Defining, Screening, and Validating Crash Surrogate Events Using Naturalistic Driving Data

Kun-Feng Wu
Ph.D. Candidate, Civil and Environmental Engineering, Larson Institute, Penn State, University Park PA. 16802, Email: kxw930@psu.edu

Paul P. Jovanis
Professor, Civil and Environmental Engineering, Larson Institute, Penn State, University Park PA. 16802, Email: ppj2@engr.psu.edu

Submitted to the 3rd International Conference on Road Safety and Simulation, September 14-16, 2011, Indianapolis, USA

ABSTRACT

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 paper 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, the authors believe the procedure is ready for testing in other applications.

Keywords: traffic safety, crash surrogate, naturalistic driving study.

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, *Number of crashes expected to occur on an entity during a certain period of time (*λ*) = crash-to-surrogate ratio for that entity (*π*)* number of crash surrogates occurring on an entity in that time (c)* or:

$$\lambda = \pi c \tag{1}$$

This statement and its application by several researchers provide support for the view of surrogates as linked to crashes through a ratio, labeled, π.

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 (i.e. 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 1 is an example of search criteria used to identify events for the VTTI 100-car study (Dingus et al., 2005). Vehicle-based accelerometers 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 Time-To-Collision (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 1 Summary of kinematic search criteria for events in VTTI study

Trigger Type	**Description**
1. Lateral Acceleration	• Lateral accel. ≥ 0.7 g.
2. Longitudinal Acceleration	• Accel. or decel. ≥ 0.6g. • Accel. or decal. ≥ 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.

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 geographical 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 reactions time.

While some aspects of events 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 times, 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 relations ship between them and crashes.

STUDY GOALS

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 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.

METHODOLOGY

Figure 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 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.

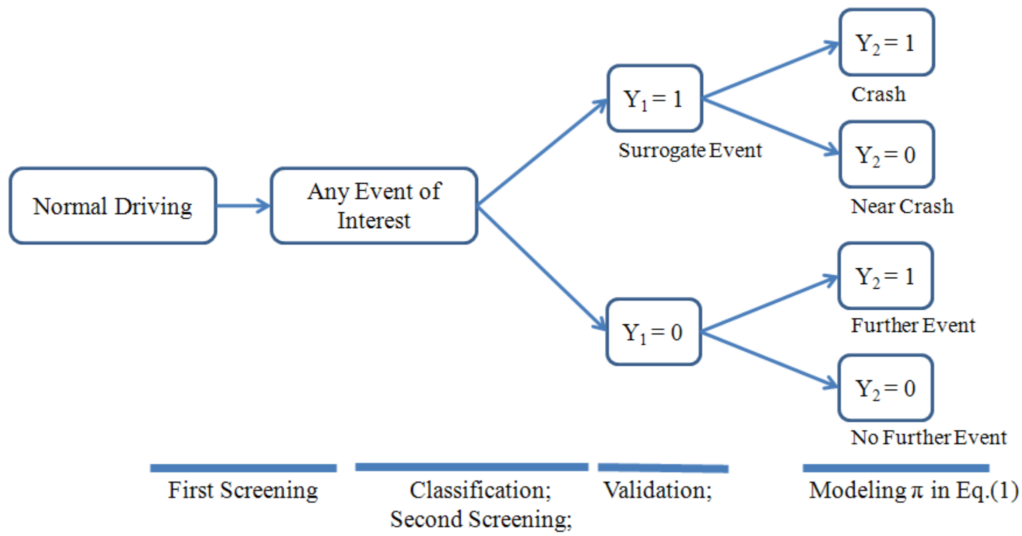


Figure 1 Conceptualization of the Relationship between Crashes and Near Crashes in Naturalistic Driving Data

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 (1).

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 $Prob(Y_2 = 1|Y_1 = 1)$, 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 would end up with crashes; this is, in fact the "π" 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-examine to be sure there are no further events of interest ($Y_2 = 1$) although this is not conducted in this paper. 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.

The Analytic Procedure

Figure 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 paper, we provide rather detailed descriptions of each step and the methods applied to our data set.

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 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.

