i \) surrogate events are observed, then the expected number of crashes is:

\[
E(X_i) = N_i P_i
\]

Considering \( P_i \) as the conditional crash probability given \( Y_i = 1 \) and event scenario \( i \), \( P_i \) can be written as:

\[
P_i = \frac{\text{Prob}(Y_2 = 1 \mid Y_1 = 1, I = i)}{\text{Prob}(Y_1 = 1, I = i)} = \frac{x_i / N_i}{N_i / N_i} = \frac{x_i}{N_i} = \pi_i
\]

Therefore, \( P_i \) is a more generalized form of the crash-to-surrogate ratio, \( \pi \), in Equation (1). The whole derivation is summarized in Table 2-3. \( P_i \) represents the conditional crash
probability for the surrogate events \((Y_i = 1)\) in terms of event scenario \(i\) (column 2). Given \(N_i\) surrogate events observed in terms of event scenario \(i\) (column 3), the probability distribution of crash counts for each event scenario can be described using a binomial distribution (column 4). The expected number of crashes can be obtained using Equation (21). The statistical model used to estimate a series of conditional crash probabilities in terms of event scenarios is referred to as an event-based model. An application of this formulation is provided with naturalistic data in Section 2.4, Table 2-7.

Table 2-3. I independent binomials.

<table>
<thead>
<tr>
<th>Conditional Crash Probability</th>
<th>Surrogate Event Count (N_i)</th>
<th>Crash Count Generating Process</th>
<th>Expected Crash Count, (E(X_i))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i = 1) (P_1) (N_1)</td>
<td>(\text{Prob}(X_1 = x_1) = \binom{N_1}{x_1} P_1^x (1 - P_1)^{N_1-x_1})</td>
<td>(E(X_1))</td>
<td></td>
</tr>
<tr>
<td>(i = 2) (P_2) (N_2)</td>
<td>(\text{Prob}(X_2 = x_2) = \binom{N_2}{x_2} P_2^x (1 - P_2)^{N_2-x_2})</td>
<td>(E(X_2))</td>
<td></td>
</tr>
<tr>
<td>(\ldots) (\ldots) (\ldots)</td>
<td>(\ldots) (\ldots) (\ldots)</td>
<td>(\ldots) (\ldots) (\ldots)</td>
<td></td>
</tr>
<tr>
<td>(i = I) (P_i) (N_i)</td>
<td>(\text{Prob}(X_i = x_i) = \binom{N_i}{x_i} P_i^x (1 - P_i)^{N_i-x_i})</td>
<td>(E(X_i))</td>
<td></td>
</tr>
</tbody>
</table>

2.3 The Data

The Virginia Tech Transportation Institute (VTTI) 100-Car Naturalistic Driving Study dataset is used for empirical testing (Dingus et al., 2005); it includes 241 primary and secondary drivers, and 12 to 13 months of data collection for each vehicle. A data acquisition system (DAS) consisting of cameras for video recording, kinematic sensors, radar, lane tracking devices, and a hard drive for data storage was installed in each vehicle.

Two unique features of naturalistic data are important for surrogate analysis:
1. Vehicles are instrumented with video camera technologies that observe the driver and the road ahead of the vehicle continuously during driving. In addition to the video, other on-board sensors continuously record vehicle accelerations in three dimensions as well as rotational motion along the same axes. Radars are also present to record proximity to other vehicles and potential obstacles on the roadway or roadside.

2. Drivers are asked to drive as they normally would (i.e. without specific experimental or operational protocols and not in a simulator or test track).

All these data are recorded and stored within an on-board DAS. The DAS for each vehicle is periodically copied into a searchable database and assembled for later analysis. Rather than relying on law enforcement officer judgment or witness recollection, the DAS can record virtually all the actions of the subject driver before, during and after each event. Because events are recorded using video and vehicle sensors, individual events of interest can generally be described with greater accuracy and reliability than using crash reports assembled after the fact (Dingus, et.al. 2005; Jovanis, et.al. 2011).

At certain points during the study, information from the DAS hard drive was received by VTTI, and triggering software was used to identify events of interest (see Table 2-4). Once the triggering events were found in the data, attributes of each event were saved from 30 seconds prior, to 10 seconds after the onset of the precipitating event. Based upon the event criteria in Table 2-4, VTTI researchers identified 69 crash and 761 near crash events during the entire study. These events of interest, as a whole, span the full range of crashes and related events including, e.g. intersection crashes and roadway departure events. To refine the scope of the investigation, we choose road departure events, the topic of several previous papers (e.g. Jovanis et.al. 2011; Jovanis et.al. 2010; Shankar, et.al. 2008). Because the focus of the dissertation was road departure events, the sample size was reduced to 13 crashes and 38 near crashes.
### Table 2-4. Summary of kinematic search criteria for events in VTTI study (Dingus, et al., 2005).

<table>
<thead>
<tr>
<th>Trigger Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Lateral Acceleration</td>
<td>• Lateral accel. ≥ 0.7 g.</td>
</tr>
</tbody>
</table>
| 2. Longitudinal Acceleration | • Accel. or decel. ≥ 0.6g.  
• Accel. or decel. ≥ 0.5 and forward TTC ≤ 4 sec.  
• 0.4g ≤ longitudinal decel. < 0.5g, forward TTC ≤ 4 sec., and forward range at the min. TTC ≤ 100 ft. |
| 3. Event Button         | • Activated by the driver by pressing a button located on the dashboard when an event occurred that he/she deemed critical.                                                                             |
| 4. Forward Time-to-Collision | • Accel. or decel. ≥ 0.5g and TTC ≤ 4 sec.  
• 0.4g ≤ longitudinal decel. < 0.5g, forward TTC ≤ 4 sec., and forward range at the min. TTC ≤ 100 ft. |
| 5. Rear Time-to-Collision | • Rear TTC ≤ 2 sec., rear range ≤ 50 feet, and absolute accel. of the following vehicle > 0.3g                                                                                                           |
| 6. Yaw rate             | • Any value greater than or equal to a plus AND minus 4 degree change in heading (i.e., vehicle must return to the same general direction of travel) within a 3 second window of time.                        |

Various aspects of the driving environment were recorded at the moment of the event, specifically at the onset of the precipitating factor, through the use of video and radar. Table 2-5 is a list of variable names, definitions, and types for run-off-road-related events. All covariates available in the VTTI data set were tested in the analysis; a series of models were used to screen more than 50 individual variables for use as potential predictors. After testing each predictor individually, a series of pair-wise and three-at-a-time models were also tried. This search for predictors resulted in a shortened list of covariates, summarized in Table 2-5.
Table 2-5. Variable definitions.

<table>
<thead>
<tr>
<th>Group</th>
<th>Variable</th>
<th>Definition</th>
<th>Variable Type</th>
<th>Descriptive Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent Variable</td>
<td>Event Outcome</td>
<td>Crash (1); Near Crash (0)</td>
<td>Binary</td>
<td>Crash=13; Near Crash=38</td>
</tr>
<tr>
<td>Situational Characteristics</td>
<td>A straight trajectory before the event started</td>
<td>Go straight (1); otherwise (0)</td>
<td>Binary</td>
<td>Go straight = 32; Otherwise = 19</td>
</tr>
<tr>
<td></td>
<td>The type of trigger for the event is yaw rate, see Table 2-4.</td>
<td>Yaw rate criteria was exceeded (1); otherwise (0)</td>
<td>Binary</td>
<td>Exceeded = 27; Otherwise = 24</td>
</tr>
<tr>
<td></td>
<td>The type of trigger for the event is lateral acceleration, see Table 2-4.</td>
<td>Lateral acceleration criteria was exceeded (1); otherwise (0)</td>
<td>Binary</td>
<td>Exceeded = 11; Otherwise = 40</td>
</tr>
<tr>
<td></td>
<td>Maximum absolute value of lateral acceleration before the end of an event</td>
<td>Continuous variable (g-force)</td>
<td>Continuous</td>
<td>Max = 2.2 g; Min = 0.06 g</td>
</tr>
<tr>
<td></td>
<td>Event duration</td>
<td>Continuous variable (second)</td>
<td>Continuous</td>
<td>Max = 12.4 sec; Min = 1.8 sec</td>
</tr>
<tr>
<td></td>
<td>Road departure (left) before the event</td>
<td>Road departure (1); otherwise (0)</td>
<td>Binary</td>
<td>Departure = 13; Otherwise = 38</td>
</tr>
<tr>
<td></td>
<td>Road departure (right) before the event</td>
<td>Road departure (1); otherwise (0)</td>
<td>Binary</td>
<td>Departure = 28; Otherwise = 23</td>
</tr>
<tr>
<td></td>
<td>Driver was distracted</td>
<td>Distracted (1); otherwise (0)</td>
<td>Binary</td>
<td>Distracted = 39; Otherwise = 12</td>
</tr>
<tr>
<td></td>
<td>The presence of driver fatigue</td>
<td>Driver fatigue (1); otherwise (0)</td>
<td>Binary</td>
<td>Fatigue = 18; Otherwise = 33</td>
</tr>
<tr>
<td>Driving Context</td>
<td>Event occurred on a horizontal curve</td>
<td>Curve (1); otherwise (0)</td>
<td>Binary</td>
<td>Curve = 21; Otherwise = 30</td>
</tr>
<tr>
<td></td>
<td>The presence of daylight</td>
<td>Daylight (1); otherwise (0)</td>
<td>Binary</td>
<td>Daylight = 29; Otherwise = 22</td>
</tr>
<tr>
<td></td>
<td>Dry pavement surface</td>
<td>Dry (1); Wet/Icy/Snowy (0)</td>
<td>Binary</td>
<td>Dry = 40; Wet Snowy = 11</td>
</tr>
<tr>
<td></td>
<td>Roadway with median</td>
<td>Divided (1); otherwise (0)</td>
<td>Binary</td>
<td>Divided = 23; Otherwise = 28</td>
</tr>
<tr>
<td></td>
<td>The event was occurred in a rural area</td>
<td>Rural area (1); otherwise (0)</td>
<td>Binary</td>
<td>Rural area = 24; Otherwise = 27</td>
</tr>
</tbody>
</table>

2.4 Application of Method to Road Departure Event Analyses with Naturalistic Driving Data

2.4.1 Model Estimation

Figure 2-2 highlights the portion of the conceptualized relationship between road departure crashes and near crashes (from Figure 2-1) that is the focus of this chapter. The conditional crash probabilities are estimated using a binary Logit model, since the alternative-specific variables are not available. The application focuses on the ability to take a set of potential
surrogate events (i.e. a set of events that include crashes and near-crashes) and differentiate the crashes and near crashes. This differentiation, described in concept in Section 2.2, allows us to estimate $\pi$, the crash-to-surrogate ratio, for a range of environmental and event conditions. Several models are estimated and a best model selected. The model is then used to compare predicted and observed crash and near-crash outcomes.

After the search of potential covariates described in Section 2.3 a model was estimated using significant predictors from all the testing; this model is shown as model 1 in Table 2-6. The remaining two models show the results of removing additional insignificant predictors. Given the limited sample size, we are less concerned about using a fixed level of significance for variable exclusion and more interested in retaining predictors that have rational sign and magnitude.

![Figure 2-2. Conditional crash probability for a surrogate event.](image)

All of the models have reasonable model fit indicated by the likelihood ratio test. Five predictors are dropped from Model 1 to develop model 2. The pseudo-$R^2$ value drops from 0.37 to 0.33 and all remaining parameters have the same sign and similar magnitudes. Then the presence of a horizontal curve is dropped in Model 3; this changes the magnitude of all remaining parameters by a factor of 2 or more, while lowering the pseudo-$R^2$ to 0.19. These trends indicate that model 2 best balances the number of parameters and model fit and is therefore interpreted in detail in the following paragraphs.
Driver distractions and fatigue were not found to significantly affect the conditional probability of a crash, given a surrogate event. Note that this is not a test of fatigue and distraction as a crash contributing factor; rather it is a test, given a surrogate event, of the association between crashes and near crashes among those events.

Surrogate events with lateral acceleration rate greater than 0.7g are 24.36 times ($\exp(3.193)$) more likely to end up with crashes than those with lower lateral accelerations. This increase in crash odds is measured for only the 1.8 to 12.4 second duration of the event itself; during this short period of time, the event probability increases by a factor of 24. Vehicles with a straight trajectory before running into the surrogate event situation are 85 percent (1-$\exp(-1.882)$) less likely to end up with crashes; however, it is only significant at 90 percent level of confidence (p-value = 0.1). Surrogate events that occurred on dry surface are 89.5 percent less likely to end up in crashes compared to wet and snowy surface. The result is also quite robust, because the coefficients only change slightly and remain strongly significant across the three models in Table 2-6. This finding supports the intuition that dry surfaces can reduce vehicle stopping distance, and hence reduce crash risk given all other things being equal. Surrogate events that occurred with the presence of a median are 70 percent less likely to end up with crashes; nevertheless, it is barely significant with a p-value of 0.2.

The event-based model also indicates that the surrogate events that occurred on a horizontal curve and in a rural area are 25.3 and 5.9 times more likely to be near crashes (i.e. less likely to be crashes) compared to other geometrics and area type respectively. Examining the data set found that there are only 4 crashes but 17 near crashes that occurred with the presence of a horizontal curve; therefore, the estimated odds ratio is close to zero (see odds ratio for the presence of a horizontal curve in Table 2-6 model 2). Similarly, there are also only four crashes but 20 near crashes that occurred in rural areas.
It is worth noting that the presence of a horizontal curve is a very influential predictor. Evidence of the importance of the presence of a horizontal curve is the change in goodness-of-fit when the variable is removed: the pseudo $R^2$ is 0.332 with horizontal curvature but 0.188 without (see model 3). Surrogate events with the presence of daylight condition are 82.9 percent less likely to become crashes.
Table 2-6. The estimated event-based models.

<table>
<thead>
<tr>
<th>Group</th>
<th>VARIABLES</th>
<th>Coef.</th>
<th>Odds Ratio</th>
<th>S.E.</th>
<th>P-value</th>
<th>Coef.</th>
<th>Odds Ratio</th>
<th>S.E.</th>
<th>P-value</th>
<th>Coef.</th>
<th>Odds Ratio</th>
<th>S.E.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Situational Characteristics</strong></td>
<td>The type of trigger for the event is yaw rate</td>
<td>-1.341</td>
<td>0.262</td>
<td>1.539</td>
<td>0.384</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lateral acceleration rate greater than 0.7g</td>
<td>2.543</td>
<td>12.723</td>
<td>20.132</td>
<td>0.108</td>
<td>3.193</td>
<td>24.353</td>
<td>1.449</td>
<td>0.028</td>
<td>3.774</td>
<td>3.208</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Event duration</td>
<td>-0.021</td>
<td>0.979</td>
<td>0.028</td>
<td>0.441</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>A straight trajectory before the event started</td>
<td>-1.775</td>
<td>0.169</td>
<td>1.233</td>
<td>0.15</td>
<td>-1.882</td>
<td>0.152</td>
<td>0.177</td>
<td>0.015</td>
<td>-0.629</td>
<td>0.533</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Road departure (left) before the event</td>
<td>0.913</td>
<td>2.492</td>
<td>2.196</td>
<td>0.677</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Road departure (right) before the event</td>
<td>0.453</td>
<td>1.573</td>
<td>1.64</td>
<td>0.782</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Driver was distracted</td>
<td>-0.071</td>
<td>0.932</td>
<td>1.375</td>
<td>0.959</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>The presence of driver fatigue</td>
<td>0.694</td>
<td>2.002</td>
<td>1.06</td>
<td>0.513</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Driving Context</strong></td>
<td>Dry pavement surface</td>
<td>-2.739</td>
<td>0.065</td>
<td>1.365</td>
<td>0.045</td>
<td>-2.258</td>
<td>0.105</td>
<td>1.091</td>
<td>0.039</td>
<td>-1.344</td>
<td>0.261</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Roadway with median</td>
<td>-1.832</td>
<td>0.16</td>
<td>1.514</td>
<td>0.226</td>
<td>-1.214</td>
<td>0.297</td>
<td>0.961</td>
<td>0.206</td>
<td>-0.848</td>
<td>0.428</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Presence of a horizontal curve</td>
<td>-3.166</td>
<td>0.042</td>
<td>1.488</td>
<td>0.033</td>
<td>-3.23</td>
<td>0.04</td>
<td>1.424</td>
<td>0.023</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>The event was occurred in a rural area</td>
<td>-1.742</td>
<td>0.175</td>
<td>1.008</td>
<td>0.084</td>
<td>-1.778</td>
<td>0.169</td>
<td>0.943</td>
<td>0.059</td>
<td>-1.171</td>
<td>0.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>The presence of daylight</td>
<td>-1.528</td>
<td>0.217</td>
<td>1.227</td>
<td>0.213</td>
<td>-1.764</td>
<td>0.171</td>
<td>1.008</td>
<td>0.08</td>
<td>-0.851</td>
<td>0.427</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Constant term</td>
<td>5.822</td>
<td>4.317</td>
<td>2.62</td>
<td>0.026</td>
<td>5.317</td>
<td>4.317</td>
<td>1.929</td>
<td>0.025</td>
<td>1.227</td>
<td>0.202</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Observations: 51 | 51 | 51
Pseudo R-squared: 0.366 | 0.332 | 0.188
Initial Log likelihood: -28.95 | 28.95 | 28.95
Convergent Log likelihood: -18.35 | -19.34 | -23.52
Likelihood Ratio Test (chi-square): 21.21 | 19.22 | 10.86
Likelihood Ratio Test (p-value): 0.069 | 0.0075 | 0.0929
2.4.2 Estimating the Expected Number of Crashes using Surrogate Events

Table 2-7 summarizes the estimation of the expected number of crashes using the surrogate events. The specific event scenario is described by vehicle movement-related variables and event attributes contained in columns two to eight of Table 2-7. The estimated conditional crash probabilities from model three (appearing in column nine) and observed surrogate events for the specific context from the data (column ten) are applied using the binomial models of Table 2-3 and Equation 21 to produce the expected number of crashes (column 11). For event scenario 24, the estimated conditional crash probability is 0.337. The context for this event is defined by the predictor variables for the model as summarized in columns two to eight of Table 2-7: the vehicle's trajectory is not straight; the vehicle's maximum lateral acceleration is less than 0.7g, on a non-dry surface; a non-divided travel way; with the a horizontal curve during daytime condition, and in a non-rural area. Given that two surrogate events have been observed, one expects $2 \times 0.337 = 0.674$ crashes. One of these two surrogate events resulted in a crash, so it seems that the prediction in this context is reasonable, at least for this data set. A comparison of columns 11 and 12 in Table 2-7 indicates good correspondence between the expected and actual number of crashes for some contexts and poorer correspondence in others. For example the correspondence in contexts 31 to 34 is quite good, while for contexts 7 and 17 it is quite poor. The authors do not wish to speculate any further about the accuracy of the method as tests with larger sample sizes are called for. Such data should be readily available in the SHRP 2 naturalistic driving study currently underway in the U.S., and in other naturalistic studies in preparation around the world (SHRP 2, 2010).