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 TPR(c) and FPR(c), respectively.

$$TPR(c) = True\ Positive\ Rate(c) = P(Z \geq c|Y = 1) \tag{2}$$

$$FPR(c) = False\ Positive\ Rate(c) = P(Z \geq c|Y = 0) \tag{3}$$

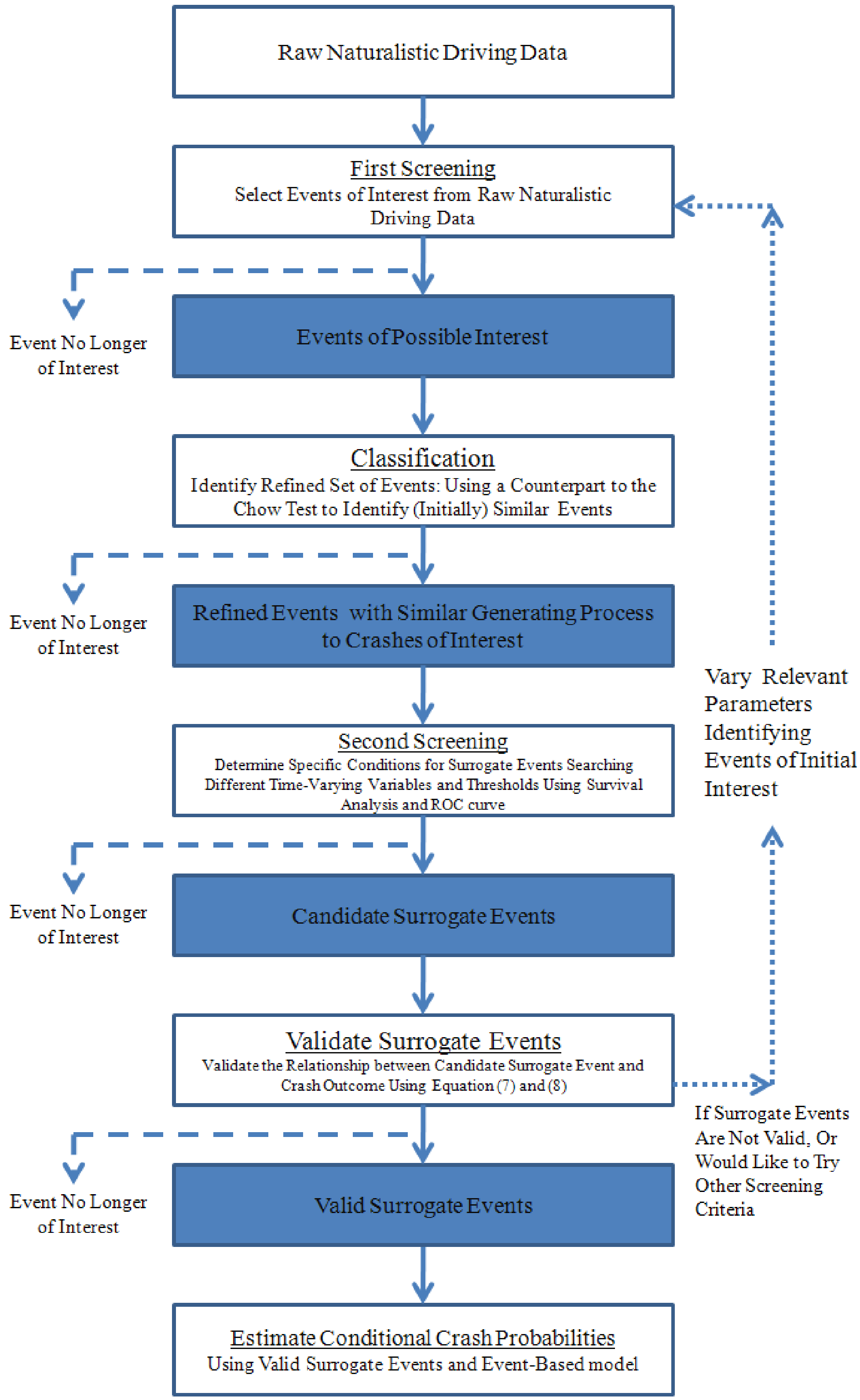


Figure 2. Analytical Procedure for Analysis and Validation of Surrogate Events

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).

<u>Receiver Operating Characteristic (ROC) Regression</u>

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 brake pedal harder than female drivers (i.e., decelerate faster), then gender is associated with the marker deceleration. Therefore, 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

$$ROC_Z(f) = \Phi\{\alpha_0 + \alpha_1 \Phi^{-1}(f) + \alpha_2 M\} \tag{4}$$

where Φ is the standard normal, *f* is a discrete set of FPR points, and α_0, α_1 and α_2 are estimated parameters. If α_2 is positive then an increase of *M* enhances the accuracy of the marker.