What is the value of these computations? The calculations illustrate that a model of crash and near-crash events can be estimated using event-based data readily available from naturalistic driving studies. Further, the model can be used to estimate the expected number of crashes given an observed number of surrogate events, and that these predicted crashes can then be compared to
actual crashes. As such, the model provides the basis for testing the efficacy of surrogates in predicting expected crash frequency, at least for this data set. Given that these data would be routinely available in naturalistic studies, this supports the use of surrogate events data in safety studies.
Table 2-7. Conditional crash probabilities based on model 2.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.0004</td>
<td>1</td>
<td>0.000</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.0024</td>
<td>1</td>
<td>0.002</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.0027</td>
<td>1</td>
<td>0.003</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.0089</td>
<td>1</td>
<td>0.009</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.0153</td>
<td>1</td>
<td>0.015</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.0334</td>
<td>3</td>
<td>0.100</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.0451</td>
<td>3</td>
<td>0.135</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.0504</td>
<td>1</td>
<td>0.050</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.0545</td>
<td>1</td>
<td>0.055</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.0565</td>
<td>1</td>
<td>0.056</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.0572</td>
<td>5</td>
<td>0.286</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.0631</td>
<td>1</td>
<td>0.063</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.1294</td>
<td>1</td>
<td>0.129</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.1678</td>
<td>3</td>
<td>0.503</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.1698</td>
<td>1</td>
<td>0.170</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.1794</td>
<td>1</td>
<td>0.179</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.1849</td>
<td>2</td>
<td>0.370</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.1999</td>
<td>3</td>
<td>0.600</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.2366</td>
<td>1</td>
<td>0.237</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.2616</td>
<td>2</td>
<td>0.523</td>
<td>1</td>
</tr>
<tr>
<td>21</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.2746</td>
<td>1</td>
<td>0.275</td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.2850</td>
<td>1</td>
<td>0.285</td>
<td>1</td>
</tr>
<tr>
<td>23</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.3337</td>
<td>1</td>
<td>0.334</td>
<td>0</td>
</tr>
<tr>
<td>24</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.3370</td>
<td>2</td>
<td>0.674</td>
<td>1</td>
</tr>
<tr>
<td>25</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.3588</td>
<td>2</td>
<td>0.718</td>
<td>0</td>
</tr>
<tr>
<td>26</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5348</td>
<td>1</td>
<td>0.535</td>
<td>1</td>
</tr>
<tr>
<td>27</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5441</td>
<td>2</td>
<td>1.088</td>
<td>0</td>
</tr>
<tr>
<td>28</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.5732</td>
<td>1</td>
<td>0.573</td>
<td>0</td>
</tr>
<tr>
<td>29</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.6501</td>
<td>1</td>
<td>0.650</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.6585</td>
<td>1</td>
<td>0.659</td>
<td>1</td>
</tr>
<tr>
<td>31</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.8868</td>
<td>1</td>
<td>0.887</td>
<td>1</td>
</tr>
<tr>
<td>32</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.9195</td>
<td>1</td>
<td>0.919</td>
<td>1</td>
</tr>
<tr>
<td>33</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.9278</td>
<td>1</td>
<td>0.928</td>
<td>1</td>
</tr>
<tr>
<td>34</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.9893</td>
<td>1</td>
<td>0.989</td>
<td>1</td>
</tr>
</tbody>
</table>
2.5 Summary and Discussion

This chapter seeks to provide a framework that will facilitate the use of surrogates in road safety analyses. The dissertation began with a review of the traffic conflicts literature and compares analytic developments in traffic conflicts to the more contemporary notion of crash surrogates. Particular attention is drawn to the parameter $\pi$, the crash to surrogate ratio (see Equation (1)). A conceptual structure for the estimation of $\pi$ forms the foundation for our analytical approach. Data on road departure crashes and near crashes from the VTTI 100-car study (Dingus, et.al. 2005) are used to test our proposed modeling structure. The findings are expressed as increases or decreases in a conditional probability: the probability of a crash, given that an event is identified as being of interest (i.e. meeting the screening criteria listed in Table 2-4). Among the empirical findings are that the conditional probability of a crash increases by a factor of 24 when there is a lateral acceleration in excess of $0.7g$ but decreases for many other factors such as the presence of a roadway median, a dry pavement and the event occurring in daylight. The accuracy of model predictions compared favorably to observed crash frequencies (see Table 2-7).

While the results are promising, there are limitations to the dissertation. The dissertation should be considered exploratory because the empirical testing was conducted with a data set of limited sample size and which was prescreened by VTTI using video review. Event screening with raw vehicle kinematic data is expected to be labor intensive and time consuming. There has been limited attention paid to statistical modeling as part of event screening (see Wu and Jovanis (2011) for an example of such modeling) and also as part of the analysis of kinematics to assess potentially safety countermeasure effectiveness (e.g. Volvo, 2005; Battelle, 2006; Fitch et al., 2008). The advantage of statistical modeling is that it provides for the ability to repeat the experiment in different settings so better controls are applied to surrogate screening and crash
kinematics. Neither of these analyses should be considered as a replacement for video review; rather they represent opportunities to be more systematic in naturalistic driving data analyses, proving opportunities for more scientific, repeatable experiments.

2.6 References


Chapter 3

Defining, Screening, and Validating Crash Surrogate Events Using Naturalistic Driving Data (Accident Analysis and Prevention, in review, forwarded by the 3rd International Conference on Road Safety and Simulation)

Naturalistic driving studies provide an excellent opportunity to better understand crash causality and to supplement crash observations with a much larger number of near crash events. The goal of this research is the development of a rigorous set of diagnostic procedures to identify and validate useful crash and near crash events that can be used in enhanced safety analyses. As such, the research seeks to apply statistical methods as part of the methodology. A way to better understand crash occurrence and identify potential countermeasures to improve safety is to learn from and use near-crash events, particularly those near-crashes that have a common etiology to crash outcomes. This chapter demonstrates that a multi-stage modeling framework can make the analysis of naturalistic driving data tractable. The procedure is tested using data from the VTTI 100-car study for road departure events. A total of 51 non-intersections and 12 intersection-related events are included in an application of the framework. While the sample sizes are limited in this empirical study, it is believed that the procedure is ready for testing in other applications.

3.1 Introduction

Considerable research has been conducted over the last 30 years on the development of crash surrogates for assessing traffic safety (Datta, 1979; Hauer, 1982; Hydén, 1987; Chin and Quek, 1997; Archer, 2004; Shankar, et al., 2008; Tarko et al., 2009; McGehee et al., 2010; Jovanis, et al., 2010; Guo et al., 2010). Nevertheless, there is limited agreement concerning fundamental issues such as the definition of a surrogate, the identification of a surrogate from
field data and the validation of particular events as crash surrogates. The lack of agreement has hindered the ability of researchers and practitioners to rigorously use crash surrogates in traffic safety studies.

One area of emerging agreement is the definition of a surrogate (Hauer, 1982; Hauer and Gardner, 1986; Davis et al., 2008; Shankar, et al., 2008; and Tarko et al., 2009). As stated by Hauer, it is,

\[
\text{Number of crashes expected to occur on an entity during a certain period of time (} \lambda \text{)} = \text{crash-to-surrogate ratio for that entity (} \pi \text{)* number of crash surrogates occurring on an entity in that time (} c \text{) or:}
\]

\[
\lambda = \pi c
\]

This statement and its application by several researchers provide support for the view of surrogates as linked to crashes through the ratio, \( \pi \).

Another perspective is provided by Grayson and Hakkert (1987) who suggest that surrogates are more than simple replacements for crashes; they believe that they should be studied for their own insights. This discussion leads one to see that the literature already reveals challenges in the use of crash surrogates; most of this literature evolved from an interest in a particular surrogate, the traffic conflicts technique, first proposed by Perkins and Harris (1967) and codified in a series of studies by Hydén (1987). Interestingly, Williams (1980) argued that the absence of standard techniques for defining surrogates in traffic conflict studies led to the production of a series of research results which were difficult to compare. One of the goals of the research by Hydén and his colleagues was the standardization of traffic conflict measurement so that results could be compared across studies.

The emerging use of naturalistic driving studies offers the unique opportunity to observe both crashes and near crash events as they occur on the road. The Strategic Highway Research Program 2 (SHRP 2) has a safety program which has recognized the importance of surrogates as a
potential enhancement to safety research and has already resulted in several studies with surrogates as at least part of their focus (e.g. SHRP 2 web site).

Naturalistic driving has been applied to studies of drivers from the regular driving population (e.g., Dingus et al., 2005), truck drivers (e.g., Hanowski et al., 2005; Hanowski et al., 2007a; Hanowski et al., 2007b), young drivers and older drivers (VTTI web site, 2010). There have also been a series of technology tests of on-board safety equipment that have used the naturalistic technique (e.g. Bogard et al., 1998; LeBlanc et al., 2006; University of Michigan Transportation Research Institute and General Motors Research and Development Center (UMTRI), 2005).

There are two distinguishing features of naturalistic driving studies. First, vehicles are instrumented with an array of sensing technologies (e.g. video cameras, radars, GPS, accelerometers, gyroscopic sensors) that observe the driver and the road ahead of the vehicle continuously during driving. As a result, events of interest such as crashes and near crashes are recorded with multiple sensors, allowing unprecedented opportunities to gain insight on crash etiology. Second, drivers are asked to drive as they normally would (i.e. without specific experimental or operational protocols and not in a simulator or test track). The period of observation can vary from several weeks to a year or more.

All these data are recorded and stored within an on-board data acquisition system (DAS). The DAS for each vehicle is periodically copied into a searchable data base and assembled for later analysis. Rather than relying on law enforcement officer judgment or witness recollection, the DAS can record virtually all the actions of the subject driver before, during and after each event. Because events are recorded using video and vehicle sensors, individual events of interest can generally be described with greater accuracy and reliability than using crash reports assembled after the fact.
Crash and near crash events in naturalistic driving are typically identified through the detection of unusual vehicle kinematics recorded electronically through accelerometers and gyroscopic sensors. Table 3-1 is an example of search criteria used to identify events for the VTTI 100-car study (Dingus et al., 2005). Vehicle-based accelerometer gyros are used to measure lateral and longitudinal acceleration and vehicle rotation; these measures are used individually or with time-to-collision (TTC) estimates from radar to initially identify potential events. The driver may also highlight a driving event by using an "event" button located in the vehicle for this purpose. Forward and rear TTC can be used with vehicle kinematics (including measurements of a target vehicle) to identify additional events. Once identified kinematically, the events are reviewed through use of forward and face video. They are retained if verified as safety-related events and discarded if not. Within each event, factors that precipitated the event, that contributed to the event, and that were associated with the event are grouped into pre-event maneuvers, precipitating factors, contributing factors, associated factors, and avoidance maneuvers. The event begins at the onset of the precipitating factors and ends after the evasive maneuvers. Data for the period shortly before, during and shortly after the event are then preserved.
Table 3-1. Summary of kinematic search criteria for events in VTTI study (Dingus, et al., 2005).

<table>
<thead>
<tr>
<th>Trigger Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Lateral Acceleration</td>
<td>• Lateral accel. $\geq 0.7$ g.</td>
</tr>
<tr>
<td>2. Longitudinal Acceleration</td>
<td>• Accel. or decel. $\geq 0.6$ g.</td>
</tr>
<tr>
<td></td>
<td>• Accel. or decel. $\geq 0.5$ and forward TTC $\leq 4$ sec.</td>
</tr>
<tr>
<td></td>
<td>• $0.4 \leq$ longitudinal decel. $&lt; 0.5$, forward TTC $\leq 4$ sec.,</td>
</tr>
<tr>
<td></td>
<td>and forward range at the min. TTC $\leq 100$ ft.</td>
</tr>
<tr>
<td>3. Event Button</td>
<td>• Activated by the driver by pressing a button located on the dashboard when</td>
</tr>
<tr>
<td></td>
<td>an event occurred that he/she deemed critical.</td>
</tr>
<tr>
<td>4. Forward Time-to-Collision</td>
<td>• Accel. or decel. $\geq 0.5$ and TTC $\leq 4$ sec.</td>
</tr>
<tr>
<td></td>
<td>• $0.4 \leq$ longitudinal decel. $&lt; 0.5$, forward TTC $\leq 4$ sec.,</td>
</tr>
<tr>
<td></td>
<td>and forward range at the min. TTC $\leq 100$ ft.</td>
</tr>
<tr>
<td>5. Rear Time-to-Collision</td>
<td>• Rear TTC $\leq 2$ sec., rear range $\leq 50$ feet,</td>
</tr>
<tr>
<td></td>
<td>and absolute accel. of the following vehicle $&gt; 0.3$ g</td>
</tr>
<tr>
<td>6. Yaw rate</td>
<td>• Any value greater than or equal to a plus AND minus 4 degree change in</td>
</tr>
<tr>
<td></td>
<td>heading (i.e., vehicle must return to the same general direction of travel)</td>
</tr>
<tr>
<td></td>
<td>within a 3 second window of time.</td>
</tr>
</tbody>
</table>

In addition to the kinematic variables discussed above, there are three other sets of data routinely collected in naturalistic driving studies:

1. Context variables – these are descriptors of the physical features, such as road and environment, at the time of the event including geometric alignment and environmental factors (e.g. rain or snow; day or night). Some geometric features may be obtained by linking on-board GPS to existing geographic information systems (e.g. roadway inventory systems maintained by most state highway departments).

2. Event attributes - attributes of the event occurring immediately prior to and during event occurrence. Examples include the occurrence of driver distraction (sometimes identified by type of distraction) and presence of fatigue.

3. Driver attributes - typically obtained during subject intake to the study and may include age, stated prior driving record, propensity to take risks when driving and physiological conditions such as vision and reaction time.

While some event aspects remain unobserved (e.g. the actions of drivers in other vehicles and events beyond the range of cameras and sensors), it is an unquestioned advantage to observe the
actions of individual drivers, over long periods of time, including crash and near-crash events involvements. Although the result is a set of potentially very rich data that offers insight to crashes and near crash events that have been previously unavailable, a challenge remains in evaluating the near crashes and seeking a clearer relationship between them and crashes.

### 3.2 Study Goal

While naturalistic driving studies provide unique opportunities for safety analyses, the challenge of standardized measurement and observation remains. A standardized definition of a surrogate is a beginning, but more is needed. There is a need to develop a standard procedure to examine the validity of the events identified by using the definitions. This validation for naturalistic data has several steps:

1. The initial screening of possible events of interest, including crashes and near crashes.
2. An assessment of the events to classify them as to type; current classification of road crash types are a useful place to begin (e.g. road departure, rear end).
3. The events remaining after initial screening and classification need to be further analyzed so that the crashes and near crashes have a consistent etiology.

One can think of this goal by comparison with medical testing and diagnosis. Physicians and other medical professionals conduct standardized tests using accepted diagnostic procedures to identify the presence of disease in patients. In road safety analysis, particularly with near crashes, the challenge is to develop valid and consistent diagnostic procedures that can be used to assess safety problems for locations in the network or drivers in the population. The key is the standardization of diagnoses so that findings may be applied across studies through the accumulation of a firm knowledge base.
The goal of this research is the development of diagnostic procedures to identify and validate useful crash and near-crash events that can be used in enhanced safety analyses.

3.3 Methodology

Figure 3-1 is a conceptualization of the analysis of surrogates, crashes, and near crashes using naturalistic driving data. Normal driving (i.e., naturalistic) leads to a series of events that may be of interest for further study based upon pre-determined screening criteria; this is the First Screening. These criteria should be set to be inclusive of many possible events, with particular care in not excluding events that may be a crash or near crash. Candidate screening criteria include those listed in Table 3-1 and possibly others. This first screening is based on an analysis of computer-stored data (likely from the DAS and other information integrated into a data base). This first screening does not require analysis of video.

Figure 3-1. Conceptualization of the relationship between crashes and safety-relevant events.
This sets the stage for Classification, which has as an outcome the grouping of crash and near crash events with similar etiologies or generating characteristics. The classification criteria include kinematic or vehicle movement-related measures (e.g., lateral acceleration rate) and event attributes (e.g., intersection location or roadway curve location). After the Classification, the Second Screening further refines the set of events of interest. Once the Second Screening is complete, the Validation determines that the events of interest for a particular study have been properly identified and separated from those not of interest, because they fail the tests for a similar etiology or crash generating process. Notice that the notation is that the events selected from the Validation (i.e. the model outcome) are called “surrogate events,” even though they include both crashes and near crashes. This allows our notation to be consistent with that of Equation (4).

At the end of Validation stage, there are two conditional probabilities of interest: the probability of a crash outcome given either branch of the tree (either $Y_1 = 1$ or $Y_1 = 0$). These conditional probabilities are explored through an event-based model. Notice that the $\text{Prob}(Y_2 = 1|Y_1 = 1)$ represents the conditional probability of a crash given an event identified as a surrogate event. The conditional crash probability is interpreted as the proportion of surrogate events that resulted in crashes; this is, in fact the “$\pi$” of equation 1. A test of the event-based model is described in a companion paper (Wu and Jovanis, 2011). The lower branch ($Y_1 = 0$) represents events deemed not of interest; these may be re-examined to be sure there are no further events of interest ($Y_2 = 1$) although this is not conducted in this dissertation. This branch is intended to capture the analysis of events that lead to crashes but do not have large kinematic signatures; these events were observed in the VTTI data, so this outcome is specifically mentioned as an area in need of specific analysis.
3.3.1 The Analytic Procedure

Figure 3-2 is an overview of the proposed framework. Each step in the procedure is described in the following section. Statistical approaches are offered at each step but these are examples; other approaches are certainly possible. The idea is to undergo a sequence of statistical tests with the overall goal of identifying crashes and a set of similar near crashes for later analyses. Because the description of the framework is central to the chapter, we provide rather detailed descriptions of each step and the methods applied to our data set.
Figure 3-2. Conceptualization of the relationship between crashes and safety-relevant events.
3.3.2 First Screening

First Screening seeks to detect possible events of interest using information collected in the DAS. One way to think about the screening of crash and near crash events is in parallel with medical diagnosis. The result of a diagnostic test can be classified as a true positive (TP), a true negative (TN), a false positive (FP), or a false negative (FN). As the names suggest, a true positive result occurs when a diseased subject is correctly classified with a positive test; a true negative is a situation where the subject does not have the disease and the test says so. Both of these outcomes are desirable. A false negative result occurs when a diseased subject tests negative; similarly, a false positive occurs when a non-diseased subject has a positive result. At this stage we want to have true positives in diagnosing crash and near crash events and true negatives in identifying events that are not safety-related or not of interest. The test threshold determines the number of true positives, true negatives, false positives and false negatives.