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?

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. 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.

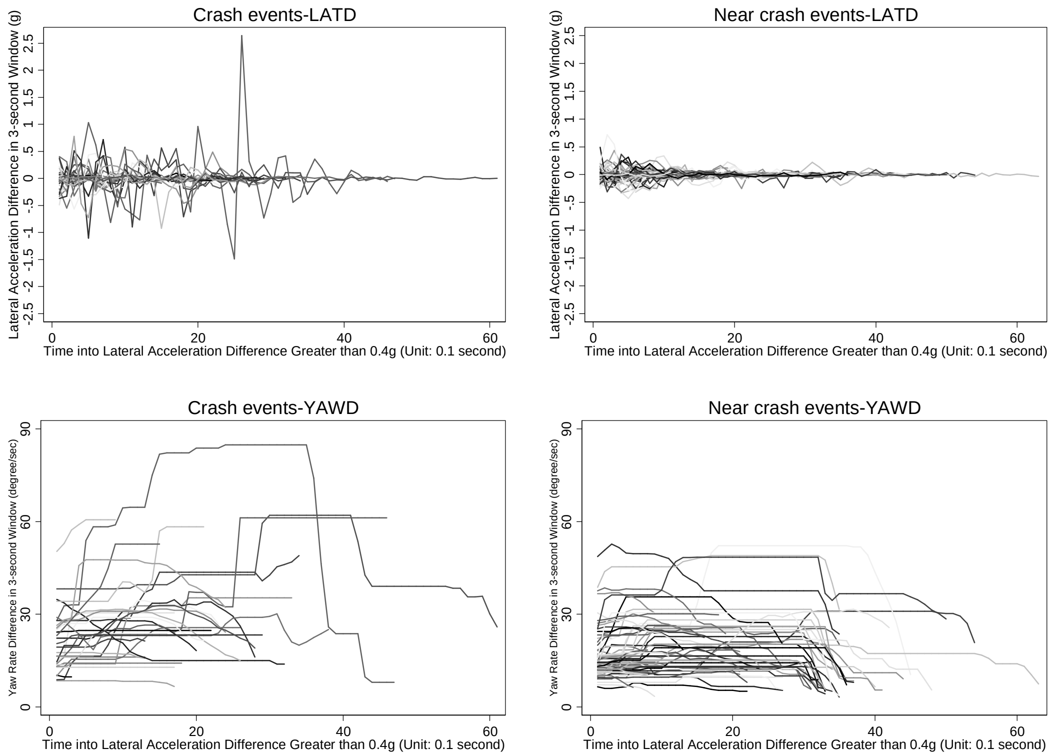


Figure 3 Illustrative Example for Data at Second Screening

One can see that the vehicle kinematics for crash events (left side of figure) 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. 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 possible 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 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 on pass through the data.

Validation

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 other 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:

$$\Pr(Y_2 = 1|Y_1 = 1, \mathbf{X}) \cong 1 \tag{5}$$

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 positive one.

$$\mathrm{Cov}(Y_1, Y_2|\mathbf{X_1}, \mathbf{X_2}) \cong 1 \tag{6}$$

where $\mathbf{X}_1$ and $\mathbf{X}_2$ represents factors that affect Y_1 and Y_2 respectively.

Equation (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**)**.

$$\Pr(Y_2 = 1|Y_1 = 1, \mathbf{X}) \gg 0 \tag{7}$$

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

$$\mathrm{Cov}(Y_1, Y_2|\mathbf{X_1}, \mathbf{X_2}) \gg 0 \tag{8}$$

In this study, a bivariate Probit model is applied to test the 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 $\Pr(Y_2 = 1|Y_1 = 1, \mathbf{X})$; 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 Equation (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 Equation (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).

<u>Bivariate Probit Model</u>

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, otherwise), and Y_2=1 indicates a crash occurrence (Y_2=0, near crash). The surrogate event and crash generating processes can be written as:

$$Y_1^* = \mathbf{X}_1'\boldsymbol{\beta}_1 + \varepsilon_1,\ Y_1 = 1\ if\ Y_1^* > 0, 0\ otherwise \tag{9}$$

$$Y_2^* = \mathbf{X}_2'\boldsymbol{\beta}_1 + \varepsilon_2,\ Y_2 = 1\ if\ Y_2^* > 0, 0\ otherwise \tag{10}$$

$$E(\varepsilon_1|\mathbf{X_1}, \mathbf{X_2}) = E(\varepsilon_2|\mathbf{X_1}, \mathbf{X_2}) = 0 \tag{11}$$

$$Var(\varepsilon_1|\mathbf{X_1}, \mathbf{X_2}) = Var(\varepsilon_2|\mathbf{X_1}, \mathbf{X_2}) = 1 \tag{12}$$

$$\mathrm{Cov}(Y_1, Y_2|\mathbf{X_1}, \mathbf{X_2}) = \mathrm{Cov}(\varepsilon_1, \varepsilon_2|\mathbf{X_1}, \mathbf{X_2}) = \rho \tag{13}$$

And the bivariate normal cumulative density function is

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

$$\phi_2(z_1, z_2, \rho) = \frac{e^{-\left(\frac{1}{2}\right)(x_1^2 + x_2^2 - 2\rho x_1 x_2)/(1-\rho^2)}}{2\pi(1-\rho^2)^{1/2}} \tag{15}$$

Estimating the Conditional Crash Probability Using Valid Surrogate Events

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

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 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 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 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 2.

Table 2 Variable Definitions

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

DATA ANALYSIS

This section demonstrates how the whole procedure for screening, identifying, and validating surrogate events is implemented as shown in Figure 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.

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 event of interest. As shown in the left panel in Figure 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 event 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 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 3-second window were selected as the marker (the first screening criterion) for examining their accuracy for detecting event of interest (period B in Figure 4). The application of the ROC analysis is summarized in Figure 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.

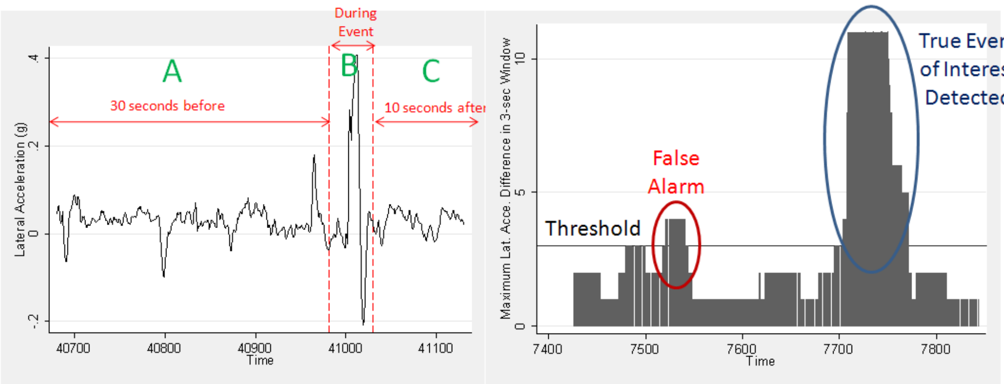
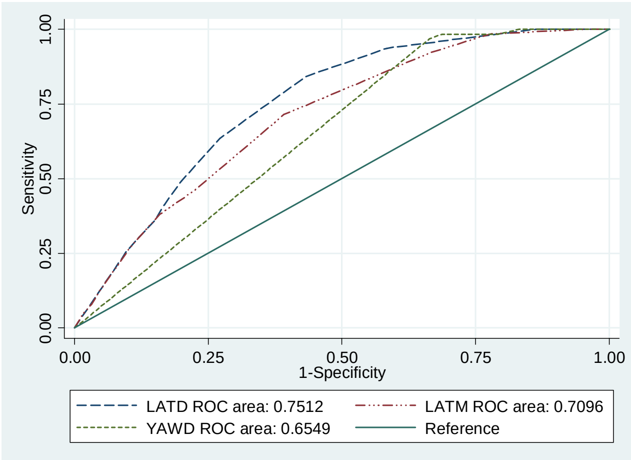


Figure 4 (Left) Event of Interest vs. (Right) Event Not of Interest: The Impact of The Selection of the Threshold.



- Ho: area(LATD) = area(LATM)
  chi2(1) = 5.25, p-value = 0.02
- Ho: area(LATD) = area(YAWD)
  chi2(1) = 9.72, p-value = 0.002

Figure 5 ROC Curves for LATD, LATM, and YAWD

Note that at this step, the goal is to detect as many as 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 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 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 10 percentage points less specificity. This confirms that lateral acceleration rate difference can perform better than maximum lateral acceleration.