3.3.2.1 Receiver Operating Characteristic (ROC) Curve

One way to examine tradeoffs with the 4 outcomes is with the Receiver Operating Characteristic (ROC) Curve, which can be conceptualized as determining the optimal diagnostic point (Peat and Barton, 2005). The ROC technique is commonly used in medical science to handle this problem (e.g. Swets, 1988; Centor, 1991; Obuchowski, 2003; Pepe, 2003). We first define a threshold $c$ for a marker $Z$ as positive if $Z > c$, or as negative if $Z < c$. A marker in the medical field indicates a diagnostic test score for a variable used to discriminate between a diseased and non-diseased subject. In our safety analysis, the marker is the variable used to identify the event of interest in First Screening. A marker could be a kinematic variable or a
combination of kinematic variables, context variable, and event attributes. Let the corresponding true and false positive rate at the threshold $c$ be $\text{TPR}(c)$ and $\text{FPR}(c)$, respectively.

$$\text{TPR}(c) = \text{True Positive Rate}(c) = P(Z \geq c|Y = 1)$$  

(2)

$$\text{FPR}(c) = \text{False Positive Rate}(c) = P(Z \geq c|Y = 0)$$  

(3)

As the threshold $c$ increases, both the false positive and true positive rate decreases.

Generally, the thresholds of the criteria should be set to include a high proportion of events of interest (i.e. high sensitivity). The desired goal is to achieve an acceptable sensitivity (correctly detect event of interest), say at least 90 percent, at the maximum specificity (minimum false alarm rate).

3.3.2.2 Receiver Operating Characteristic (ROC) Regression

Conveniently for safety studies, some medical researchers (Janes and Pepe, 2008) have found that some covariates, $M$, that are associated with disease can also impact the marker $Z$, and hence impact the inherent discriminatory accuracy of the marker (i.e. the ROC curve). For example, if male drivers tend to depress the brake pedal harder than female drivers (i.e., decelerate faster), then gender is associated with the marker deceleration. Therefore, the threshold of the marker may better discriminate events of interest for female drivers than for male drivers because female drivers will have severe decelerations less often. ROC regression methods can be used to test and handle this situation, where covariates affect the screening of events of interest (Pepe 2000; Alonzon and Pepe, 2002). Implementation proceeds in two steps: (1) model the distribution of the marker among controls as a function of covariates, and calculate the case
percentile values; and (2) model the cumulative density function of the ROC curve as a function of covariates. The ROC curves can therefore be modeled parametrically by using

\[ \text{ROC}_2(f) = \Phi\{\alpha_0 + \alpha_1 \Phi^{-1}(f) + \alpha_2 M\} \] (4)

where \( \Phi \) is the standard normal, \( f \) is a discrete set of FPR points, and \( \alpha_0, \alpha_1 \) and \( \alpha_2 \) are estimated parameters. If \( \alpha_2 \) is positive then an increase of \( M \) enhances the accuracy of the marker.

3.3.3 Classification

Once initial events are identified, there is a need to statistically distinguish different event types. Here we seek crashes with similar contributing factors and etiologies. A counterpart to the Chow test as suggested by Greene (2003) is proposed to undertake this step. The procedure tests whether the log-likelihood for a pooled-dataset model is significantly different from the sum of log-likelihoods for reduced dataset models. The result of the classification is the division of events into groups with similar etiologies; many different groups can be identified but it is expected that most studies, at least initially, will use two different crash types. There is a need to conduct a second, more refined, screening of the events to identify even more similar and consistent crash and near crash events by answering: What is a good marker? What is a good threshold?

3.3.4 Second Screening

To provide readers a better sense of the data at this step, vehicle lateral acceleration and yaw rate difference measured using a three second time window are presented in Figure 3-3. The lateral acceleration rate difference is the difference between the minimum and maximum lateral deceleration within the window (3 seconds in this case). Each individual trace is a separate event.
Figure 3-3. Illustrative Example for Data at Second Screening.

One can see that the vehicle kinematics for crash events (left side of Figure 3-3) tend to be more volatile than that for near crashes. Therefore, one can expect that a well-defined trigger should be able to identify the crashes. Notice in particular the figure in the lower right corner of Figure 3-3. The near crash events have vehicle traces that are higher than those for the crash events in the lower left corner. In concept, these are the events we are seeking to identify: events that are similar enough to crash events, but did not result in a crash outcome. Because the focus now is time-varying variables, and the crash risk over time during the events is also of interest, survival analysis is well-suited for detecting influential factors during the event. It is not only the duration of the event, per se, that is interesting, but also the likelihood that the event will end in "the next period" given that it has lasted as long as it has (Greene, 2003).

Different types of events would essentially be triggered by different vehicle movement-related variables and event attributes. As an example, lateral acceleration rate may play a more
important role in run-off-road than in rear-end events. The challenge in identifying an effective vehicle movement-related measure is that it is time-dependent and interacts with other event attributes during the event. The response variable can be translated into time-to-failure, where crash occurrence and the effects of time-varying covariates are of interest. Survival models have been used in several transportation studies (e.g. Jovanis and Chang, 1989; Hensher and Mannering, 1994) and they fit well in this analysis paradigm.

At this step, the original trigger criteria should be refined, since the initial criteria are simply like an entry threshold to sort events of interest. The refined thresholds should be determined differently for each type of event. The ROC curve can be applied to identify a threshold that has the best ability to correctly classify crashes and near crashes. Although there is no definitive formula for determining the most suitable cut-off point, the general guidance at this step is that one needs an ability to effectively filter out true negatives in order to "diagnose" similar surrogate events, though at the expense of not losing true crashes. However, those true crash events lost here can possibly indicate crash events that are not similar to the near crash events defined. Finally, one may have more than one surrogate measure with specific thresholds. With a large sample, a surrogate event can be identified based on more than one surrogate measure. The use of multiple screening criteria is suggested by the feedback loop in Figure 3-2. It is suggested that criteria be tested one at a time, with specific threshold levels and that the validity of the near-crash to crash relationship be tested. The feedback returning to the first screening may be used to change the kinematic trigger, the time window used to compute variable values or some combination. With our small sample, we provide only one pass through the data.
3.3.5 Validation

3.3.5.1 General Discussion of Validation

To validate whether an event of interest is a surrogate event, it is best to start with the definition of a surrogate event. Generally, a surrogate event represents a circumstance in which a driver needs to recover to normal driving by either adopting evasive maneuvers (Amundsen and Heden, 1977) or another appropriate response, otherwise a crash is likely (e.g. Shankar et al., 2008). Ideally, a set of conditions \( Y_1 = 1 \) that define a perfect surrogate event can be written as:

\[
Pr(Y_2 = 1|Y_1 = 1, X) \equiv 1
\]  

(5)

where crashes would definitely occur as the event satisfies the conditions of \( Y_1 \) in terms of event attributes and context variables. Moreover, Equation (5) implies that the association/correlation between \( Y_1 \) and \( Y_2 \) is a positive one.

\[
\text{Cov}(Y_1, Y_2|X_1, X_2) \equiv 1
\]  

(6)

where \( X_1 \) and \( X_2 \) represents factors that affect \( Y_1 \) and \( Y_2 \), respectively.

Equations (5) and (6) provide guidelines for defining a valid surrogate event. First, though it is not necessary to have every such event ending up with a crash, the conditional crash probability for a valid but weak surrogate event should still be significantly greater than zero, as shown in Equation (7).

\[
Pr(Y_2 = 1|Y_1 = 1, X) \gg 0
\]  

(7)

And there should be a significantly positive association/correlation between a crash and surrogate event, as shown in Equation (8).

\[
\text{Cov}(Y_1, Y_2|X_1, X_2) \gg 0
\]  

(8)

In this chapter, a bivariate Probit model is applied to test Equation (8) using the Tetrachoric correlation, a correlation measure for a pair of binary variables \( Y_1 \) and \( Y_2 \). To test
Equation (7), a Probit model is first applied to model the relationship \( \Pr(Y_2 = 1 | Y_1 = 1, X) \); endpoint transformation is then applied to construct confidence intervals for the conditional crash probability for each event (please refer to Xu and Long (2005) for more details). Only events satisfying Equations (7) and (8) will be referred to as surrogate events, and will be carried into the next step. It should be noted that a well-defined surrogate event should not only satisfy Equations (7) and (8), but also five general criteria: consistency with the basic definitions of a surrogate (Tarko, 2005; Sevensson, 1998); correlated with the clinical meaningful outcome (Tarko et al., 2009; Davis and Swenson, 2006; Davis et al., 2008; Shankar et al., 2008; Jovanis et al., 2010; McGehee et al., 2010; Guo et al., 2010; Hauer, 1999); have a statistical and causal relationship to crashes (Sevensson, 1998; Guo et al., 2010); fully capture the effect of the treatment in a way similar to how the treatment would affect crashes (Hauer, 1999; Shankar et al., 2008; Tarko et al., 2009); and, be useful as a "marker" indicating a time scale underpinning (Shankar et al., 2008; see Wu and Jovanis, 2011 for additional discussion).

### 3.3.5.2 Bivariate Probit Model

To test Equation (8), let whether an event will be deemed as a surrogate event \( (Y_1) \) and whether the surrogate event ends up in a crash \( (Y_2) \) be two latent processes; the Tetrachoric correlation is appropriate for analyzing multivariate relationships between the dichotomous variables. The Tetrachoric correlation for binary variables estimates the Pearson correlation of the latent continuous variables. Since the occurrence of surrogate events affects crash risk, a bivariate Probit model is suitable in terms of this situation (Greene, 2003). Formally, \( Y_1 = 1 \) indicates an event passing all specific conditions through first screening, classification, and second screening \( (Y_1=0, \text{ otherwise}) \), and \( Y_2=1 \) indicates crash occurrence \( (Y_2=0, \text{ near crash}) \). The surrogate event and crash generating processes can be written as:
\[ Y_1^* = X_1' \beta_1 + \epsilon_1, \ Y_1 = 1 \text{ if } Y_1^* > 0, 0 \text{ otherwise} \]  
(9)

\[ Y_2^* = X_2' \beta_1 + \epsilon_2, \ Y_2 = 1 \text{ if } Y_2^* > 0, 0 \text{ otherwise} \]  
(10)

\[ E(\epsilon_1|X_1, X_2) = E(\epsilon_2|X_1, X_2) = 0 \]  
(11)

\[ Var(\epsilon_1|X_1, X_2) = Var(\epsilon_2|X_1, X_2) = 1 \]  
(12)

\[ Cov(Y_1, Y_2|X_1, X_2) = Cov(\epsilon_1, \epsilon_2|X_1, X_2) = \rho \]  
(13)

And the bivariate normal cumulative density function is written as:

\[ Prob(X_1 < x_1, X_2 < x_2) = \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} \phi_2(z_1, z_2, \rho) \, dz_1 \, dz_2 \]  
(14)

\[ \phi_2(z_1, z_2, \rho) = \frac{e^{-\left(\frac{1}{2}\right) \left( z_1^2 + z_2^2 - 2\rho z_1 z_2 \right) / (1 - \rho^2) \right)}}{2\pi (1 - \rho^2)^{1/2}} \]  
(15)

3.3.6 Estimating the Conditional Crash Probability Using Valid Surrogate Events

At this step, one simply uses valid surrogate events to estimate the conditional crash probability \( \Pr(Y_2 = 1|Y_1 = 1, X) \) in terms of a variety of event scenarios. A generalized formulation to specify the conditional crash probability is developed in Wu and Jovanis (2011).

3.4 The Data

A subset of the Virginia Tech Transportation Institute (VTTI) 100-Car Naturalistic Driving Study dataset is applied to test the framework (Dingus et al., 2005). In the 100-car study 241 primary and secondary drivers drove for 12 to 13 months following the naturalistic driving protocols described in section 1. Based upon the event criteria in Table 3-1, VTTI researchers identified 69 crashes, 761 near crashes and 8295 critical events during the entire study. A focus
on road departure events led to a sample size of 21 single-vehicle-conflict crashes and 42 near crashes. Various aspects of the driving environment were recorded at the moment of the event, specifically at the onset of the precipitating factor, through the use of video and radar. Table 3-2 is a list of variable names, definitions, types, and data sources. All covariates available in the VTTI data set were tested in the analysis. The predictors shown in Table 3-2 are those which extensive modeling indicated were most consistently associated with event outcomes. Literally hundreds of models were explored to produce the reduced set of predictors in Table 3-2.

Table 3-2. Variable Definitions.

<table>
<thead>
<tr>
<th>Group</th>
<th>Variable</th>
<th>Measurement</th>
<th>Variable Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kinematic Variable</td>
<td>Vehicle lateral acceleration rate (Lat)</td>
<td>Measured every tenth of a second</td>
<td>Time Varying</td>
</tr>
<tr>
<td></td>
<td>• LATD</td>
<td>Maximum lateral acceleration rate difference within 3-second window</td>
<td>Time Varying</td>
</tr>
<tr>
<td></td>
<td>• LATM</td>
<td>Instantaneous maximum lateral acceleration rate within 3-second window</td>
<td>Time Varying</td>
</tr>
<tr>
<td></td>
<td>Vehicle longitudinal deceleration rate</td>
<td>Measured every tenth of a second</td>
<td>Time Varying</td>
</tr>
<tr>
<td></td>
<td>Vehicle yaw rate (Yaw)</td>
<td>Measured every tenth of a second</td>
<td>Time Varying</td>
</tr>
<tr>
<td></td>
<td>• YAWD</td>
<td>Maximum change of yaw rate within 3-second window</td>
<td>Time Varying</td>
</tr>
<tr>
<td></td>
<td>Vehicle speed</td>
<td>Measured every 3 to 10 tenth of a second</td>
<td>Time Varying</td>
</tr>
<tr>
<td>Event Attributes</td>
<td>Presence of driver fatigue</td>
<td>Fatigue (1); otherwise (0)</td>
<td>Time Independent</td>
</tr>
<tr>
<td>Context Variable</td>
<td>Event occurred on a horizontal curve</td>
<td>Curve (1); otherwise (0)</td>
<td>Time Independent</td>
</tr>
<tr>
<td></td>
<td>The presence of daylight</td>
<td>Daylight (1); otherwise (0)</td>
<td>Time Independent</td>
</tr>
</tbody>
</table>

3.5 Data Analysis

This section demonstrates how the whole procedure for screening, identifying, and validating surrogate events is implemented as shown in Figure 3-2. We conduct the analysis of the data as we would with an actual data set, but in this application, we can assess the accuracy of our framework because we have verified surrogate event etiologies as shown in the Appendix.
3.5.1 First Screening

Given the raw naturalistic driving data, the first task is to screen events of interest. We use all information in the 63 "trips" obtained from VTTI 100-car study to examine how the selection of first screening criteria would affect the accuracy of detecting events of interest. As shown in the left panel in Figure 3-4, since the data for each trip consists of 30 seconds before the event, during the event, and 10 seconds after the event, data from periods A and C are seen as events not of interest and data from period B for both crash and near crash events are considered as observations with events of interest. As long as the pre-specified first screening criteria can "hit" at least one of the observations in data chunk B, the event of interest would be detected. In other words, the threshold would be more effective if it could pick out the one extreme lateral acceleration in Figure 3-4, without detecting the “false alarm” shown in the right panel of the figure.

Maximum lateral acceleration difference greater than 0.4g (LATD), maximum lateral acceleration (LATM) and maximum change of yaw rate (YAWD) within a 3-second window were selected as the marker (the first screening criterion) for examining their accuracy for detecting the event of interest (period B in Figure 3-4). The application of the ROC analysis is summarized in Figure 3-5. The 45 degree line (the solid line) indicates the reference line; the greater the area between the ROC curve and the reference line, the better the accuracy of the marker. If the ROC area for a marker is not significantly greater than 0.5, then the discriminating ability for the marker is no better than random guess. It was found that lateral deceleration difference performs significantly better than maximum lateral acceleration. These test results suggest that the use of maximum difference within a time window can enhance the marker's accuracy. Meanwhile, lateral deceleration performs significantly better than yaw rate difference.
Figure 3-4. (Left) Event of Interest vs. (Right) Event Not of Interest: The Impact of The Selection of the Threshold.

- $H_0: \text{area}(\text{LATD}) = \text{area}(\text{LATM})$
  $\chi^2(1) = 5.25$, $p$-value $= 0.02$

- $H_0: \text{area}(\text{LATD}) = \text{area}(\text{YAWD})$
  $\chi^2(1) = 9.72$, $p$-value $= 0.002$

Figure 3-5. ROC Curves for LATD, LATM, and YAWD

Note that at this step, the goal is to detect as many true events of interest without including too many false alarms. As an example, one of the trigger criteria used by VTTI researchers is maximum lateral acceleration greater than or equal to $0.7g$; Table 3-3 indicates that this criteria can achieve 90 percent specificity (only 10 percent false alarms), but at the expense of only 27 percent sensitivity (only 27 percent true events of interest detected). VTTI did not lose the other 73 percent of events of interest; they used other trigger criteria (as shown in Table 3-1) to enhance the overall sensitivity. Similarly, if one uses lateral acceleration rate difference greater than $0.7g$, the sensitivity is almost doubled, though at the expense of ten percentage points less
specificity. This confirms that lateral acceleration rate difference can perform better than maximum lateral acceleration.

In this chapter, we will carry events with LATD greater than 0.4g during the entire event into the next step as a demonstration of this procedure. Using LATD greater than 0.4g, there are a total of 99 events detected from the 63 trips. The longest event lasted for 6.3 seconds, the shortest one lasted for 0.2 second, and the average event duration is 2.6 seconds. These 99 events will be carried to the classification stage to test the need of further classification.

Table 3-3. ROC Curves Analysis for LATM and LATD

<table>
<thead>
<tr>
<th>Cut-off point</th>
<th>LATM Sensitivity</th>
<th>LATM Specificity</th>
<th>LATD Sensitivity</th>
<th>LATD Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;= 0.0g</td>
<td>100.00%</td>
<td>0.00%</td>
<td>100.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>&gt;= 0.1g</td>
<td>100.00%</td>
<td>4.80%</td>
<td>100.00%</td>
<td>2.40%</td>
</tr>
<tr>
<td>&gt;= 0.2g</td>
<td>98.41%</td>
<td>22.40%</td>
<td>100.00%</td>
<td>13.60%</td>
</tr>
<tr>
<td>&gt;= 0.3g</td>
<td>92.06%</td>
<td>33.60%</td>
<td>96.83%</td>
<td>28.00%</td>
</tr>
<tr>
<td>&gt;= 0.4g</td>
<td>71.43%</td>
<td>60.80%</td>
<td>93.65%</td>
<td>41.60%</td>
</tr>
<tr>
<td>&gt;= 0.5g</td>
<td>46.03%</td>
<td>77.60%</td>
<td>84.13%</td>
<td>56.80%</td>
</tr>
<tr>
<td>&gt;= 0.6g</td>
<td>38.10%</td>
<td>84.00%</td>
<td>63.49%</td>
<td>72.80%</td>
</tr>
<tr>
<td>&gt;= 0.7g</td>
<td>26.98%</td>
<td>89.60%</td>
<td>49.21%</td>
<td>80.00%</td>
</tr>
<tr>
<td>&gt;= 0.8g</td>
<td>12.70%</td>
<td>95.20%</td>
<td>41.27%</td>
<td>83.20%</td>
</tr>
<tr>
<td>&gt;= 0.9g</td>
<td>7.94%</td>
<td>96.80%</td>
<td>36.51%</td>
<td>84.80%</td>
</tr>
<tr>
<td>&gt;= 1.0g</td>
<td>4.76%</td>
<td>98.40%</td>
<td>25.40%</td>
<td>90.40%</td>
</tr>
<tr>
<td>&gt;1.0g</td>
<td>0.00%</td>
<td>100.00%</td>
<td>0.00%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

3.5.2 Classification

One of the characteristics of valid surrogate events is the similarity among them, no matter whether they result in crashes or near crashes. This suggests that similar surrogate events should have similar crash probabilities in terms of context. Figure 3-6 shows that the crash probabilities for VTTI single-vehicle conflict events are quite different depending on whether the events occurred at intersections or non-intersections. This suggests that some of or all the regression coefficients will be different for the two subsets of the data.
In order to test whether the crash generating process for intersection-related and non-intersection-related events are the same, the equivalent of a Chow test for structural change was applied (see Figure 3-6). Due to the small sample size (63 total events), only lateral acceleration rate difference was included to model the crash probabilities. The log-likelihood for the pooled model is -43.97, as shown in Table 3-4. The log-likelihoods for the model based on single vehicle run-of-road and intersection related events are -7.15 and -33.07, respectively. The log-likelihood for the unrestricted model with separate coefficients is thus the sum, -40.22. The chi-square statistic for testing the two restrictions of the pooled model is twice the difference, \( LR = 2\times[-40.22-(-43.97)] = 7.5 \). The 95 percent critical value for the chi-square distribution with two degrees of freedom is 5.99 (the p-value of this chi-square test is 0.02). Therefore, at this significant level, the hypothesis that the constant term and LATD are the same for both types of event-based model is rejected. That is, there is significant structural change between the event-based models for intersection-related and non-intersection-related events, and hence the model including both types of events would be inconsistent. An interesting finding is also revealed in Table 3-4, which is that the same LATD would cause higher crash probability at intersections.
than non-intersections. As a result, 81 non-intersection-related single vehicle conflict events will be carried into second screening stage.

Table 3-4. Chow Test for Intersection v.s. Non-intersection Related Events.

<table>
<thead>
<tr>
<th>Dependent Variable: crash occurrence</th>
<th>Pooled Model</th>
<th>Non-Intersection</th>
<th>Intersection</th>
</tr>
</thead>
<tbody>
<tr>
<td>LATD</td>
<td>0.475</td>
<td>0.376</td>
<td>0.883</td>
</tr>
<tr>
<td>(S.E.)</td>
<td>(0.116)</td>
<td>(0.132)</td>
<td>(0.371)</td>
</tr>
<tr>
<td>p-value</td>
<td>0.000</td>
<td>0.004</td>
<td>0.017</td>
</tr>
<tr>
<td>Constant</td>
<td>-4.355</td>
<td>-4.002</td>
<td>-5.963</td>
</tr>
<tr>
<td>(S.E.)</td>
<td>(0.867)</td>
<td>(0.979)</td>
<td>(2.513)</td>
</tr>
<tr>
<td>p-value</td>
<td>0.000</td>
<td>0.000</td>
<td>0.018</td>
</tr>
<tr>
<td>Observations</td>
<td>99</td>
<td>81</td>
<td>18</td>
</tr>
<tr>
<td>Pseudo R-squared</td>
<td>0.181</td>
<td>0.113</td>
<td>0.427</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-43.97</td>
<td>-33.07</td>
<td>-7.150</td>
</tr>
</tbody>
</table>

3.5.3 Second Screening

At this step, to identify non-crash events that are similar to crashes, we are not only looking for a threshold that is best predicting crash occurrence, but also a marker that is influential to crash risk during the event. Even with a refined sample that has gone through the previous two steps, the results from the survival analysis still suggests that in terms of the magnitude of the estimated coefficient, LATD is more influential as a time-varying covariate than LATM and YAWD. The coefficient of 0.12 is interpreted to mean that those events with higher LATD once entering a situation where LATD greater than or equal to 0.4g have a higher risk of having a crash, as shown in Table 3-5. The greater ROC area for LATD than for YAWD also points to the same result (Ho: area (LATD) = area (YAWD), chi2 (1) = 12.06, p-value = 0.0005), as shown in Figure 3-7. Note that, if one solely relies on ROC techniques at this step, the time-varying effects of either kinematic variables, event attributes, or geometric alignment cannot be captured.
Table 3-5. Survival Analysis in Second Screening.

|       | Coef. | Std. Err. | z   | P>|z| | 95% CI  |
|-------|-------|-----------|-----|-----|---------|
| LATD  | 0.12  | 0.24      | 0.50| 0.62| -0.35   |
| LATM  | 0.03  | 0.24      | 0.12| 0.90| -0.44   |
| YAWD  | 0.07  | 0.33      | 0.21| 0.83| -0.58   |

Figure 3-7. ROC Curve for LATD and YAWD at Second Screening.

Table 3-7 summarizes threshold testing for LATD greater than 0.7g, which was selected at the second screening for two reasons: (1) with over 90 percent specificity, the cut-off point is high enough to leave almost no crash events; and (2) the marginal increase of specificities for LATD from greater than or equal to 0.7g to greater than or equal to 0.8g is only 2.5 percent, but the sensitivity decreased by 17 percent, suggesting LATD greater than or equal to 0.7g can both provide decent sample size and specificity.

Table 3-6. ROC Curves Analysis for LATD at Second Screening.

<table>
<thead>
<tr>
<th>Cut-off point</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Near Crashes</th>
<th>Crashes</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;= 0.4g</td>
<td>100.00%</td>
<td>0.00%</td>
<td>24</td>
<td>3</td>
</tr>
<tr>
<td>&gt;= 0.5g</td>
<td>82.61%</td>
<td>35.53%</td>
<td>21</td>
<td>2</td>
</tr>
<tr>
<td>&gt;= 0.6g</td>
<td>73.91%</td>
<td>67.11%</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>&gt;= 0.7g</td>
<td>73.91%</td>
<td>80.26%</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>&gt;= 0.8g</td>
<td>56.52%</td>
<td>82.89%</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>&gt;= 0.9g</td>
<td>56.52%</td>
<td>86.84%</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>&gt;= 1.0g</td>
<td>47.83%</td>
<td>93.42%</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>&gt;1.0g</td>
<td>0.00%</td>
<td>100.00%</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>
3.5.3 Validation

Use of the bivariate Probit model to test the tetrachoric correlation with LATD greater than or equal to 0.7g, indicates that the correlation is significantly greater than zero. In other words, Equation (8) holds (Likelihood-ratio test of $\rho = 0$: chi2 (1) = 7.55, p-value = 0.006). The results in Table 3-7 also indicate that higher speed will increase the probability of exceeding an LATD of 0.7g during an event. This is a useful connection between driver behavior (i.e. speed choice) and event outcome. Conversely, reducing speed during an event would reflect higher deceleration rate, which would reduce the probability of exceeding LATD greater than 0.7g during an event, and hence reduce the probability of crash occurrence. Daylight condition would reduce the probability of crash, though not significant.

Table 3-7. Bivariate Probit Model for Crash Occurrence and Events with LATD Greater Than or Equal to 0.7g.

<table>
<thead>
<tr>
<th>Equation</th>
<th>Coef.</th>
<th>Std. Err.</th>
<th>z</th>
<th>P&gt;z</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>LATD $\geq$ 0.7g</td>
<td>Deceleration Rate (g)</td>
<td>-1.09</td>
<td>0.81</td>
<td>-1.35</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>Speed (mph)</td>
<td>0.01</td>
<td>0.01</td>
<td>1.33</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>-1.10</td>
<td>0.36</td>
<td>-3.03</td>
<td>0.00</td>
</tr>
<tr>
<td>Crash Occurrence</td>
<td>Daytime Condition</td>
<td>-0.36</td>
<td>0.31</td>
<td>-1.15</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>-0.80</td>
<td>0.23</td>
<td>-3.53</td>
<td>0.00</td>
</tr>
<tr>
<td>$\rho$</td>
<td></td>
<td>0.65</td>
<td>0.18</td>
<td></td>
<td>0.15</td>
</tr>
</tbody>
</table>

This chapter considers the lower bound of the confidence intervals less than 0.1 as not satisfying Equation (7). It suggests that events with LATD greater than 0.7g but less than 0.8g, and events with LATD greater than 0.8g but less than 0.9g during daytime conditions are invalid surrogate events. Hence, in this chapter, the specific conditions for defining surrogate events are events that are: (1) with LATD greater than or equal to 0.4g during the entire event, (2) non-intersection related, and (3) with LATD greater or equal to 0.9g and events with LATD between 0.8g to 0.9g during nighttime conditions. A review of narratives accompanying the VTTI data
revealed that the identified grouping of crashes and near crashes appear to be qualitatively similar (see Appendix). The narratives indicate that a common kinematic maneuver by the driver is that the drivers undertook an abrupt evasive maneuver to avoid hitting a roadside object no matter whether the event ended up with a crash or near crash.

### 3.5.5 Conditional Crash Probabilities

The event-based model was constructed to model the conditional crash probabilities using valid surrogate events and instrumental variable Probit model to handle potential endogeneity (Cameron and Trivedi, 2005). The goodness of fit (Wald $\chi^2(2) = 5.01$; Prob $> \chi^2 = 0.08$) of the event-based model in Table 3-8 shows the appropriateness of this model specification. Although the suspected endogeneity is not statistically significant (Wald test of exogeneity: $\chi^2(1) = 1.28$, Prob $> \chi^2 = 0.26$), this model is still a more generalized form of the regular Probit model. Based on this event-based model, the conditional crash probabilities in terms of a variety of combinations of LATD and daytime condition are estimated and shown in Table 3-9.

**Table 3-8. The Event-based Model Using Valid Surrogate Events.**

| Coef. | Std. Err. | z   | P>|z| | 95% CI    |
|-------|-----------|-----|-----|---------|
| **Stage 2- Dependent Variable: Crash occurrence** |
| LATD  | 2.85      | 1.32| 2.15| 0.03    | 0.25     | 5.45    |
| Daytime Condition | -0.37 | 0.82 | -0.45 | 0.66    | -1.98    | 1.24    |
| Constant | -3.43 | 1.85 | -1.85 | 0.06    | -7.06    | 0.20    |
| **Stage 1- Dependent Variable: LATD** |
| Daytime Condition | 0.40 | 0.21 | 1.87 | 0.06    | -0.02    | 0.82    |
| Speed (mph)      | -0.01 | 0.01 | -1.03 | 0.30    | -0.03    | 0.01    |
| Deceleration Rate (g) | -0.63 | 0.22 | -2.79 | 0.01    | -1.07    | -0.19   |
| Constant         | 1.24 | 0.45 | 2.73 | 0.01    | 0.35     | 2.12    |
The predicted conditional crash probabilities are based on all predictors in Table 3-8, including LATD, daytime condition, vehicle average speed, and maximum deceleration rate during the event, hence the predicted probability for the same scenario will be somewhat different due to different vehicle average speed and maximum deceleration rate during the event. The lower and upper bound conditional crash probabilities for each scenario can be therefore constructed. As an example, for event scenario one, the average conditional crash probability is predicted as 0.08, meaning that for every 100 events satisfying this conditions, eight crashes are expected. The lower and upper bound conditional crash probabilities were constructed based the two surrogate events falling into this scenario. Given two such surrogate events observed, we expect to see 0.16 crashes, and there is actually no crash satisfying this condition observed. The ranges of the conditional crash probability for scenario four and five are large, partly because of small sample size and some crash events containing extreme vehicle kinematics. As comparing scenario two to three, and four to five, it was found that given the same LATD, events occurred during daytime condition have lower crash probability than during nighttime.

Table 3-9. Conditional Crash Probabilities Using Valid Surrogate Events.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Daytime/Nighttime</th>
<th>LATD</th>
<th>Average Conditional Crash Probability</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>Surrogate Event Observed</th>
<th>Crashes Expected</th>
<th>Crashes Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Night</td>
<td>&gt;= 0.8g</td>
<td>0.08</td>
<td>0.07</td>
<td>0.08</td>
<td>2</td>
<td>0.16</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>Night</td>
<td>&gt;= 0.9g</td>
<td>0.13</td>
<td>0.13</td>
<td>0.13</td>
<td>1</td>
<td>0.13</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>Day</td>
<td>&gt;= 0.9g</td>
<td>0.09</td>
<td>0.07</td>
<td>0.12</td>
<td>5</td>
<td>0.45</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>Night</td>
<td>&gt;= 1.0g</td>
<td>0.57</td>
<td>0.26</td>
<td>0.86</td>
<td>3</td>
<td>1.71</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>Day</td>
<td>&gt;= 1.0g</td>
<td>0.56</td>
<td>0.13</td>
<td>1.00</td>
<td>8</td>
<td>4.48</td>
<td>4</td>
</tr>
</tbody>
</table>
3.6 Conclusion and Discussion

3.6.1 Conclusion

Naturalistic driving studies provide an excellent opportunity to better understand crash causality and to supplement crash observations with a much larger number of near crash events. The goal of this research is the development of a rigorous set of diagnostic procedures to identify and validate useful crash and near crash events that can be used in enhanced safety analyses. As such, the research seeks to apply statistical methods as part of the methodology. A way to better understand crash occurrence and identify potential countermeasures to improve safety is to learn from and use near-crash events, particularly those near-crashes that have a common etiology to crash outcomes. This chapter demonstrates that a multi-stage modeling framework can make the analysis of naturalistic driving data tractable, without substantial use of video screening.

The dissertation begins by defining a crash surrogate as a crash or near crash event, consistent with the stream of work published since the 1980’s concerning the traffic conflicts technique. A standardized definition is a beginning, but more is needed to fully utilize the analysis potential of both naturalistic driving studies and near crash events. A standard procedure is developed that is applicable to virtually any naturalistic driving data that contains a stream of vehicle kinematic and context data. The procedure seeks to identify valid near crash events by:

1. The initial screening of possible events of interest, including crashes and near crashes. The input to this part of the process is expected to be an entire set of vehicle kinematic data for all vehicles in a study. Knowledge of crashes within the data steam is required, but this seems a reasonable expectation given past studies and the current experience in the U.S. Strategic Highway Research Program 2 (SHRP 2) Safety Program. As long as the kinematic signatures of the crashes are known, along with
GPS-based location information, the proposed procedure should be able to extract an initial set of candidate near crashes for subsequent processing. In this study, First Screening was conducted using Receiver Operating Characteristic (ROC) methods.

2. Once the initial screening is complete, the procedure calls for a classification of events to group those with similar etiology. The classification should be applicable to different road or driver crash types. The result of the classification is a reduced set of crashes and near crashes that are closer in etiology than those identified at the end of the First Screening.

3. The events remaining after initial screening and classification undergo a second screening and a series of tests to identify the most effective screening kinematic and context variables. A test is conducted of different triggering levels for individual variables or combinations of variables. The procedure ends with a validation test comparing crashes and near crashes in a common statistical model and then the estimation of crash and near crash probabilities and expected values so that modeled estimates can be compared to those from the data themselves. A number of statistical methods may be used in these steps including ROC regression and survival theory, especially due to the time-dependencies in the markers for events.

The framework is applied data from the VTTI 100-car study for road departure events. A total of 63 events are included in the study: 51 non-intersections and 12 intersection-related. While the sample sizes are limited in the empirical study, the authors believe the procedure is ready for testing in other applications. While the team has implemented specific statistical approaches in each of the steps of the procedure (See Figure 3-2), we believe the process is flexible enough to accommodate a range of methods.

With the appropriate caveats concerning the analysis of a single data set, the empirical findings include:
• Introducing the use of maximum difference within a time window on crash and near-crash markers offers advantages as to improvements in sensitivity (correctly detecting event of interest), with good specificity (minimum false alarm rate). In this case, the marker maximum difference in lateral acceleration (in a 3-second window) achieved the highest level of sensitivity and best specificity. This testing used the ROC method of analysis.

• Using ROC regression, the presence of driver fatigue was found to increase a marker's accuracy, but was found not to be statistically significant. Given the small sample sizes in this study, lack of statistical significance is not surprising; this example indicates that kinematic markers along with driver attributes may yield superior performance compared to kinematics alone.

• For single vehicle conflict events, there is a need to separate events occurring at intersections and non-intersections.

• In this study, the specific conditions for defining surrogate events are events that are: (1) detected using a maximum lateral acceleration difference of greater than or equal to 0.4g during the entire event duration; (2) non-intersection related; and, (3) have a maximum lateral acceleration difference of greater than or equal to 0.9g/event with a maximum lateral acceleration difference between 0.8g to 0.9g during nighttime conditions.

• For valid surrogate events, the same maximum lateral acceleration difference during daytime has a lower crash probability than during nighttime.
3.6.2 Discussion

One can think of the accomplishment of the study goal by comparison with medical testing and diagnosis. Physicians and other medical professionals conduct standardized tests using accepted diagnostic procedures to identify the presence of disease in patients. In road safety analysis, particularly with surrogates, the challenge is to develop valid consistent diagnostic procedures that can be used to assess safety problems for locations in the network or drivers in the population. The key is the standardization of diagnoses so that findings may be applied across studies through the accumulation of a firm knowledge base.

A recent study found that the contribution of treating near-crash events as crash observations can reduce the standard errors for the estimation of the effects of crash contributing factors because of the increase in the sample size (Guo et al., 2010). The review of the diverse traffic conflicts literature (e.g. Williams, 1981; Hauer, 1982; Grayson and Hakkert, 1987; Hauer, 1999) suggests there are additional potential benefits including:

- Given well-defined surrogate events (the output of the Validation step), it should be possible to use the models to assess what factors influence the conditional probability of a crash outcome and then, what countermeasure would be helpful in reducing crash probability. It was not possible to conduct this assessment due to limitations in sample size, but data from the SHRP Naturalistic Driving Field Study should provide ample data for such a test.
- Given the difference between crash and near-crash event outcomes, it would be interesting to conduct additional diagnosis of the factors that stop a surrogate event from becoming a crash given that both events share similar generating processes.
- Given the surrogate-to-crash evolution process, it would be useful to determine the triggering of near-crash events during normal driving. We can thus better understand
what we can do to reduce the probability of near-crash event occurrence, and hence crash occurrence.

It is hoped that this chapter has offered some useful suggestions on the use of crash and near-crash data from naturalistic driving studies that will be useful in improving our knowledge of road safety.

3.7 References


SHRP 2 website for naturalistic driving data: 
http://www.trb.org/StrategicHighwayResearchProgram2SHRP2/Public/Pages/RFP_S08_Resources_and_Reference_Material_487.aspx


Chapter 4

Exploring Relevant Issues when Screening and Analyzing Naturalistic Driving Data (Submitting to Accident Analysis & Prevention)

This study responds to the need for exploring the impact of the screening procedure on identifying and defining well-defined surrogate events, consisting of crashes and near crashes with common etiologies, using naturalistic driving data. The identification and definition of a surrogate event are associated with the selection of screening criteria and the designation of event duration. The exploration tests an algorithm developed in a previous paper (Wu and Jovanis, 2011). The algorithm first searches for events with common etiology from raw naturalistic driving data, and then identifies surrogate events among those events. This testing varies screening criteria systematically to explore the implications of criteria variation on event selection. The testing is conducted using data from the Virginia Tech 100-car study for road departure events. A total of 51 non-intersections and 12 intersection-related run-off-road events are included in an application of the testing. It was found that it is difficult to screen events of interest or identify surrogate events using a single measure or a universal threshold. A combination of a set of screening criteria would be beneficial for overall detecting ability and accuracy. Lastly, the performance of measures used in detecting events of interest and identifying surrogate events was found to vary in terms of context.