In this study, we will carry events with LATD greater than 0.4g during the entire events into the next step as a demonstration of this procedure. Using LATD greater than 0.4g, there are total 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 classification stage to test the need of further classification.

Table 3 ROC Curves Analysis for LATM and LATD

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

Classification

One of the characteristics of valid surrogate events is the similarity among them, no matter whether they end up with crashes or near crashes. This suggests that similar surrogate events should have similar crash probabilities in terms of context. Figure 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-intersection segments. This suggests that some of or all the regression coefficients are different for the two subsets of the data.

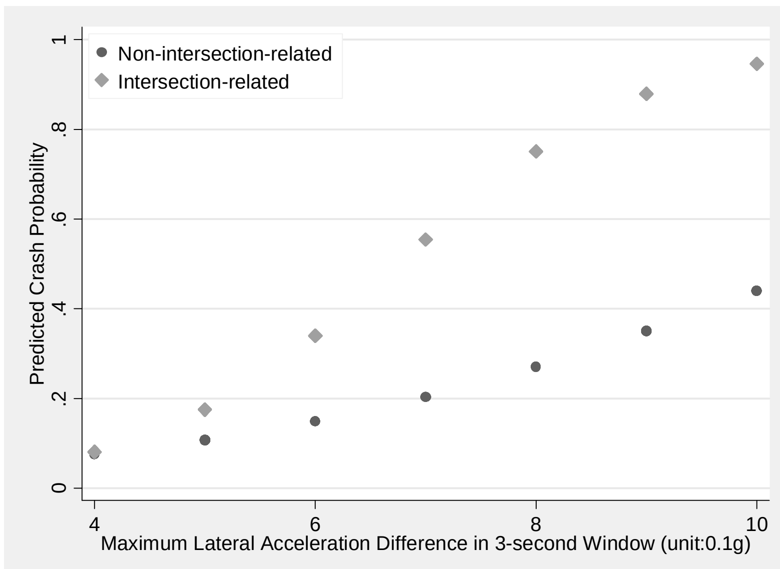


Figure 6. Structural Difference on Crash Probabilities between Events with Different Context

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 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 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 coefficient 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*[-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 4, it 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 4 Chow Test for Intersection v.s. Non-intersection Related Events

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

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 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 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 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	0.59
LATM	0.03	0.24	0.12	0.90	-0.44	0.50
YAWD	0.07	0.33	0.21	0.83	-0.58	0.72

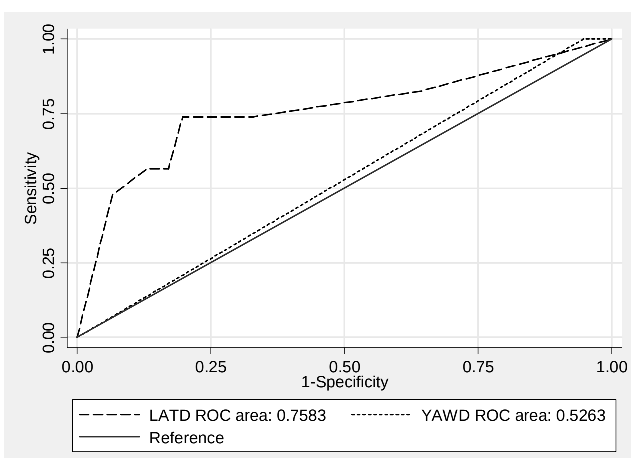


Figure 7 ROC Curve for LATD and YAWD at Second Screening

Table 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 6 ROC Curves Analysis for LATD at Second Screening

Cut-off point	Sensitivity	Specificity	Near Crashes	Crashes
>= 0.4g	100.00%	0.00%	24	3
>= 0.5g	82.61%	35.53%	21	2
>= 0.6g	73.91%	67.11%	8	0
>= 0.7g	73.91%	80.26%	2	2
>= 0.8g	56.52%	82.89%	2	0
>= 0.9g	56.52%	86.84%	5	1
>= 1.0g	47.83%	93.42%	5	4
>1.0g	0.00%	100.00%	0	2

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 $\rho = 0$: chi^2 (1) = 7.55, p-value = 0.006). The results in Table 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 probability of crash occurrence. Daylight condition would reduce the probability of crash, though not significant.