4.1 Introduction

The goal of crash surrogate studies is to identify well-defined surrogate events to enhance traffic safety analysis. Surrogate events are safety-relevant events that are strongly associated
with crash risk, and they consist of crashes and near crashes with common etiology. Although a wide range of different data can be used (e.g. street cameras), naturalistic driving studies are considered to be the most promising. Nevertheless, it is very costly to investigate every second of naturalistic driving data, since the size of the database is enormous. As an example, the Virginia Tech (VTTI) 100-car study collected approximately two million vehicle miles and 43,000 hours of driving data (Dingus et al., 2005). Therefore, at the very beginning, it is necessary to set up pre-screening criteria to extract safety-relevant events for further analysis. The challenge of the pre-set screening criteria is that the variation of screening criteria would essentially lead to different sets of safety-relevant events identified, and hence different sets of surrogate events. Therefore, this study seeks to explore relevant issues when screening and identifying surrogate events using naturalistic driving data.

This study is a continuation of previous work (Wu and Jovanis, 2011a; 2011b), which seeks to identify surrogate events in naturalistic driving data (Dingus et al., 2005). In Wu and Jovanis (2011a), an event-based approach was proposed to connect crash and surrogate events using the crash-to-surrogate ratio. In Wu and Jovanis (2011b), a multi-stage procedure was developed to screen and validate surrogate events. The multi-stage procedure includes:

1. First screening stage to detect safety-relevant events, referred to as event of interests, from raw naturalistic driving data,
2. Classification stage to separate events of interest with structural differences,
3. Second screening stage to search for events of interest that are strongly associated with crash risk, and
4. Validation stage to validate the relationship between surrogate events and crashes.

This testing selects different screening criteria systematically, and explores the implications of criteria variation on event selection.
To illustrate relevant issues, start with the following two cases extracted from raw naturalistic driving data, as shown in Figure 4-1. Note that a measurement duration of 100 represents 10 seconds. In Figure 4-1, one is a crash event (left panel), and the other is a near crash events (right panel), verified by VTTI research using video. Both cases consist of measurements before, during, and after the events. To simplify discussions, only the longitudinal acceleration, lateral acceleration, and vehicle speed over time are plotted. The left y-axis indicates the g-force of longitudinal and lateral acceleration, and the right y-axis represents vehicle's longitudinal speed in miles per hour (MPH). The review of video indicates that these two cases are:

Case 1: "Subject driver appears drowsy. He obtains some aerosol air freshener from his glove box, sprays it, and begins to put it back in the glove box when the vehicle runs off the road on the right, hitting the curb (Dingus et al., 2005)." The narrative describes what was happening in the left panel of Figure 4-1.

Case 2: " Subject driver is shifting gears and appears to be looking in the center mirror when the vehicle runs off the road on the right side (Dingus et al., 2005)." The narrative describes what was happening in the right panel of Figure 4-1.

In this example, there were three screening criteria considered at the first screening stage:

A. speed difference greater than 10 miles per hour in a short period;
B. lateral acceleration greater than 0.4g; and
C. longitudinal acceleration greater than 0.2g.

As one will see in the following, there are several relevant issues that need to be considered.
The first issue is how to assess the performance of the screening criteria to correctly detect events of interest, and how to enhance their performance. Although Wu and Jovanis (2011b) have proposed applying the Receiver Operating Characteristic (ROC) curve technique to quantify and compare the performance among screening criteria, there is a need to better understand how one screening criterion can enhance the other(s) on the overall accuracy. As an example, if one only has screening criterion A initially, and criteria B and C are suggested to enhance the overall accuracy of detecting true events of interest, the first challenge is how one can assess the incremental value of adding screening criterion B or C, so that one can better determine whether it is necessary to impose additional screening criteria.

Another challenge is that the screening criteria's capability to correctly detect true safety-relevant events would vary in terms of many factors. This phenomenon is also prevailing in medical science (e.g. Janes and Pepe, 2009). As an example, had the driver in case two tended to depress the brake pedal harder than the driver in case one (i.e., decelerate faster), then driver attributes are associated with deceleration force. Had we used screening criterion C to screen events of interest, it may better to discriminate surrogate events for the driver in case two than for the driver in case one from normative driving, because the driver in case one will have severe decelerations less often. In other words, the observation of more volatile longitudinal acceleration for case two than that for case one could be attributed to different driving behaviors across
drivers. The implication is that one may need to set up different thresholds in terms of different groups of drivers, or even roadway functional class and daytime/nighttime conditions. With a better understanding of the factors that affect screening criteria performance, insights can lead to modifications of criteria to improve their performance.

The second issue is that different screening criteria for detecting events of interest would lead to different sets of events being identified, and we are interested in how one set of events is different from the others, as all of them are derived from the same data. Based on the three criteria assumed above, there will be two events flagged in the left panel of Figure 4-1 and several events being flagged in the right panel. Combining the information from the video (such as the narratives), and kinematic variables, the question is: how does one determine whether these two events in the left panel and the several events in the right panel belong to the same event? If they belong to separate events, then the next question is: how does one determine where the starting point is and what the end point is for the events? One can now have a better sense that the way we screen events of interest and the way we define the span of the events would essentially lead to different sets of events.

The last issue is an outcome of decisions made in response to the second issue: how do different sets of events affect analysis, and how does one assess how one set of events is superior to others? If the selections of screening criteria would lead to different sets of surrogate events, one needs to check the effect on analysis results. As an example, what would be the effect of driver distraction on crash risk as comparing between two events or among several events pulled out from case one and two? More importantly, how does one define an event clearly and meaningfully while making it tractable? Therefore, if different screening criteria lead to different sets of events, then there is a need to discuss which set of surrogate events is better than the others, and what would be the implication for safety analysis.
This study seeks to explore the following issues: (1) assessing the performance of the screening criteria to detect events of interest and identify surrogate events more precisely; (2) explore the effects of changes in search criteria on the set of identified events; and (3) determining when a set of surrogate events is preferred to others.

4.1.1 Screening Procedure for Detecting Events of Interest and Designating Event Duration

In general, the screening strategies are mostly based on (1) sudden evasive maneuvers, reflecting on sudden changes of kinematic variables such as longitudinal or lateral acceleration (Battelle, 2006; Fitch et al., 2008), and (2) physics derivation, reflecting the on-going collision process, such as short time-to-collision (Dingus et al., 2005), or maximum additional time a following vehicle could have waited to brake before avoiding a rear-end crash, referred to as lagged time in Martin and Burgett (2001), Battelle (2006), and Fitch et al. (2008). It should be noted that neither Battelle (2006) nor Fitch et al. (2008) used lagged time to screen events of interest (they used it to determine event severity). Also note that the triggered events and conflict events in Dingus et al. (2005), Battelle (2006), and Fitch et al. (2008) are analogous to events of interest and surrogate events in Wu and Jovanis (2011a, 2011b). The events of interest identified through the detection of unusual vehicle kinematics recorded electronically are through accelerometers and gyroscopic sensors. Table 4-1 is an example of search criteria used to identify events for the VTTI 100-car study (Dingus et al., 2005). In Battelle (2006) and Fitch et al. (2008), events of interest are events in which a deceleration greater than 0.25g is required to avoid a collision with a lead vehicle within 1.5 seconds, and the trigger must persist for at least 0.7 seconds. Battelle (2006) and Fitch et al. (2008) then filter out non-threatening triggered events. The validation of surrogate events involves using video footage of the events to confirm the presence of a rear-end conflict. Unfortunately, the studies above provide little discussion of the
selection of the screening criteria, let alone the weakness and strengths of the screening criteria, i.e. the first two issues raised in this study.

Table 4-1. Summary of kinematic search criteria for events in VTTI study.

<table>
<thead>
<tr>
<th>Trigger Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Lateral Acceleration</td>
<td>• Lateral accel. $\geq 0.7$ g.</td>
</tr>
</tbody>
</table>
| 2. Longitudinal Acceleration  | • Accel. or decel. $\geq 0.6$ g.  
• Accel. or decel. $\geq 0.5$ and forward TTC $\leq 4$ sec.  
• $0.4g \leq$ longitudinal decel. $< 0.5g$, forward TTC $\leq 4$ sec., and forward range at the min. TTC $\leq 100$ ft. |
| 3. Event Button               | • Activated by the driver by pressing a button located on the dashboard when an event occurred that he/she deemed critical.               |
| 4. Forward Time-to-Collision  | • Accel. or decel. $\geq 0.5g$ and TTC $\leq 4$ sec.  
• $0.4g \leq$ longitudinal decel. $< 0.5g$, forward TTC $\leq 4$ sec., and forward range at the min. TTC $\leq 100$ ft. |
| 5. Rear Time-to-Collision     | • Rear TTC $\leq 2$ sec., rear range $\leq 50$ feet, and absolute accel. of the following vehicle $> 0.3g$                                  |
| 6. Yaw rate                   | • Any value greater than or equal to a plus AND minus 4 degree change in heading (i.e., vehicle must return to the same general direction of travel) within a 3 second window of time.

Once events of interest are detected, it is critical to determine the event duration for the events of interest. VTTI researchers determined event duration through the use of forward and face video. The event begins at the onset of the precipitating factors and ends after the evasive maneuvers. Although this approach may seem to be reasonable, it is very costly to go through every second of driving with respect to such an enormous data set for different research questions. Second, researchers solely rely on the use of forward and face video to determine whether the events are safety-related events, which is not only subjective, but also intractable and variable. Worse, if there were a need to change the guidelines for the researchers, it would be costly to redo the analysis.

This study does not disagree with VTTI's approach, but provides another alternative. Consider a general case where the duration of an event of interest begins at the onset of a measure exceeding a predetermined threshold and ends after another measure drops below another threshold. This measure is not necessary to be a single variable or constant over time. VTTI's
approach can be considered as a special case under this definition. The measures that VTTI researchers used at the onset are factors that are safety-related, referred to as precipitating factors, and the measure at the end is the completion of an evasive maneuver.

4.1.2 Assessment of A Refined Set of Surrogate Events

The development of a refined set of surrogate events has received only cursory treatment in past naturalistic driving data analyses, i.e. the data set was only screened once (e.g. VTTI 100-car study; UMTRI; Battelle, 2006; Fitch et al., 2008; Wu and Jovanis, 2011b). One question that can be posed is: how sure are we that the refined set of surrogate events would be the best data set for surrogate analysis? Before answering this question, one may need to go back one step to ask why and how to use surrogate events. The rationale for the use of surrogate events is that given well-defined surrogate events, it should be possible to assess expected number of crashes, what factors influence the conditional probability of a crash outcome, and what countermeasure would be helpful in reducing conditional crash probability (Hauer, 1982; Hydén, 1987; Chin and Quek, 1997; Archer, 2004; Tarko et al., 2009; McGehee et al., 2010 (Strategic Research Highway Program, SHRP 2); Fitch et al., 2008; Wu and Jovanis, 2011a; Wu and Jovanis, 2011b). Therefore, to assess how good a refined set of surrogate events are, it can be translated into (a) how similar the surrogate events are, and ultimately, (b) how good they are at predicting crash occurrence.

Assessing the similarity among surrogate events is challenging. Consider an event of interest with T time frames recorded and N variables, including kinematic variables, event attributes, driving environment, and driver attributes, where the basic set of variable dimension is T by N. As an example, for two events of interest with exact event duration of 30 seconds being recorded at every 0.1 second, then T is 300. Meanwhile, suppose that there are "only" 100
variables recorded (e.g. vehicle speed, lane position, etc.), N is 100. Hence, there are 300*100 = 30,000 dimensions for evaluating the similarity between these two events of interest. Even these 30,000 dimensions are similar for these two events of interest, the incorporation of changes of every single predictor over time would almost make the exactness between two events impossible. This is not saying the use of surrogate events to analyze naturalistic driving data is useless; conversely, there is a need to reduce the dimensions needed to identify the similarity. Nevertheless, the ultimate goal is to have well-defined surrogate events that can be used to precisely predict crash occurrence. In other words, although assessing similarity among surrogate events is difficult, the prediction accuracy for crash occurrence would be relatively easy to compare and has practical meaning.

There are a variety methods to assess prediction accuracy, but model goodness-of-fit and cross validation are the most common approaches (e.g. Hilbe, 2010). Comparing model goodness-of-fit tends to result in over-fitting a model that cannot be used for prediction of non-model data from a greater population, and cannot be compared across different samples (Hilbe, 2010). Many researchers urge the use of a validation data set to help confirm the use of the model to a greater population (e.g. Hilbe, 2010). This is particularly the case when the naturalistic driving data represents a sample of observation within a population. A validation data set is based on data that are not used in constructing the model. Specifically, a researcher can withhold a percentage of data from the multi-stage procedure. Once the selected data have been modeled, the withheld data - the validation data - are processed for prediction and comparison. Therefore, without predicting values of the response beyond the range specified by the fitted model, it is of great use if the model can be extrapolated to a greater population.

This paper builds on earlier research (Wu and Jovanis, 2011a; 2011b) and seeks to explore and discuss (1) how varying screening or trigger criteria would lead to different sets of surrogate events, and, (2) how to assess the appropriateness of a refined set of surrogate events
for surrogate analysis. The next section of this chapter describes the methodology used in this study, followed by a description of the particular data used in the analysis, the Virginia Tech 100 Car data set. A discussion of results follows with conclusions and suggestions for future research.

### 4.2 Methodology

Most of the analysis methods and tools have been described in Wu and Jovanis (2011a; 2011b). This section only appends additional methods used in this study, including the method to deal with the separation problem, a method to model true positive and false fractions, and a discussion of the incremental value of a test for screening.

#### 4.2.1 The Separation Problem

The separation problem occurs as the binary response variable is perfectly separated by a single risk factor or by a non-trivial linear combination of risk factors, which is analogous to the "sampling zero" problem in categorical data analysis (Albert and Anderson, 1984). The event-based approach extended by Wu and Jovanis (2011a) suggests that the binary outcome response variable model such as logistic/probit models can provide the flexibility to either test the need for classification or model the conditional crash probability. Nevertheless, often times, we see that some crash contributing factors only occur during a crash event state. Meanwhile, with the breaking down of events of interest by event attributes to sort out similar events, the small to medium-sized data set situations, i.e. no crash events in certain categories, may also often lead to such a problem. Computationally, it means that although the likelihood converges, there may be at least one parameter estimate that is infinite (e.g. Albert and Anderson, 1984; Santner and Duffy, 1986; Lesaffre and Albert, 1989; Hirji et al., 1989; Clarkson and Jennrich, 1991; Heinze
An infinite estimate can be regarded as extremely inaccurate (Lesaffre and Albert, 1989).

There are several options that can be considered to mitigate the problem. However, as discussed in Heinze and Schemper's (2002) paper, the following five options are unsatisfactory. The options are listed along with the reason in the parenthesis explaining why they are not satisfactory. They are: (1) omitting the risk factor from the model (loss of information); (2) change to different type of model (difficult to interpret and compare across models); (3) use of an ad hoc adjustment (difficult to interpret and compare across models); (4) exact logistic regression (difficult to deal with continuous risk factors); and (5) standard analysis with the estimate set to "high" value, i.e. set the value of parameter estimate of that iteration at which the log-likelihood changed by less than $10^{-6}$ (a high value of the estimate implies extreme inflation of its variance, leading to insignificant Wald test).

Heinze and Schemper (2002) proposed a method derived by Firth (1993) to overcome this problem, which is considered as a better alternative than the five options above (e.g. Hilbe, 2010). Heinze and Schemper penalize the logistic log-likelihood by half of the logarithm of the determinant of the information matrix. Mathematically, the logistic formulation of the fit ($Y=1$) is given as

$$\Pr(y_i = 1|x_i, \beta) = \pi_i = \frac{1 + \exp \left(-\sum_{r=1}^{k} x_{ir} \beta_r \right)}{1 + \exp \left(-\sum_{r=1}^{k} x_{ir} \beta_r \right)}$$

(5)

where $r$ is the risk factor or predictor, and $k$ is the total number of model predictors. The usual score equation

$$U(\beta_r) = \sum_{i=1}^{n} \left( y_i - \pi_i \right) x_{ir} = 0$$

(6)

is modified to appear as:
where \( h_i \)'s are the \( i \)th diagonal element of the "hat" matrix \( H = W^{\frac{1}{2}}X(X^TWX)^{-1}X^TW^{\frac{1}{2}} \), with \( W = \text{diag}\{\pi_i(1 - \pi_i)\} \). The estimates can then be obtained iteratively in the usual way until convergence is obtained. In short, Firth (1993) showed that the influence of the penalty function \( h_i(0.5 - \pi_i)x_{ir} \) is asymptotically negligible, and hence the estimate of \( \beta_r \) remains consistent. This method essentially splits each original observation \( i \) into two new observations having values \( y_i \) and \( 1 - y_i \) with iteratively updated weights \( (1 + 0.5h_i) \) and \( 0.5h_i \) respectively. The splitting of each original observation into a response and a non-response therefore guarantees finite estimates. This method thus completely eliminates the separation problem.

4.2.2 Generalized Linear Models for Modeling True and False Fractions

The performance of a screening criterion consists of true positive fraction (TPF), sensitivity, and false positive fraction (FPF), one minus specificity. To model the true and false positive fractions

\[
TPF(Z) = P(Y_1 = 1|Y_2 = 1, Z) \quad (8)
\]

\[
FPF(Z) = P(Y_1 = 1|Y_2 = 0, Z) \quad (9)
\]

Where \( Y_1 = 1 \) represents an event satisfying a pre-set screening criterion, \( Z \), whereas \( Y_2 = 1 \) represents the event ending in a crash. A generalized linear model for binary outcomes can be used, refer to Pepe (2003) for more details. Separate models for TPF and FPF can be employed as follows:
Popular choices for the link function are the logit link, \( g(t) = \log\left(\frac{t}{1 - t}\right) \), or the log link, \( g(t) = \log(t) \), which is easier for interpretation than the logit link.

### 4.2.3 The Incremental Value of a Test for Prediction

It is not viable to rely on only a single measure to detect either an event of interest or surrogate event, and hence it is necessary to understand how to complement the overall screening. Pepe (2003) provides a convenient framework for evaluating the incremental predictive value of a test beyond information contained in other sources. Suppose that the result of a simple test, denoted by \( Y_A \), is available and that another test \( Y_B \) can be performed in addition. One can quantify the additional information provided by the second test as \( P(Y_1 = 1|Y_A = 1, Y_B) \) compared with \( P(Y_1 = 1|Y_A = 1) \). One can also compare \( P(Y_1 = 1|Y_A = 0, Y_B) \) with \( P(Y_1 = 1|Y_A = 0) \). By comparing \( P(Y_1 = 1|Y_A = 1, Y_B) \) with \( P(Y_1 = 1|Y_A = 1) \), one can determine, among events testing positive with \( Y_A \), if \( Y_B \) provides additional predictive information.

### 4.3 The Data

The Virginia Tech Transportation Institute (VTTI) 100-Car Naturalistic Driving Study dataset is used for empirical testing (Dingus et al., 2005); it includes 241 primary and secondary drivers, and 12 to 13 months of data collection for each vehicle. A data acquisition system (DAS) consisting of cameras for video recording, kinematic sensors, radar, lane tracking devices, and a hard drive for data storage was installed in each vehicle.

Two unique features of naturalistic data are important for surrogate analysis:
1. Vehicles are instrumented with video camera technologies that observe the driver and the road ahead of the vehicle continuously during driving. In addition to the video, other on-board sensors continuously record vehicle accelerations in three dimensions and well as rotational motion along the same axes. Radars are often present to record proximity to other vehicles and potential obstacles on the roadway or roadside.

2. Drivers are asked to drive as they normally would (i.e. without specific experimental or operational protocols and not in a simulator or test track). The period of observation can vary from several weeks to a year or more.

All these data are recorded and stored within an on-board data acquisition system (i.e. DAS). The DAS for each vehicle is periodically copied into a searchable database and assembled for later analysis. Rather than relying on law enforcement officer judgment or witness recollection, the DAS can record virtually all the actions of the subject driver before, during and after each event. Because events are recorded using video and vehicle sensors, individual events of interest can generally be described with greater accuracy and reliability than using crash reports assembled after the fact (Dingus, et.al. 2005; Jovanis, et.al. 2011).

At certain points during the study, information from the DAS hard drive was received by VTTI, and triggering software was used to identify events of interest. Once the triggering events were found in the data, attributes of each event were saved from 30 seconds prior, to 10 seconds after the onset of the precipitating event. Based upon the event criteria in Table 4-1, VTTI researchers identified 69 crashes and 761 near crashes events during the entire study. These events of interest, as a whole, span the full range of crashes and related events including, e.g. intersection crashes and roadway departure events. To refine the scope of the investigation, we choose road departure events, the topic of several previous papers (e.g. Jovanis et al. 2011; Jovanis et al. 2010; Shankar, et.al. 2008). Because the focus of the study was road departure
events, the sample size was reduced to 13 crashes and 38 near crashes. This study uses the same data analyzed in Wu and Jovanis (2011b). Six measures were considered:

- Maximum lateral acceleration difference within 1-second window, Lat10D,
- Maximum lateral acceleration difference within 3-second window, Lat30D,
- Maximum instantaneous lateral acceleration, Lat01M,
- Maximum lateral acceleration within 1-second window, Lat10M,
- Maximum lateral acceleration within 3-second window, Lat30M, and
- Maximum yaw rate difference within 3-second window, Yaw30D.

4.4 Data Analysis

This multi-stage procedure seeks to identify well-defined surrogate events with similar etiologies, and hence the crash-to-surrogate ratio, referred to as conditional crash probability, can be modeled and studied. 63 single vehicle conflict events collected by VTTI 100-car naturalistic driving study were used to explore how the variation of screening criteria might affect the estimated conditional crash probability. In order to address the issues, the six measures described above were selected as the only measure used at the first screening stage. Hence these six samples will go through the multi-stage procedure proposed in Wu and Jovanis (2011b). All of the issues discussed in previous sections will be addressed along the way.

4.4.1 First Screening

The first issue, how do we assess the performance of the screening criteria, and how do we enhance their performance, will be addressed in this section. The goal of this stage is to detect as many real events of interest (acceptable sensitivity) without obtaining too many false alarms.
The ROC curve was applied to determine which measure is best in detecting events of interest, and to select a threshold for each measure. The greater the ROC area, the better the performance of a measure to correctly discriminate events of interest from the raw naturalistic driving data. As shown in Figure 4-2, in terms of the ROC area, Lat10D was found to be the best measure in detecting events of interest, followed by Lat01M, Lat10M, Lat30D, Lat30M, and then Yaw30D. Lat10D is significantly better than Lat01M, Lat01M is significantly better than Lat10M, Lat10M is significantly better than Lat30D, and Lat30D is significantly better than Lat30M. Nevertheless, Lat30M is not found to be significantly better than Yaw30D. As comparing Lat30D to Lat30M and Lat10D to Lat10M, it is clear that the concept of window can enhance the overall performance. Since Lat10D is better than Lat30D, Lat01M is better than Lat10M, and Lat10M is better than Lat30M, it seems that the shorter the window, the better the detecting ability. The results above is intuitive since all of which reflect drivers' abruptly evasive maneuvers. It is therefore recommended to use Lat10D to detect ROR events at the very beginning.

Figure 4-2. ROC Curves at First Screening.
A threshold for each measure is chosen in terms of sensitivity and specificity, as shown in Appendix B. For the Lat10D measure, a value that was greater than 0.4g was selected. For the Lat01M measure, a value that was greater than 0.3g was selected. For the Lat10M measure, a value that was greater than 0.3g was selected. For the Lat30D measure, a value that was greater than 0.4g was selected. For the Lat30M measure, a value that was greater than 0.3g was selected. Finally, for the Yaw30D measure, a value that was greater than 4 degrees/second was selected. Although the threshold for each screening criterion is determined based on sensitivity and specificity, the environment in which they are performed can influence the performance of correctly detecting a true event of interest.

It is well known that crash risk varies in terms of driver attributes, event attributes, and context. There is no secret that the performance of a screening criterion would also vary in terms of these factors. Therefore, it is important to identify and understand the influence of such factors to optimize conditions for using these screening criteria. The ROC regression is then applied to assess how the presence of driver fatigue would affect the overall ROC area. As shown in Table 4-2, it was found that the presence of driver fatigue could increase the performance of correctly discriminating events of interest from normal driving for Lat10D and Lat30M. Taking a closer look, it was found that the presence of driver fatigue enhanced the ability to correctly detect events of interest, i.e. increased TPF. As opposed to that, events that occurred in rural areas reduced the measure's TPF, as shown in Table 4-3. Although only several parameters estimated are marginally significant, a consistent sign and magnitude among all six samples suggests the plausibility that screening events of interest by sudden evasive maneuver may be useful in some occasions, but not in all situations. Table 4-4 also shows that the detection of events of interest from drivers who are older than 50, and are male drivers, are more likely to be correct compared to that from other drivers, i.e. less likely to be false alarms or decreased FPF. Although the
presence of driver fatigue shows a consistent coefficient sign increasing TPF, it is inconsistently affecting FPF, which is why it only significantly impacts the Lat10D and Lat30M sample.

Table 4-2. Factors affecting ROC Performance.

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>Lat10D</th>
<th>Lat30D</th>
<th>Lat01M</th>
<th>Lat10M</th>
<th>Lat30M</th>
<th>Yaw30D</th>
</tr>
</thead>
<tbody>
<tr>
<td>alpha_0</td>
<td>2.098</td>
<td>1.109***</td>
<td>1.751***</td>
<td>1.385***</td>
<td>0.719***</td>
<td>0.867***</td>
</tr>
<tr>
<td>(S.E.)</td>
<td>(1.363)</td>
<td>(0.284)</td>
<td>(0.334)</td>
<td>(0.263)</td>
<td>(0.221)</td>
<td>(0.233)</td>
</tr>
<tr>
<td>alpha_1</td>
<td>0.893</td>
<td>1.435***</td>
<td>1.317***</td>
<td>1.197***</td>
<td>1.315***</td>
<td>1.477***</td>
</tr>
<tr>
<td>(S.E.)</td>
<td>(0.967)</td>
<td>(0.242)</td>
<td>(0.244)</td>
<td>(0.204)</td>
<td>(0.180)</td>
<td>(0.182)</td>
</tr>
<tr>
<td>Presence of Fatigue</td>
<td>5.226**</td>
<td>0.314</td>
<td>1.850</td>
<td>1.341</td>
<td>0.693***</td>
<td>-0.104</td>
</tr>
<tr>
<td>(S.E.)</td>
<td>(2.162)</td>
<td>(0.282)</td>
<td>(1.957)</td>
<td>(0.873)</td>
<td>(0.261)</td>
<td>(0.282)</td>
</tr>
<tr>
<td>Observations</td>
<td>188</td>
<td>188</td>
<td>188</td>
<td>188</td>
<td>188</td>
<td>188</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 4-3. Factors affecting TPF and FPF at first screening.

**Factors affecting TPF at first screening**

<table>
<thead>
<tr>
<th></th>
<th>Lat10D04</th>
<th>Lat30D04</th>
<th>Lat01M03</th>
<th>Lat10M03</th>
<th>Lat30M03</th>
<th>Yaw30D4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatigue</td>
<td>1.840</td>
<td>1.599</td>
<td>2.018</td>
<td>2.018</td>
<td>1.822</td>
<td>-0.651</td>
</tr>
<tr>
<td>S.E.</td>
<td>1.501</td>
<td>1.509</td>
<td>1.489</td>
<td>1.489</td>
<td>1.497</td>
<td>1.988</td>
</tr>
<tr>
<td>P-value</td>
<td>0.220</td>
<td>0.290</td>
<td>0.175</td>
<td>0.175</td>
<td>0.223</td>
<td>0.743</td>
</tr>
<tr>
<td>Rural</td>
<td>-0.707</td>
<td>-0.307</td>
<td>-0.325</td>
<td>-0.325</td>
<td>0.0697</td>
<td>-0.270</td>
</tr>
<tr>
<td>S.E.</td>
<td>0.886</td>
<td>0.945</td>
<td>0.822</td>
<td>0.822</td>
<td>0.885</td>
<td>1.988</td>
</tr>
<tr>
<td>P-value</td>
<td>0.428</td>
<td>0.745</td>
<td>0.693</td>
<td>0.693</td>
<td>0.937</td>
<td>0.892</td>
</tr>
<tr>
<td>Constant Term</td>
<td>2.200</td>
<td>2.205</td>
<td>1.819</td>
<td>1.819</td>
<td>1.823</td>
<td>4.156</td>
</tr>
<tr>
<td>S.E.</td>
<td>0.665</td>
<td>0.667</td>
<td>0.576</td>
<td>0.576</td>
<td>0.577</td>
<td>1.617</td>
</tr>
<tr>
<td>P-value</td>
<td>0.009946</td>
<td>0.009943</td>
<td>0.00158</td>
<td>0.00158</td>
<td>0.00159</td>
<td>0.0102</td>
</tr>
<tr>
<td>Observations</td>
<td>63</td>
<td>63</td>
<td>63</td>
<td>63</td>
<td>63</td>
<td>63</td>
</tr>
</tbody>
</table>

**Factors affecting FPF at first screening**

<table>
<thead>
<tr>
<th></th>
<th>Lat10D04</th>
<th>Lat30D04</th>
<th>Lat01M03</th>
<th>Lat10M03</th>
<th>Lat30M03</th>
<th>Yaw30D4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Driver older than 50</td>
<td>-0.466</td>
<td>-0.0385</td>
<td>-0.295</td>
<td>-0.348</td>
<td>-0.409</td>
<td>-0.339</td>
</tr>
<tr>
<td>S.E.</td>
<td>0.750</td>
<td>0.435</td>
<td>0.492</td>
<td>0.449</td>
<td>0.447</td>
<td>1.067</td>
</tr>
<tr>
<td>P-value</td>
<td>0.534</td>
<td>0.929</td>
<td>0.549</td>
<td>0.438</td>
<td>0.360</td>
<td>0.751</td>
</tr>
<tr>
<td>Driver with Age 30-50</td>
<td>-0.0692</td>
<td>0.460</td>
<td>0.411</td>
<td>-0.508</td>
<td>0.555</td>
<td>-0.275</td>
</tr>
<tr>
<td>S.E.</td>
<td>0.969</td>
<td>0.702</td>
<td>0.679</td>
<td>0.703</td>
<td>0.777</td>
<td>1.639</td>
</tr>
<tr>
<td>P-value</td>
<td>0.943</td>
<td>0.513</td>
<td>0.545</td>
<td>0.470</td>
<td>0.475</td>
<td>0.867</td>
</tr>
<tr>
<td>Male Driver</td>
<td>-0.336</td>
<td>-0.00351</td>
<td>-0.116</td>
<td>-0.126</td>
<td>0.253</td>
<td>-0.0358</td>
</tr>
<tr>
<td>S.E.</td>
<td>6.018</td>
<td>0.397</td>
<td>0.428</td>
<td>0.400</td>
<td>0.417</td>
<td>1.076</td>
</tr>
<tr>
<td>P-value</td>
<td>0.580</td>
<td>0.993</td>
<td>0.786</td>
<td>0.752</td>
<td>0.544</td>
<td>0.973</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.699</td>
<td>0.299</td>
<td>-0.753</td>
<td>-0.224</td>
<td>0.615</td>
<td>3.258</td>
</tr>
<tr>
<td>S.E.</td>
<td>0.463</td>
<td>0.326</td>
<td>0.348</td>
<td>0.327</td>
<td>0.339</td>
<td>0.878</td>
</tr>
<tr>
<td>P-value</td>
<td>0.000242</td>
<td>0.360</td>
<td>0.0305</td>
<td>0.492</td>
<td>0.0695</td>
<td>0.000207</td>
</tr>
<tr>
<td>Observations</td>
<td>113</td>
<td>113</td>
<td>113</td>
<td>113</td>
<td>113</td>
<td>113</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-35.80</td>
<td>-72.68</td>
<td>-64.19</td>
<td>-71.25</td>
<td>-67.01</td>
<td>-13.99</td>
</tr>
</tbody>
</table>
Although ROC curve and regression has been shown to be useful to quantify the performance of measures in separating events of interest from normal driving, another enhancement to the performance can be made through combining multiple screening measures. To demonstrate, suppose that the only screening measure in this study is Yaw30D greater than 4 degrees/sec (Yaw30D04). As shown in Table 4-4, the predictive value of Yaw30D04 alone is 35% (exp(-1.072)). If events that are detected by Yaw30D04 are also tested with Lat30D04 (Lat30D greater than 0.4g), a positive result increases the probability that the event is a true event of interest by a multiplicative factor of 1.18 to 1.43 (exp(0.173) and exp(0.36), respectively). A negative Lat30D04 test decreases the probability by a multiplicative factor of 10% to 56%. Thus the Lat30D04 provides statistically significant predictive information to the Yaw30D04.

Table 4-4. The incremental value of Lat30D on Yaw30D.

|       | Coef. | Std. Err. | z   | P>|z|   | 95% CI     |
|-------|-------|-----------|-----|------|-----------|
| Lat30D04=0 | -1.493 | 0.462     | -3.230 | 0.001 | -2.398  | -0.588    |
| Lat30D04=1  | 0.267  | 0.048     | 5.600  | 0.000 | 0.173   | 0.360     |
| _cons     | -1.072 | 0.102     | -10.460| 0.000 | -1.273  | -0.871    |

The left panel in Figure 4-3 shows that a different screening criterion would essentially lead to different sets of events. To demonstrate how varying the definition of event duration affects analysis results, we consider the simplest case where the duration of an event of interest begins at the onset of a certain measure exceeding a predetermined threshold and ends after the same measure falls below the same threshold. As an example, as shown in Figure 4-3, two events of interest were detected when applying a threshold of Lat10D greater than 0.4g during the event. One of them lasted for 0.3 seconds, which is in the "30-second before period"; the other lasted for about 1 second around the "event end." The reference line indicates the time point at which the event ended according to the identification by VTTI researchers. Similarly, three events of interest were detected when applying a threshold of Lat10M greater than 0.3g, and so on so forth.
The advantages of this definition are that (1) the comparison across events is standardized, and (2) these events are kinematically similar. 76 events of interest satisfying \( \text{Lat}10\text{D} > 0.4\text{g} \) threshold during an event were carried to the second stage. 203 events of interest satisfying \( \text{Lat}01\text{M} > 0.3\text{g} \) threshold during an event were carried to the second stage. 101 events of interest satisfying \( \text{Lat}10\text{M} > 0.3\text{g} \) threshold during an event were carried to the second stage. 99 events of interest satisfying \( \text{Lat}30\text{D} > 0.4\text{g} \) threshold during an event were carried to the second stage. 90 events of interest satisfying \( \text{Lat}30\text{M} > 0.3\text{g} \) threshold during an event were carried to the second stage. 188 events of interest satisfying \( \text{Yaw}30\text{D} > 4 \text{degree/second} \) threshold during an event were carried to the classification and second screening stage.

Figure 4-3. Illustration of Detected Event of Interest After First Screening for Each Measure.
4.4.2 Classification, Second Screening, and Validation

The second issue, how different screening criteria for detecting events of interest would lead to different sets of events being identified, will be addressed in this section. The classification, second screening, and validation stages seek to identify surrogate events with common etiology. Therefore, the specific conditions for identifying surrogate events are events that satisfy first screening criteria, pass classification, meet second screening criteria and validation procedure, as shown in Table 4-5, which summarizes the results from all tests and models at each stage for each sample, and will be discussed at some length.

At the classification stage, all six samples were tested using Chow tests. The results indicated the need to break down events of interest by intersection and non-intersection for all samples. After removing intersection-related events from events screened from the previous stage, the number of events left can be seen in Table 4-5, which also summarizes the number of tests and models, and will be discussed at some length.

At second screening, the screening criteria are refined and hence will be different from the criteria at first screening. As an example, the ROC curve and survival analysis suggested the use of maximum lateral acceleration difference within 3-second window (Lat30D) as the influential time-dependent measure at second screening stage for the Lat10D sample, and Lat30D greater than 0.7g was selected as the threshold in terms of sensitivity and specificity. Likewise, for all other samples, the corresponding second criteria are shown in Table 4-5.

The analysis for factors that affect TPF and FPF are similar to the discussion in section 4.1, and the results are shown in Appendix C. For TPF, it was found that driver distraction and the presence of a median increases TPF, though not statistically significant. As opposed to this finding, these two factors were also found to be increasing FPF at the same time. And hence, it was not able to find any factor that could improve the overall ROC area at second screening, not
shown here. It should be noted that the separation problem is common due to small sample size at this stage, so the TPF and FPF modeling are based on firth logit.

At the validation stage, as suggested in Wu and Jovanis (2011b), the validation involves the use of the bivariate Probit model to test the tetrachoric correlation, and the use of the end point transformation method to further exclude events in which the lower bound of the confidence intervals less than 0.1. The tetrachoric correlation is significantly positive for each sample, indicating significant association between crash outcome and second screening criteria. The Lat01M sample was excluded entirely, since there were no potential surrogate events in this sample satisfying the criteria above. After going through this step, a set of specific conditions were determined for the rest of each sample. Events of interest that satisfy the specific conditions are referred to as valid surrogate events, see Table 4-5 for more details about the specific conditions for each sample.

The right panel in Figure 4-3 illustrates the exclusion of events of interest from the first screening stage though validation stage. For the Lat10D sample, there is only one event left after the validation stage. All the events in the Lat01M sample are gone. For the Lat10M sample, there is only one event left, and so on so forth. All the events left from each sample in the right panel are surrogate events and all are centered to the time point at which the event ends which was determined by VTTI researchers, the red reference line.
Table 4-5. Summary of Outputs from All Stages.