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

	Coef.	Std. Err.	z	P>z	95% CI	
LATD >= 0.7g Equation						
Deceleration Rate (g)	-1.09	0.81	-1.35	0.18	-2.68	0.50
Speed (mph)	0.01	0.01	1.33	0.18	-0.01	0.03
Constant	-1.10	0.36	-3.03	0.00	-1.81	-0.39
Crash Occurrence Equation						
Daytime Condition	-0.36	0.31	-1.15	0.25	-0.97	0.25
Constant	-0.80	0.23	-3.53	0.00	-1.25	-0.36
ρ	0.65	0.18			0.15	0.88

This study 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 at daytime condition are invalid surrogate events. Hence, in this study, the specific conditions for defining surrogate events are events that are: (1) with LATD greater than or equal to 0.4g during the entire events, (2) non-intersection related, and (3) with LATD greater or equal to 0.9g and events with LATD between 0.8g to 0.9g at nighttime condition. A review of narratives accompanying the VTTI data revealed that the identified grouping of crashes and near crashes appear to be qualitatively quite 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.

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 8 shows the appropriateness of this model specification. Although the suspected endogeneity is not statistically significant (Wald test of exogeneity: chi2(1) = 1.28, Prob > $chi^2 = 0.26$), this model is still a more generalized form of 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 9.

Table 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 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, 8 crashes is 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 9 Conditional Crash Probabilities Using Valid Surrogate Events

Scenario	Daytime/ Nighttime	LATD	Average Conditional Crash Probability	Lower Bound	Upper Bound	Surrogate Event Observed	Crashes Expected	Crashes Observed
1	Night	>= 0.8g	0.08	0.07	0.08	2	0.16	0
2	Night	>= 0.9g	0.13	0.13	0.13	1	0.13	0
3	Day	>= 0.9g	0.09	0.07	0.12	5	0.45	1
4	Night	>= 1.0g	0.57	0.26	0.86	3	1.71	2
5	Day	>= 1.0g	0.56	0.13	1.00	8	4.48	4

CONCLUSION AND DISCUSSION

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 paper demonstrates that a multi-stage modeling framework can make the analysis of naturalistic driving data tractable, without substantial use of video screening.

The paper 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 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 occuring 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 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.
- For valid surrogate events, the same maximum lateral acceleration difference, during daytime has a lower crash probability than during nighttime.

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 of the sample size (Guo et al., 2010). The team's 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 (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.

It is hoped that this paper 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.

REFERENCE

Archer, J. (2004). Methods for the Assessment and Prediction of Traffic Safety at Urban Intersections and their Application in Micro-simulation Modeling, Royal Institute of Technology.

Bogard, S., P. Fancher, R. Ervin, M. Hagan, and Z. Bereket (1998). *Intelligent Cruise Control Field Operational Test, Final Report*, Volume III, prepared for National Highway Traffic Safety Administration.

Cameron, A. C., Trivedi, P. K., (2005). *Microeconometrics: Methods and Applications*, Cambridge University Press, 1034 pages.

Chin, H.C., Quek, S.T. (1997). Measurement of traffic conflicts. *Safety Science*. 26 (3), pp. 169–185.

Datta, T.K. (1979). Accident surrogates for use in analyzing highway safety hazards. In: Proceedings, Second International Traffic Conflict Technique Workshop, pp. 4–20.

Dingus, T.A., S.G. Klauer, V.L. Neale, A. Petersen, S.E. Lee, J. Sudweeks, M.A. Perez, J. Hankey, D. Ramsey, S. Gupta, C. Bucher, Z.R. Doerzaph, J. Jermeland, and R.R. Knipling (2005). *The 100-Car Naturalistic Driving Study, Phase II—Results of the 100-Car Field Experiment*, National Highway Traffic Safety Admin (DOT HS 810 593).

Grayson, G.B., Hakkert, A.S. (1987). Accident Analysis and Conflict Behaviour, In J. Rotherngatter and R.A. de Bruine (Eds.), Road-user and Traffic Safety, Van Gorcum. pp. 27-59.
Greene, W.H. (2003). Econometric Analysis. 5_{th} edition, Prentice Hall, New York.