<table>
<thead>
<tr>
<th>Stage</th>
<th>Description</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Lat10D</td>
</tr>
<tr>
<td>First Screening</td>
<td>ROC area</td>
<td>0.948</td>
</tr>
<tr>
<td></td>
<td>Criteria</td>
<td>Lat10D &gt; 0.4g during entire event</td>
</tr>
<tr>
<td></td>
<td># of events of interest</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>Min/Max duration (sec.)</td>
<td>0.1/3.1</td>
</tr>
<tr>
<td>Classification (intersection vs. non-intersection)</td>
<td>Chow test</td>
<td>Separate</td>
</tr>
<tr>
<td></td>
<td>Chow test (Prob &gt; chi2)</td>
<td>.008</td>
</tr>
<tr>
<td></td>
<td># of eligible events</td>
<td>60</td>
</tr>
<tr>
<td>Second Screening</td>
<td>Criteria</td>
<td>Max Lat30D &gt; 0.7g</td>
</tr>
<tr>
<td></td>
<td>Rho (p-value for LR test)</td>
<td>0.480 (0.0271)</td>
</tr>
<tr>
<td>Validation</td>
<td>Further Exclusion</td>
<td>• Lat30D &lt; 0.8g &amp; Daylight=1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All events were excluded.</td>
</tr>
<tr>
<td></td>
<td>Specific conditions</td>
<td>• Lat10D &gt; 0.4g during event</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Non-intersection related</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Lat30D &gt; 0.7g &amp; Daylight=0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Lat30D &gt; 0.9g &amp; Daylight=1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Yaw30D &gt; 4°/sec during event</td>
</tr>
<tr>
<td></td>
<td># of surrogate events</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td># of crash events</td>
<td>9</td>
</tr>
</tbody>
</table>
4.4.3 Conditional Crash Probability - Event-based model

The final issues, the effects of different sets of events on the analysis results, and assessing how a set of events is superior to others, are discussed in this section. As discussed in section 1.2, to assess how good a refined set of surrogate events are, it can be translated into (a) how similar the surrogate events are, and ultimately, (b) how good they are at predicting crash occurrence.

To see how similar the validated surrogate events are, Figure 4-4 and Figure 4-5 show the kinematic trajectory for both crash and near-crash events screened through the Lat10D and Yaw30D samples. Each individual trace is a separate event, and only one kinematic variable is shown for simple comparison, which is lateral acceleration difference in a 3-second window. One can see that these traces have a similar shape, but crash events tend to have a higher magnitude. For the Lat10D sample, most of the traces are like an inverse "z." For the Yaw30D sample, most of the traces are either like n or m, but note that the event duration for crashes are shorter than that for near crashes. The similarity in shape is intuitive. The multi-stage procedure sets different criteria at different stages, and therefore it is simply like a program that slices similar pieces from each original trace.
To see how different sets of surrogate events would affect analysis result, a separate event-based model was built for each sample using the same predictors, as shown in Appendix E.
The event-based model was constructed by instrumental variable probit model using valid surrogate events. Visually the results are shown in Figure 4-6, and the six event scenarios are:

- Event Scenario 1: Lat30D greater than 0.7g during nighttime condition.
- Event Scenario 2: Lat30D greater than 0.8g during nighttime condition.
- Event Scenario 3: Lat30D greater than 0.9g during nighttime condition.
- Event Scenario 4: Lat30D greater than 0.9g during daytime condition.
- Event Scenario 5: Lat30D greater than 1.0g during nighttime condition.
- Event Scenario 6: Lat30D greater than 1.0g during daytime condition.

Though the number of valid surrogate events and crash events are different for different samples, the conditional crash probabilities in terms of event scenarios are similar and consistent. The average conditional crash probabilities for event scenario one range from 0.13 for the Lat10 M sample to 0.35 for the Lat30M sample. Note that this difference is incomparable, since the specific conditions for each sample are different, as shown in Table 4-5. The effects of crash contributing factors on the conditional crash probability are similar in some cases, indicated by the slopes between two connected dots. As an example, when comparing scenarios one and two, the line slopes for all samples are similar, suggesting the effect of the increase in maximum lateral acceleration from 0.7g to 0.8g on the conditional crash probability are similar for all samples. Note that the effect of every 0.1g increase in maximum lateral acceleration on the conditional crash probability is non-linear, which is why the same observation does not apply when comparing event scenarios three and five. Nevertheless, the confidence intervals for the slopes are overlapped, suggesting insignificant difference. Overall, the event-based model for each sample is similar to the others.
Due to the limitations of data (only 63 short trips are available in this study, and they have been screened by VTTI researchers), it is not surprising to see the event-based models in terms of different samples are still similar. On the one hand, these 63 events screened by VTTI researchers reflect similar kinematic maneuvers, which is that the drivers undertook an abrupt evasive maneuver to avoid hitting a roadside object no matter whether the event ended up with a crash or near crash. On the other hand, the similarity may be mainly attributed to the small sample size.

The other way to assess how good a refined set of surrogate events are is to assess how good they are at predicting crash occurrences. Figure 4-7 shows the observed and expected number of crashes, the product of predicted conditional crash probability and observed surrogate event counts. Visually, Lat10M sample seems to have the best fit. Although chi-square can be computed to see how the predicted values fit the observe values, the chi-square cannot be
compared across different samples because of different set of observations and specific conditions.

Figure 4-7. Expected vs. observed number of crashes.

To determine the best set of specific conditions for a single vehicle ROR event, a cross-validation was explored to decide which event-based model in term of different sets of specific conditions can best predict the number of crashes. The procedure begins with splitting raw naturalistic data into two groups. One is used to implement the whole procedure above, and the other is used to compare the predictability. To demonstrate, this study randomly split 63 events into two groups, and compared Lat10D and Yaw30 sample. Figure 4-8 shows that the other half of events is better predicted by the model built by Lat10D sample than Yaw30D from the first half’s data. It was also found that the value of summation of (observed - expected)^2/expected for Lat10 is 0.92 with 5 groups, and 2.5 with 6 groups for Yaw30D sample. Although these two numbers are not statistically comparable, it is suggested that the set of specific conditions of
Lat10D sample can produce more valid surrogate events than Yaw30D sample, as shown in Figure 4-8.

![Graph showing cross-validation: Lat10D vs. Yaw30D.](image)

Figure 4-8. Cross-validation: Lat10D vs. Yaw30D.

### 4.5 Conclusion and Discussion

This study explored and discussed relevant issues when screening and analyzing naturalistic driving data. Among the findings are:

- It is recommended to use Lat10D to detect single vehicle conflict events in the very beginning.

- In terms of sensitivity and specificity, Lat10D greater than 0.4g, Lat01M greater than 0.3g, Lat10M greater than 0.3g, Lat30D greater than 0.4g, Lat30M greater than 0.3g, Yaw30D greater than 4 degree/second were found to be suitable as first screening criteria.
Although the threshold for each screening criterion is determined based on sensitivity and specificity, the environment in which they are performed can influence their performance of correctly detecting true events of interest. It was found that the presence of driver fatigue could increase the performance for Lat10D and Lat30M.

Lat30D04 provides statistically significant predictive information to the Yaw30D04. This finding implies that multiple metric surrogate measures are better than single metrics to detect events of interest.

Different screening criteria would essentially lead to different sets of events.

The ROC curve and survival analysis suggested the use of Lat30D as the influential time-dependent measure at the second screening stage for Lat10D (Lat30D > 0.7g), Lat30D (Lat30D > 0.7g), Lat01M (Lat30D > 0.6g), and Lat10M (Lat30D > 0.7g) sample, and the use of Lat10D for Lat30M (Lat10D > 0.7g) and Yaw30D (Lat30D > 0.6g) sample.

The Lat10D sample can produce more valid surrogate events than the Yaw30D sample.

Although this study screened events of interest kinematically, it should be noted that this is not the only way to define an event of interest. The first screening criteria can consist of event attributes, kinematic variables, time-dependent variables, context, driver attributes, and a variety of first screening criteria can be tried.

The selections of screening or trigger criteria may lead to different sets of surrogate events be identified. It seems to be difficult to screen events of interest using a single measure or a universal threshold. In other words, the performance of detecting events of interest would vary in terms of context, and a test may have some incremental value for the overall detecting ability as combining with other tests. This finding implies that multiple metric surrogate measures are better than single metric to detect event of interest.
Although this study has demonstrated a cross-validation scheme for selecting the best set of surrogate events, it should be noted that even this examination may not be sufficient. This study suggests to further compare the crashes predicted from the procedure described above with the number observed on the roads. Please refer to Section 5.2.2 for more details.

Finally, this study strongly suggests that all issues raised should be discussed in the future naturalistic driving data analysis.

4.6 References


Chapter 5

Conclusions and Recommendations for Future Research

5.1 Conclusion

Although abundant resources have been devoted to for prevention and severity reduction of traffic crashes, a method for better understanding crash occurrence is to utilize surrogate events. The most valuable application of crash surrogates is not only to predict future crashes but also to identify inappropriate driving behaviors, roadway design and operational deficiencies, or to find some other countermeasures that may help reduce crash risk (Hauer, 1999; Grayson and Hakkert, 1987). The practical use of surrogate events is also to learn what makes the difference between crash and surrogate events, which are promising for determining factors that can break the surrogate-to-crash process, i.e. stop a surrogate event from becoming a crash. Other benefits of using surrogate events within traffic safety research include:

- Reducing the time needed to develop a sufficient sample size for analysis. Using surrogate events should allow the analyst to estimate safety and study crash contributing factors by observing a large number of events in a short period of time. This responds to a common practical complaint of needing to wait too long for sufficient crash events to occur before undertaking action.

- Developing models with greater prediction precision. Using well-defined surrogate events, it should be possible to use models to assess what factors influence the conditional probability of a crash outcome and then, what countermeasure would be helpful in reducing crash probability. These models would have greater precision in
their estimate of crash contributing factors if crash events were supplemented by carefully screened surrogate events, creating a larger validated sample of events.

- Understanding the factors that contributing to crashes. Given the difference between crash and near-crash event outcomes, it would be interesting to conduct additional diagnoses of the factors that stop a surrogate event from becoming a crash given that both events share similar generating processes. Similarly, it would be useful to study the occurrence of crash surrogates during normal driving to better understand how surrogates are triggered. We can thus improve countermeasures to reduce the probability of crash surrogate occurrence, and hence crash occurrence.

Although surrogate events could be measured through a variety of methods, they are best studied in a naturalistic driving study. The objective of chapter two is to formulate and test an analytical paradigm for surrogate analysis using naturalistic driving data. The relationship between surrogate events and crashes is first established through a set of desirable criteria that integrates desirable attributes from a diverse literature for defining surrogate events, and a definitional link between the expected number of crashes, the number of observed surrogate events, and the crash-to-surrogate ratio. The general criteria for defining a surrogate event are summarized from the literature, and are discussed in Section 2.1.2. This chapter also provides a flexible specification for estimating the crash-to-surrogate conversion factor using a wide range of variables. Although the link is definitional, it requires surrogate events to be well-defined, i.e. they should have common etiologies. Therefore, a multi-stage algorithm is developed and tested to identify well-defined surrogate events, as discussed in Chapters three and four.

The multi-stage procedure starts with raw naturalistic driving data (including vehicle kinematic, video, and location information) and screening of the events identified as potential surrogate events. The challenges of analyzing surrogate events using naturalistic driving data is to
screen, identify, and validate surrogate events. As argued by Williams (1980), the absence of standard techniques for defining surrogate events has produced a series of research results which are difficult to compare, and hence the available results of research cannot be cumulated to improve our knowledge. This problem may be overcome by adopting standard definitions for surrogate events. Chapters three and four propose a sequence of statistical tests with the overall goal of validating surrogate events. A systematic procedure has been developed to identify, screen, and validate crash surrogates that have a common etiology with crash outcomes (Chapter three), which forms the core of the surrogate analysis in this dissertation. The procedure has been iteratively tested with a small data set from the 100-car study (Chapter four). The outcomes of this procedure are sets of specific conditions that identify well-defined surrogate events.

This dissertation has also introduced several useful concepts and methods to effectively implement the whole procedure. Both ROC curve and regression techniques are useful in assessing the performance of markers and their corresponding thresholds in differentiating events. Sensitivity and specificity are shown to be useful measures for quantifying the evaluation. ROC regression can enhance researchers' understanding in factors that would affect screening criteria's capability in correctly discriminating events. Nevertheless, the cautionary note regarding the use of all screening criteria is that: different screening criterion would essentially lead to different sets of events, and hence it is important to iteratively test how the variation would affect analysis results. Other findings for enhancing markers' performance include:

- Introducing the use of maximum difference within a time window on kinematic measures that are used to screen events offers advantages as to improvements in sensitivity and specificity.
- Multiple metric surrogate measures are better than single metrics for detecting an event of interest.
Although the results from this dissertation should be considered exploratory due to limited sample size, the analysis framework and some findings are ready to be carried out with a larger naturalistic driving data set. Among the findings are:

- It is recommended to use Lat10D to detect single vehicle run-off-road surrogate events at the very beginning.
- The use of driver fatigue as a predictor was found to increase a marker's accuracy in detecting single vehicle run-off-road surrogate events.
- For single vehicle run-off-road events, there is a need to separate events occurring at intersections and non-intersections.
- Although somewhat strict, the specific conditions for defining single vehicle run-off-road surrogate events are events that are: (1) detected using a maximum lateral acceleration difference within a second window greater than or equal to 0.4g during entire event duration; (2) non-intersection related; and, (3) have a maximum lateral acceleration difference of greater than or equal to 0.9g/events with a maximum lateral acceleration difference between 0.8g to 0.9g during nighttime conditions (see Table 4-5).

5.2 Recommendation for Future Research

While the results are promising, there are limitations to the dissertation: (1) The impacts of the selection of screening criteria and designation of event duration on analysis should be examined with a larger data set; (2) Much more can be done to refine the identification and definition of surrogate events in naturalistic driving data; and (3) Further validation using historical crash data is needed. Since all the issues above rely on the use of SHRP2 naturalistic
driving data in concert with Highway Safety Information System, section 5.2.2 describes a research plan regarding how to utilize the two data sets.

5.2.1 Refine the Identification and Definition of Surrogate Events

While the dissertation develops and tests a modeling framework for estimation of $\pi$ by detecting sudden evasive maneuvers, much more can be done to refine the identification and definition of surrogate events in naturalistic driving data. The screening strategies are mostly based on (1) sudden evasive maneuvers, and (2) physics derivation. Sudden evasive maneuver reflect sudden changes of kinematic variables such as longitudinal or lateral acceleration. One should recognize that there may be events in which no kinematic trigger is apparent (see Hydén (1987) for an early discussion of this issue). Some distraction or fatigue-related events are examples of such events. Hence, solely relying on sudden evasive maneuvers to detect surrogate events is insufficient. Physics derivation reflects on-going collision process, such as short time-to-collision, or maximum additional time a following vehicle could have waited to brake before avoiding a rear-end crash, referred to as lagged time in Martin and Burgett (2001). Since modeling the on-going collision process requires a wide range of variables along with relevant parameters, numerous assumptions are often imposed, resulting in inaccurate models. Therefore, at this moment, neither of these analyses should be considered as a replacement for video review until the definition of surrogate events are fully examined and validated. Caution should be also paid on the use of instrumented videos to define the events, which is not only subjective, but also intractable and incomparable, so that the knowledge cannot be accumulated. A feedback mechanism which allows researchers to build up on previous findings is therefore desired (e.g. Chapter three).
Although event screening with raw vehicle kinematic data is expected to be labor intensive and time consuming, tailored screening for different research questions may be necessary. Since different screening criterion selection and event duration designation would essentially lead to different sets of events, as demonstrated in Chapter four, one should keep track of how one set of events is different from the others, as all of them are derived from the same data. There has been limited attention paid to statistical modeling as part of event screening and also as part of the analysis of kinematics to assess potentially safety countermeasure effectiveness. The advantage of statistical modeling is that it provides for the ability to repeat the experiment in different settings so better controls are applied to surrogate screening and crash kinematics. The statistical modeling represents opportunities to be more systematic in naturalistic driving data analyses, providing opportunities for more scientific, repeatable experiments.

5.2.2 Further Cross-Validation Using Historical Crashes

The proposed framework (Figure 5-1) extends research already completed in the joint use of crash events and surrogate events in a structural framework. The proposed research uses a flexible framework that can be adapted to the evolving needs of highway safety research advancement and to the investigation of data from on-going naturalistic driving studies. Data analyses are proposed using Highway Safety Information System (HSIS) data from North Carolina and Washington (and SHRP 2 S-04 if available), along with naturalistic driving data from SHRP 2 Naturalistic Driving Study data collection sites in Durham and Seattle. The methods described may also be applied to other naturalistic driving data and crash records.

The idea is to undergo a sequence of statistical tests with the overall goal of validating surrogate events and facilitating their use in enhanced safety analyses. The process is grounded in the fundamental relationship of Equation (1) in Chapter one, as shown in the first box in Figure
5-1(Chapter two). The multi-stage procedure depicted in the second box is a series of statistical tests that start with raw naturalistic driving data (including vehicle kinematic, video, and location information) and screen the events identified as potential surrogate events. The outcomes of this procedure are sets of specific conditions that identify surrogate events with common etiology.

In the second box, the proposed second validation compares the crashes predicted from the procedure described above with the number of crashes observed in HSIS at segments/intersections in Durham and Seattle. We are thus checking which conversion table can best predict the number of crashes on the roads. It should be noted that although some variables/attributes are not observed in HSIS, which is the fundamental deficiency in HSIS, that should not hinder the utility of this procedure. Once the surrogate events are validated for a
context, one can use the valid surrogate events for advanced traffic safety analysis as shown in
the third box, which will be discussed in the next section.

Despite the variation of the analysis process, it usually involves three recursive steps,
described in Figure 5-1. The recommended data analysis process is recursive, meaning that
returning to a specific step several times may be necessary. Specifically, to utilize SHRP2
naturalistic driving data in concert with HSIS data, the analysis plan is suggested as the
following:

**Step 1:**
1. Split the whole SHRP2 data into two portions.
2. Use VTTI, UMTRI or a portion of SHRP2 data to construct initial screening criteria
   for different types of events.
3. Review the video for those events detected using the initial screening criteria.
4. Fine tune the initial screening criteria to incorporate many other variables.
5. Conduct classification, second screening, validation, and estimate a conversion table
   using each subsample.
6. Cross-validate using the other portion of data (first validation).

**Step 2:**
1. Count the number of surrogate events occurred on the road, and turn into population
   basis.
2. Validate candidate conversion tables using HSIS data (second validation).

**Step 3:**
1. Use validated conversion table to identify SWiPs.
2. Use validated conversion table to evaluate effectiveness of countermeasures.
3. Study driving performance using the validated conditions for defining surrogate
   events.
4. If there is any new findings or new methods to identify surrogate events, repeat step one and two.

5.2.3 The Use of Surrogate Events and Learning from Surrogate Events

Once the surrogate events are validated for a context, one can use the valid surrogate events for advanced traffic safety analysis. One application is to use surrogate event counts to predict the expected number of crashes and identify sites with promise for improvement (SWiPs) by knowing the number of surrogate events at each segment/intersection. Another application is to evaluate countermeasure effectiveness where we can compare the π across different event scenarios. Lastly, the specific conditions for defining surrogate events can then be further tested and studied through driving performance assessment (e.g. the specific conditions can then be used to defined "risky driving" in driving simulators), which is promising to improve our knowledge about inappropriate driving behaviors, driver responses, roadway design and operational deficiencies. This dissertation suggests that there are additional potential opportunities for future research, including:

- Given well-defined surrogate events (the output of the validation step), it should be possible to use the models to assess what factors influence the conditional probability of a crash outcome and then, what countermeasure would be helpful in reducing crash probability. It was not possible to conduct this assessment due to limitations in sample size, but data from the SHRP Naturalistic Driving Field Study should provide ample data for such a test.
- Given the difference between crash and near-crash event outcomes, it would be interesting to conduct additional diagnosis of the factors that stop a surrogate event from becoming a crash given that both events share similar generating processes (this
is similar in concept to some of the work conducted by Gary Davis of University of Minnesota for SHRP 2 and others).

- Given the surrogate-to-crash evolution process, it would be useful to determine the triggering of near-crash events during normal driving. We can thus better understand what we can do to reduce the probability of near-crash event occurrence, and hence crash occurrence.
Appendix A Event Narratives

- **Surrogate Event 1**: crash event. Subject driver loses control of vehicle in the snow. The vehicle spins 180 degrees counter clockwise while moving longitudinally and laterally and the passenger side of the vehicle hits a snow bank off the opposite side of the road.
- **Surrogate Event 2**: crash event. Subject driver is asleep and hits the median.
- **Surrogate Event 3**: near crash. Subject driver is going too fast, and nearly hits the median on the other side.
- **Surrogate Event 4**: crash. Subject driver is adjusting the radio while driving. At the last minute he moves into a left turn lane. The left turn lane is separated from his initial lane by a median. When he moves into the turn lane he hits the median.
- **Surrogate Event 5**: near crash. Subject driver has just inserted a cd into cd player and is closing the cd case as he veers off the road to the right.
- **Surrogate Event 6**: crash. Subject driver is singing and appears to be driving too fast on wet roads while making a right turn. She loses control of the vehicle and ends up on the median to the left of the road she turns on to.
- **Surrogate Event 7**: near crash. Subject driver looks out his left window (no other traffic is present and he appears to be looking at the scenery). The road curves and the vehicle runs off the right side of the road.
- **Surrogate Event 8**: near crash. Subject driver is dialing phone and crossing over double yellow line. Then, the road curves and the vehicle runs off the road on the right.
- **Surrogate Event 9**: near crash. Subject driver is looking at a piece of paper as he drives under an overpass. The road curves to the left and the vehicle veers left and nearly hits the left median.
- **Surrogate Event 10**: near crash. Subject driver appears fatigued and is negotiating a curve to the right while on an exit/entrance ramp. He is driving too fast and goes off the road on the left side.
- **Surrogate Event 11**: crash. Subject driver hits patch of ice on the roadway and vehicle slides over double lane line on the left. Subject driver overcorrects and vehicle swerves across right lane and off onto right shoulder hitting the guardrail.
- **Surrogate Event 12**: crash. Subject driver appears drowsy and possibly under the influence of drugs or alcohol. He falls asleep and the vehicle drifts off the right side of the road, nearly hitting a parked vehicle. Then the vehicle goes up onto the curb. The vehicle travels on the curb hitting a mailbox before returning to the roadway and nearly hitting another parked vehicle.
- **Surrogate Event 13**: near crash. Subject driver is driving in the left lane. The median to his left has snow plowed against it in places. It appears that the vehicle hits either the median on the left or snow covering it. Video data is missing for the forward view. Changed the "Traffic Flow" variable after reviewing on satellite map.
- **Surrogate Event 14**: near crash. Subject vehicle is traveling on snowy roadway and appears to get tire caught in snow on right side of roadway which causes him to hit the right median.
- **Surrogate Event 15**: near crash. Subject driver is drowsy and falling asleep while driving. The vehicle runs off the road on the right.
- **Surrogate Event 16**: crash. Subject driver appears drowsy. He obtains some aerosol air freshener from his glove box, sprays it, and begins to put it back in the glove box when the vehicle runs off the road on the right, hitting the curb.
• Surrogate Event 17: near crash. Subject driver appears drowsy and is looking at a book he has placed near the steering wheel while driving in a construction zone. The road curves to the right and the vehicle runs off the right side of the path created by the construction barrels.

• Surrogate Event 18: near crash. Subject driver falls asleep while driving and the vehicle runs off the road on the right.

• Surrogate Event 19: near crash. Subject driver appears drowsy and the vehicle runs off the road on the right side and almost hits a telephone pole.
# Appendix B

## Sensitivity and Specificity Each Sample (Unit: Percentage)

<table>
<thead>
<tr>
<th>Cut-off point</th>
<th>Lat10D Sensitivity</th>
<th>Lat01M Specificity</th>
<th>Lat10M Sensitivity</th>
<th>Lat30D Specificity</th>
<th>Lat30M Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\geq 0.0g)</td>
<td>100.00</td>
<td>0.00</td>
<td>100.00</td>
<td>0.00</td>
<td>100.00</td>
</tr>
<tr>
<td>(\geq 0.1g)</td>
<td>100.00</td>
<td>8.00</td>
<td>100.00</td>
<td>18.40</td>
<td>100.00</td>
</tr>
<tr>
<td>(\geq 0.2g)</td>
<td>100.00</td>
<td>44.80</td>
<td>96.83</td>
<td>55.20</td>
<td>96.83</td>
</tr>
<tr>
<td>(\geq 0.3g)</td>
<td>95.24</td>
<td>75.20</td>
<td>90.48</td>
<td>69.60</td>
<td>90.48</td>
</tr>
<tr>
<td>(\geq 0.4g)</td>
<td>92.06</td>
<td>89.60</td>
<td>68.25</td>
<td>91.20</td>
<td>84.80</td>
</tr>
<tr>
<td>(\geq 0.5g)</td>
<td>55.56</td>
<td>97.60</td>
<td>38.10</td>
<td>99.20</td>
<td>97.60</td>
</tr>
<tr>
<td>(\geq 0.7g)</td>
<td>44.44</td>
<td>98.40</td>
<td>26.98</td>
<td>100.00</td>
<td>99.20</td>
</tr>
<tr>
<td>(\geq 0.9g)</td>
<td>33.33</td>
<td>99.20</td>
<td>4.76</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>(\geq 1.0g)</td>
<td>23.81</td>
<td>99.20</td>
<td>0.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>&gt;1.0g</td>
<td>0.00</td>
<td>100.00</td>
<td>0.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>
Appendix C Factors affecting TPF and FPF at Second screening

Factors affecting TPF at second screening.

<table>
<thead>
<tr>
<th></th>
<th>Lat10D</th>
<th>Lat30D</th>
<th>Lat01M</th>
<th>Lat10M</th>
<th>Lat30M</th>
<th>Yaw30D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lat30D07</td>
<td>Lat30D07</td>
<td>Lat30D07</td>
<td>Lat30D06</td>
<td>Lat10D07</td>
<td>Lat10D06</td>
</tr>
<tr>
<td>Distraction</td>
<td>0.754</td>
<td>0.754</td>
<td>0.885</td>
<td>1.263</td>
<td>0.754</td>
<td>0.754</td>
</tr>
<tr>
<td>S.E.</td>
<td>1.068</td>
<td>1.068</td>
<td>1.145</td>
<td>1.196</td>
<td>1.068</td>
<td>1.068</td>
</tr>
<tr>
<td>P-value</td>
<td>0.480</td>
<td>0.480</td>
<td>0.440</td>
<td>0.291</td>
<td>0.480</td>
<td>0.480</td>
</tr>
<tr>
<td>Presence of Median</td>
<td>2.507</td>
<td>2.507</td>
<td>2.769</td>
<td>2.406</td>
<td>2.507</td>
<td>2.507</td>
</tr>
<tr>
<td>S.E.</td>
<td>1.587</td>
<td>1.587</td>
<td>1.644</td>
<td>1.637</td>
<td>1.587</td>
<td>1.587</td>
</tr>
<tr>
<td>P-value</td>
<td>0.114</td>
<td>0.114</td>
<td>0.0920</td>
<td>0.142</td>
<td>0.114</td>
<td>0.114</td>
</tr>
<tr>
<td>Constant</td>
<td>0.0196</td>
<td>0.0196</td>
<td>-0.315</td>
<td>0.0186</td>
<td>0.0196</td>
<td>0.0196</td>
</tr>
<tr>
<td>S.E.</td>
<td>0.757</td>
<td>0.757</td>
<td>0.830</td>
<td>0.760</td>
<td>0.757</td>
<td>0.757</td>
</tr>
<tr>
<td>P-value</td>
<td>0.979</td>
<td>0.979</td>
<td>0.704</td>
<td>0.981</td>
<td>0.979</td>
<td>0.979</td>
</tr>
<tr>
<td>Observations</td>
<td>21</td>
<td>21</td>
<td>18</td>
<td>19</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-8,799</td>
<td>-8,799</td>
<td>-7,799</td>
<td>-7,570</td>
<td>-8,799</td>
<td>-8,799</td>
</tr>
</tbody>
</table>

Factors affecting FPF at second screening.

<table>
<thead>
<tr>
<th></th>
<th>Lat10D</th>
<th>Lat30D</th>
<th>Lat01M</th>
<th>Lat10M</th>
<th>Lat30M</th>
<th>Yaw30D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lat30D07</td>
<td>Lat30D07</td>
<td>Lat30D07</td>
<td>Lat30D06</td>
<td>Lat10D07</td>
<td>Lat10D06</td>
</tr>
<tr>
<td>Distraction</td>
<td>-0.885</td>
<td>-0.349</td>
<td>-1.138</td>
<td>-0.540</td>
<td>-0.225</td>
<td>-0.0958</td>
</tr>
<tr>
<td>S.E.</td>
<td>0.587</td>
<td>0.583</td>
<td>0.365</td>
<td>0.517</td>
<td>0.667</td>
<td>0.543</td>
</tr>
<tr>
<td>P-value</td>
<td>0.131</td>
<td>0.549</td>
<td>0.00184</td>
<td>0.297</td>
<td>0.735</td>
<td>0.860</td>
</tr>
<tr>
<td>Presence of Median</td>
<td>0.389</td>
<td>0.676</td>
<td>0.573</td>
<td>0.166</td>
<td>0.772</td>
<td>0.772</td>
</tr>
<tr>
<td>S.E.</td>
<td>0.571</td>
<td>0.565</td>
<td>0.362</td>
<td>0.472</td>
<td>0.644</td>
<td>0.492</td>
</tr>
<tr>
<td>P-value</td>
<td>0.496</td>
<td>0.231</td>
<td>0.113</td>
<td>0.726</td>
<td>0.231</td>
<td>0.116</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.237</td>
<td>-1.412</td>
<td>-0.787</td>
<td>-0.381</td>
<td>-1.740</td>
<td>-2.325</td>
</tr>
<tr>
<td>S.E.</td>
<td>0.574</td>
<td>0.603</td>
<td>0.323</td>
<td>0.532</td>
<td>0.717</td>
<td>0.572</td>
</tr>
<tr>
<td>P-value</td>
<td>0.680</td>
<td>0.0191</td>
<td>0.0148</td>
<td>0.473</td>
<td>0.0152</td>
<td>4.82e-05</td>
</tr>
<tr>
<td>Observations</td>
<td>55</td>
<td>78</td>
<td>185</td>
<td>82</td>
<td>69</td>
<td>167</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-31.55</td>
<td>-36.17</td>
<td>-88.01</td>
<td>-48.46</td>
<td>-29.01</td>
<td>-55.12</td>
</tr>
</tbody>
</table>
## Appendix D Surrogate events for each sample

<table>
<thead>
<tr>
<th>Lat10D</th>
<th>Lat10M</th>
<th>Lat30D</th>
<th>Lat30M</th>
<th>Yaw30D</th>
</tr>
</thead>
<tbody>
<tr>
<td>case</td>
<td>crash</td>
<td>case</td>
<td>crash</td>
<td>case</td>
</tr>
<tr>
<td>8307</td>
<td>1</td>
<td>8307</td>
<td>1</td>
<td>8307</td>
</tr>
<tr>
<td>8338</td>
<td>1</td>
<td>8338</td>
<td>1</td>
<td>8338</td>
</tr>
<tr>
<td>8541</td>
<td>0</td>
<td>8541</td>
<td>0</td>
<td>8541</td>
</tr>
<tr>
<td>8549</td>
<td>1</td>
<td>8549</td>
<td>1</td>
<td>8549</td>
</tr>
<tr>
<td>8551</td>
<td>1</td>
<td>8551</td>
<td>1</td>
<td>8551</td>
</tr>
<tr>
<td>8562</td>
<td>0</td>
<td>8562</td>
<td>0</td>
<td>8562</td>
</tr>
<tr>
<td>8567</td>
<td>1</td>
<td>8567</td>
<td>1</td>
<td>8567</td>
</tr>
<tr>
<td>8582</td>
<td>0</td>
<td>8582</td>
<td>0</td>
<td>8582</td>
</tr>
<tr>
<td>8591</td>
<td>0</td>
<td>8591</td>
<td>0</td>
<td>8591</td>
</tr>
<tr>
<td>8643</td>
<td>0</td>
<td>8643</td>
<td>0</td>
<td>8643</td>
</tr>
<tr>
<td>8686</td>
<td>0</td>
<td>8686</td>
<td>0</td>
<td>8686</td>
</tr>
<tr>
<td>8693</td>
<td>0</td>
<td>8693</td>
<td>0</td>
<td>8693</td>
</tr>
<tr>
<td>8712</td>
<td>1</td>
<td>8712</td>
<td>1</td>
<td>8712</td>
</tr>
<tr>
<td>8765</td>
<td>0</td>
<td>8765</td>
<td>1</td>
<td>8765</td>
</tr>
<tr>
<td>8765</td>
<td>1</td>
<td>8765</td>
<td>1</td>
<td>8765</td>
</tr>
<tr>
<td>8812</td>
<td>1</td>
<td>8812</td>
<td>1</td>
<td>8812</td>
</tr>
<tr>
<td>8836</td>
<td>0</td>
<td>8836</td>
<td>0</td>
<td>8836</td>
</tr>
<tr>
<td>8910</td>
<td>0</td>
<td>8910</td>
<td>0</td>
<td>8910</td>
</tr>
<tr>
<td>8920</td>
<td>1</td>
<td>8920</td>
<td>1</td>
<td>8920</td>
</tr>
<tr>
<td>8946</td>
<td>0</td>
<td>8946</td>
<td>0</td>
<td>8946</td>
</tr>
<tr>
<td>9046</td>
<td>0</td>
<td>9046</td>
<td>0</td>
<td>9046</td>
</tr>
<tr>
<td>9081</td>
<td>0</td>
<td>9081</td>
<td>0</td>
<td>9081</td>
</tr>
<tr>
<td>9081</td>
<td>0</td>
<td>9081</td>
<td>0</td>
<td>9081</td>
</tr>
</tbody>
</table>
## Appendix E Conditional Crash Probabilities for Each Sample

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Lat10D</th>
<th>Lat30D</th>
<th>Daylight</th>
<th>Average Conditional Crash Probability</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>Surrogate Event Observed</th>
<th>Crashes Expected</th>
<th>Crashes Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0.19</td>
<td>0.16</td>
<td>0.20</td>
<td>3</td>
<td>0.56</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0.22</td>
<td>0.21</td>
<td>0.23</td>
<td>2</td>
<td>0.45</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0.28</td>
<td>0.28</td>
<td>0.28</td>
<td>1</td>
<td>0.28</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>0.14</td>
<td>0.12</td>
<td>0.16</td>
<td>5</td>
<td>0.69</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0.56</td>
<td>0.37</td>
<td>0.74</td>
<td>5</td>
<td>2.82</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>1</td>
<td>0</td>
<td>0.46</td>
<td>0.17</td>
<td>1.00</td>
<td>9</td>
<td>4.13</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Lat10M</th>
<th>Lat30D</th>
<th>Daylight</th>
<th>Average Conditional Crash Probability</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>Surrogate Event Observed</th>
<th>Crashes Expected</th>
<th>Crashes Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0.13</td>
<td>0.11</td>
<td>0.14</td>
<td>2</td>
<td>0.25</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0.17</td>
<td>0.17</td>
<td>0.17</td>
<td>1</td>
<td>0.17</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0.21</td>
<td>0.21</td>
<td>0.21</td>
<td>1</td>
<td>0.21</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>0.14</td>
<td>0.12</td>
<td>0.16</td>
<td>5</td>
<td>0.71</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0.51</td>
<td>0.30</td>
<td>0.70</td>
<td>3</td>
<td>1.52</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>1</td>
<td>0</td>
<td>0.51</td>
<td>0.18</td>
<td>1.00</td>
<td>8</td>
<td>4.10</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Lat30D</th>
<th>Lat30M</th>
<th>Daylight</th>
<th>Average Conditional Crash Probability</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>Surrogate Event Observed</th>
<th>Crashes Expected</th>
<th>Crashes Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0.24</td>
<td>0.19</td>
<td>0.27</td>
<td>3</td>
<td>0.72</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0.31</td>
<td>0.29</td>
<td>0.33</td>
<td>2</td>
<td>0.62</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
<td>1</td>
<td>0.41</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>0.10</td>
<td>0.08</td>
<td>0.13</td>
<td>5</td>
<td>0.52</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0.79</td>
<td>0.57</td>
<td>0.95</td>
<td>3</td>
<td>2.36</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>1</td>
<td>0</td>
<td>0.54</td>
<td>0.15</td>
<td>1.00</td>
<td>8</td>
<td>4.33</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Yaw30D</th>
<th>Lat30D</th>
<th>Daylight</th>
<th>Average Conditional Crash Probability</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>Surrogate Event Observed</th>
<th>Crashes Expected</th>
<th>Crashes Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0.18</td>
<td>0.15</td>
<td>0.20</td>
<td>3</td>
<td>0.53</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0.28</td>
<td>0.24</td>
<td>0.33</td>
<td>3</td>
<td>0.84</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0.39</td>
<td>0.39</td>
<td>0.39</td>
<td>1</td>
<td>0.39</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>0.10</td>
<td>0.08</td>
<td>0.12</td>
<td>5</td>
<td>0.48</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0.81</td>
<td>0.65</td>
<td>0.89</td>
<td>3</td>
<td>2.44</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>1</td>
<td>0</td>
<td>0.55</td>
<td>0.14</td>
<td>1.00</td>
<td>8</td>
<td>4.37</td>
<td>4</td>
</tr>
</tbody>
</table>
VITA
Kun-Feng Wu

EDUCATION

- M.A. Economics, National Taiwan University, September 2003 – June 2005.

REFEREED JOURNAL PAPERS


AWARDS:

- Research Associateship Award, National Research Council, National Academy, 2012.