Guo, F., Klauer, K.G., Hankey, J.M., Dingus, T.A., 2010. Near crashes as Crash surrogate for Naturalistic driving studies, *Transportation Research Record*, No. 2147, pp. 66-74.

Hanowski, R.J., M. A. Perez, and T. A. Dingus (2005). "Driver distraction in long-haul truck drivers", *Transportation Research Part F: Traffic Psychology and Behaviour*, Volume 8, Issue 6, pp. 441-458.

Hanowski, R.J., J. Hickman, M. C. Fumero, R.L. Olson, and T.A. Dingus (2007a). "The sleep of commercial vehicle drivers under the 2003 hours-of-service regulations", *Accident Analysis & Prevention*, Volume 39, Issue 6, pp. 1140-1145.

Hanowski, R.J., J. Hickman, W.W. Wierwille, and A. Keisler (2007b). "A descriptive analysis of light vehicle-heavy vehicle interactions using in situ driving data", *Accident Analysis & Prevention*, Volume 39, Issue 1, Pages 169-179.

Hauer, E. (1982). Traffic Conflicts and Exposure. *Accident Analysis & Prevention*, 14(5), pp. 359-364.

Hauer, E., Garder, P. (1986). Research into the Validity of the Traffic Conflicts Technique. *Accident Analysis & Prevention*, 18(6), pp. 471-481.

Hensher, D.A., Mannering, F.L. (1994). Hazard Based Duration Models and Their Application to Transport Analysis. *Transport Reviews*, 14(1), pp. 63-82.

Hauer, E. (1999). Safety in Geometric Design Standards. white paper.

Hydén, C. (1987). The Development Of A Method For Traffic Safety Evaluation: The Swedish Traffic Conflicts Technique. Department of Traffic Planning and Engineering, Lund University, Lund, Sweden.

Jovanis, P.P., Shankar, V., Aguero-Valverde, J., Wu, K., Greenstein, A. (2010), Analysis of Existing Data: Prospective Views on Methodological Paradigms, Draft Final Report, Larson Transportation Institute, Department of Civil and Environmental Engineering, Pennsylvania

State University, University Park, PA, Prepared for The Strategic Highway Research Program 2, Transportation Research Board of The National Academies.

LeBlanc, D., J. Sayer, C. Winkler, S. Bogard, J. Devonshire, M. Mefford, M. Hagan, A. Bareket, R. Goodsell, and T. Gordon (2006). *Road departure crash warning system (RDCW) field operational test final report.* Washington, DC: U.S. Department of Transportation.

McGehee, D. V., Boyle, L. N., Hallmark, S. J., Lee, D., Neyens, D. M., Ward, N. J. (2010). S02 Integration of Analysis Methods and Development of Analysis Plan Phase II Report.

Peat, J, Barton, B. (2005). Medical Statistics: A guide to data analysis and critical appraisal, Oxford, U.K.: Blackwell Publishing Ltd.

Pepe, M.S. (2003). The statistical evaluation of medical tests for classification and prediction, Oxford University Press, New York, NY.

Shankar, V., P.P. Jovanis, J. Agüero, and F. Gross (2008). Analysis of Naturalistic Driving Data: A Prospective View on Methodological Paradigms. *Transportation Research Record*, 2061, pp.1-8.

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

Swets, J.A. (1988). Measuring the accuracy of diagnostic systems. Science, 240, pp. 1285-1293.

Tarko, A.P., Davis, G, Saunier, N., Sayed, T., Washington, S. (2009). Surrogate Measures of Safety, white paper.

University of Michigan Transportation Research Institute and General Motors Research and Development Center (2005). *Automotive Collision Avoidance System Field Operational test report: Methodology and Results*, Final Research Report, Contract DTNH22-99-H-07019, prepared for National Highway Traffic Safety Administration, 928p.

Web site: http://www.vtti.vt.edu/casr-research-details.html#teen-nds Searched July 22, 2010.

Williams, M. J. (1981). Validity of the traffic conflicts technique. *Accident Analysis & Prevention*, 13(2), pp.133-135.

Wu, K., Jovanis, P.P. (2011). The Relationships of Crashes and Crash-Surrogate Events in Naturalistic Driving, in Review, *Accident Analysis & Prevention*.

Xu, J., Long, S. (2005). Confidence Intervals for Predicted Outcomes in Regression Models for Categorical Outcomes. *The Stata Journal*, 5(4), pp. 537-559.

Appendix

- 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.